Heft 7
München, Oktober 1982

International Federation of Surveyors
– F I G –

Proceedings

Survey Control Networks

Meeting of Study Group 5B
7th – 9th July, 1982
Aalborg University Centre, Denmark

Editors: K. Borre
W. M. Welsch

SCHRIFTENREIHE

Wissenschaftlicher Studiengang Vermessungswesen
Hochschule der Bundeswehr München

HSBw

Heft 7
München, Oktober 1982

International Federation of Surveyors
– F I G –

Proceedings

Survey Control Networks

Meeting of Study Group 5B
7th – 9th July, 1982
Aalborg University Centre, Denmark

Editors: K. Borre
W. M. Welsch

SCHRIFTENREIHE

Wissenschaftlicher Studiengang Vermessungswesen
Hochschule der Bundeswehr München

HSBw

Der Druck dieses Heftes wurde aus Haushaltsmitteln der Hochschule der Bundeswehr München gefördert.

Verantwortlich für die Herausgabe der Schriftenreihe:
Prof. Dr. G. Neugebauer
Dipl.-Ing. J. Peipe

Bezugsnachweis:

Studiengang Vermessungswesen
im Fachbereich Bauingenieur- und Vermessungswesen
Hochschule der Bundeswehr München
Werner-Heisenberg-Weg 39
8014 Neubiberg

ISSN 0173-1009

PREFACE

Study Group 5B of the International Federation of Surveyors (FIG) was called into existence in 1980 when it became obvious that the work of the huge Commission 5 of FIG could not reasonably be continued without splitting up its activities into several study groups. One of them has become Study Group 5B on "Survey Control Networks". During the Symposium of Commission 5 on "Automated Processing of Surveying Data" in Varna, Bulgaria, 1980, the idea was born to hold a meeting of SG 5B in Aalborg, Denmark.

The plan could be realized at the University Centre of Aalborg from 7th - 9th July, 1982. The general subject of the meeting has been dedicated to the quality of networks according to the recommendation of the FIG-Congress in Montreux, 1981, to emphasize the evaluation of control networks under the aspects of stochasticity. The five main topics: Data Processing and Adjustment, Interaction of Stochastic and Functional Models, Quality of Networks, Network Densification, and Optimization of Control Networks, can be seen from this point of view.

The meeting intended to stimulate information exchange among researchers and practitioners. Fifty participants coming from fifteen countries, discussed theoretical and practical problems. The attempt was made to have a good theory applied to a better practice. Moreover, with respect to the XVII. FIG-Congress in Sofia, Bulgaria, 1983, another purpose of the meeting was to clarify the state of art and to work out those points which should be treated with particular emphasis in the next future. As a result, in the closing session the participants agreed to the recommendations listed below. Thus, the meeting functioned as a sign-post: marking the stand-point and showing the direction.

The proceedings have been edited in the most simple way by printing the manuscripts delivered by the authors ready for photo reproductions. This procedure provides a publication as fast as possible.

The meeting itself and the publication of the proceedings was sponsored by DEN DANSKE LANDINSPEKTØRFORENING, the AALBORG UNIVERSITY CENTRE and the HOCHSCHULE DER BUNDESWEHR MÜNCHEN. The organizers are grateful to the

sponsors. However the thank of the editors goes particularly to the authors and to all those unnamed persons who directly or indirectly contributed to the meeting which was held in a very competent, friendly and relaxed atmosphere.

Kai Borre
Walter M. Welsch

TABLE OF CONTENTS

		PAGE
PREFACE		1
PROGRAMME OF SESSIONS		5
RECOMMENDATIONS		7
LIST OF PARTICIPANTS		9
PAPERS:		
ANDERSON, E.G.,	Towards total optimization of surveying and mapping systems	13
ANDERSSON, O., ENGSAGER, K., WENG, W.L.,	Computation and data-management of blocks of medium size geodetic networks	33
AUGATH, W.,	Accuracy and reliability measures concerning design and qualification of densification networks	51
BANOV, B.,	A special method to derive a criterion matrix	65
BARBARELLA, M., UNGUENDOLI, M.,	Analysis of some densification networks in Italy	79
BENCIOLINI, B., BETTI, B., MUSSIO, L.,	A remark on the application of the GPS procedure to very irregular graphs	97
BLACHNITZKY, K.,	Status and problems of official horizontal densification networks as shown by the Federal Republic of Germany	103
BUITEN, H.J., RICHARDUS, P.,	Junction of control surveys by adjustment compared with coordinate transformation	115
CROSILLA, F.,	A criterion matrix for the second order design of control networks	143
CROSS, P., FAGIR, A.,	Procedures for the first and second order design of vertical control networks	159
DARE, P., VANÍČEK, P.,	Strength analysis of horizontal networks using strain	181
FÖRSTNER, W.,	Systematic errors in photogrammetric point determination	197
DE HEUS, H.,	Data-snooping in control networks	211
DE HEUS, H.,	Quality related problems of densification networks	225
HRADILEK, L.,	Horizontal and vertical survey control by three-dimensional triangulation	239

		PAGE
KRUMM, F.,	Criterion matrices for estimate quantities	245
MIERLO, J. van,	Difficulties in defining the quality of geodetic networks	259
NIEMEIER, W.,	Principal component analysis and geodetic networks - some basic considerations	275
NINKOV, T.,	A new method of land surveying networks optimization	293
PAPO, H.B., PERELMUTER, A.,	Densification of a network as a special case of deformation analysis	301
PELZER, H.,	Influence of systematic effects in stochastic and functional models	309
PEROVIĆ, G.,	Some estimable functions in geodetic networks	321
PODER, K.,	Data processing and adjustment	327
QUEE, H.,	On the stochasticity of alignment-functions for automated track maintenance machines	339
REMMER, O.,	Modelling errors in geometric levelling	355
SCHMITT, G.,	Optimization of control networks - state of the art	373
SJÖBERG, L.,	Station adjustment of directions using generalized inverses	381
WELSCH, W.M.,	Network densification - problems and solutions	401
ZLATANOW, G.,	The increment method	415

PROGRAMME OF THE MEETING

DATA PROCESSING AND ADJUSTMENT

July 7, 1982 09.20 - 12.20

Chairman: K. PODER

K. PODER, Data processing and adjustment, review paper

L. HRADILEK, Horizontal and vertical survey control by three-dimensional triangulation, contributed paper

L. SJÖBERG, Station adjustment of directions using generalized inverses, contributed paper

H.M. de HEUS, Data snooping in control networks, contributed paper

O. ANDERSSON, K. ENGSAGER, W. WENG, Computation and data management of blocks of medium size geodetic networks, contributed paper, presented by O. ANDERSSON

INTERACTION OF STOCHASTIC AND FUNCTIONAL MODELS

July 7, 1982 13.30 - 15.30

Chairman: H. PELZER

H. PELZER, Influence of systematic effects in stochastic and functional models, review paper

W. FÖRSTNER, Systematic errors in photogrammetric point determination, contributed paper

W. NIEMEIER, Principal component analysis and geodetic networks - some basic considerations,
contributed paper, not presented

O. REMMER, Modelling errors in geometric levelling, contributed paper

• • •

G. ZLATANOV, The increment method, personal paper

B. BENCIOLINI, B. BETTI, L. MUSSIO, A remark on the application of the GPS procedure to very irregular graphs, personal paper, presented by L. MUSSIO

QUALITY OF NETWORKS

July 8, 1982 09.00 - 12.00

Chairman: J. van MIERLO

J. van MIERLO, Difficulties in defining the quality of geodetic networks, review paper

H.M. de HEUS, Quality related problems of densification networks, contributed paper

H. QUEE, On the stochasticity of alignment-functions for automated track maintenance machines, contributed paper

P. DARE, P. VANÍČEK, Strength analysis of horizontal networks using strain, personal paper, presented by P. DARE

NETWORK Ddensification

July 9, 1982 09.00 - 12.00

Chairman: W.M. WELSCH

W.M. WELSCH, Network densification - problems and solutions, review paper

K. BLACHNITZKY, Status and problems of official horizontal densification networks as shown by the Federal Republic of Germany, contributed paper

M. BARBARELLA, M. UNGUENDOLI, Analysis of some densification networks in Italy, personal paper, presented by M. UNGUENDOLI

W. AUGATH, Accuracy and reliability measures concerning design and qualification of densification networks, contributed paper

H.J. BUITEN, P. RICHARDUS, Junction of control surveys by adjustment compared with coordinate transformation, contributed paper, presented by P. RICHARDUS

H.B. PAPO, A. PERELMUTER, Densification of a network as a special case of deformation analysis, personal paper, presented by A. PERELMUTER

OPTIMZATION OF CONTROL NETWORKS

July 9, 1982 13.00 - 16.00

Chairman: G. SCHMITT

G. SCHMITT, Optimization of control networks - state of the art, review paper

B. BANOV, A special method to derive a criterion matrix, contributed paper, presented by G. ZLATANOV

E.G. ANDERSON, Towards total optimization of surveying and mapping systems, personal paper

P. CROSS, A. FAGIR, Procedures for the first and second order design of vertical control networks, contributed paper, presented by P. CROSS

F. KRUMM, Criterion matrices for estimable quantities, contributed paper

F. CROSILLA, A criterion matrix for the second order design of control networks, personal paper

• • •

The personal papers

T. NINKOV, A new method of land surveying networks optimization

G. PEROVIĆ, Some estimable functions in geodetic networks

were not presented.

RECOMMENDATIONS

The participants agreed that the following subjects should be emphasized in the next future:

DATA PROCESSING AND ADJUSTMENT

The development of integrated DP-systems with respect to

1. Objective criteria
   - 1.1 Scope (e.g. general purposes, engineering, cadastral, scientific)
   - 1.2 Data types (e.g. coordinate systems, observation data units etc.)
   - 1.3 Capacity
   - 1.4 Numerical quality
   - 1.5 Data base interfacing
   - 1.6 Language portability

2. Quality information

3. Processing information and reporting
   - 3.1 Standard deviation
   - 3.2 Numerical precision
   - 3.3Blunder detection, data snooping, outliers

4. Special modules
   - 4.1 Normal equations
   - 4.2 Weight assignements
   - 4.3 Preliminary coordinates

5. Connection to physical geodesy

INTERACTION OF STOCHASTIC AND FUNCTIONAL MODEL

1. The extension of the stochastic as well as the functional model of networks considering systematic effects (i.e. refraction-, instrumental-, calibration-, centering-errors).
2. Theoretical investigations and practical experiments for detecting and eliminating systematic effects in large networks.

QUALITY OF NETWORKS

1. The sensitivity and reliability of a network with respect to the testing of deformation models.
2. The use of criterion matrices in the design of densification networks. Localization of weak parts of a network. The use of substitute matrices.
3. Application of strain analysis techniques with respect to the reliability of a network.

NETWORK DENSIFICATION

1. Criteria for "best" solutions considering different circumstances under which network densification has to be carried out.
2. The possibilities which photogrammetric networks can add to terrestrial methods of network densificartion, as a partner of comparable capability.
3. How to find a compromise for the diverging results of network designs which had to be established for an optimum of accuracy and reliability respectively.
4. Numerical comparison of the results of different network densification techniques, using a set of real data.

OPTIMIZATION OF CONTROL NETWORKS

1. Development of criterion matrices being appropriate and realizable.
2. Reliability criteria for the Second Order Design considering additional constrain.
3. Problems of reliability when reducing a Second Order Design for the optimization of the First Order Design.
4. Third Order Design of densification networks.

LIST OF PARTICIPANTS

Alberda, J.E.,	Delft University of Technology, Afd. Geodesie, Thijsseweg 11, NL-2629 JA Delft, The Netherlands
Anderson, E.G.,	Division of Surveying Engineering, Univ. of Calgary, 2500, University Drive NW, Calgary, Alb., Canada, T2N 1N4
Andersson, O.,	Geodætisk Institut, Svanemøllevej 85, DK-2900 Hellerup, Denmark
Augath, W.,	Landesvermessungsamt Niedersachsen, Lönsweg 2, D-3257 Springe 1, Fed. Rep. of Germany
Barbarella, M.,	University of Bologna, Viale Risorgimento 2, I-40136 Bologna, Italy
Blachnitzky, K.,	Bayerisches Landesvermessungsamt, Alexandrastraße 4, D-8000 München 22, Fed. Rep. of Germany
Borre, K.,	Institute of Development and Planning, Aalborg University Center, Fibingerstraede 11, DK-9100 Aalborg, Denmark
Coker, O.,	Kenting Africa, 53, Lawson Street, P.O. Box 3660, Lagos, Nigeria
Crosilla, F.,	Istituto di Strade e Trasporti, Università di Trieste, Via Valerio, 6/1, I-34127 Trieste, Italy
Cross, P.A.,	North East London Polytechnic, Forest Road, London E17 4JB, England
Dare, P.,	Dept. Surveying Science, University of Toronto, Erindale Campus, Mississauga, Ontario, Canada, L5L 1C6
Denekamp, J.,	Hoofdirectie Kadaster, Waltersingel 1, NL-7314 NK Apeldoorn, The Netherlands
Diehl, W.,	Technische Universität Berlin, Straße des 17. Juni 135, Sekr. H21, D-1000 Berlin, Fed. Rep. of Germany
Eeg, J.,	Geodætisk Institut, Gamlehave Alle 22, DK-2920 Charlottenlund, Denmark
Engsager, K.,	Geodætisk Institut, Svanemøllevej 85, DK-2900 Hellerup, Denmark
Fagir, A.H.,	Land Surveying Department - NELP, Forest Road, London E17 4JB, England
Funcke, G.,	Institut für Markscheidewesen, TU Clausthal, Erzstr. 18, D-3392 Clausthal-Zellerfeld, Fed. Rep. of Germany
Förstner, W.,	Institut für Photogrammetrie, Keplerstr., 11, D-7000 Stuttgart 1, Fed. Rep. of Germany
Gervaise, J.,	CERN - Geneva CH-1211 Geneva 23, Switzerland
Gründig, L.,	Universität Stuttgart, Keplerstr., 11, D-7000 Stuttgart 1, Fed. Rep. of Germany
Heer, R.,	Sonderforsch. Bereich 1491 P3, Universität Hannover, Nienburger Straße 1, D-3000 Hannover 1, Fed. Re. of Germany

Heus de, H.M.,	Delft Geodetic Institute, Thijsseweg 11, NL-2629 JA Delft, The Netherlands
Holsen, J.,	Institut for Geodesie og Fotogrammetri, N-7034 Trondheim NTH, Norway
Hradilek, L.,	Faculty of Sciences, Charles University, CZ-128 43 Praha 2, Albertov 6, Czechoslovakia
Jensen, J.E.,	Matrikeldirektoratet, Titangade 13, DK-2200 København N, Denmark
Krumm, F.,	Geodetic Institute, University Stuttgart, Keplerstr., 11, D-7000 Stuttgart 1, Fed. Rep. of Germany
Leonhard, T.,	Universität Hannover, SFB 149, Nienburger Straße 1, D-3000 Hannover 1, Fed. Re. of Germany
Mark, A.N.,	Technical University of Denmark, Landmalervej 7, DK-2800 Lyngby, Denmark
Mayoud, M.,	CERN - Geneva CH-1211 Geneva 23, Switzerland
Mierlo van, J.,	Geodätisches Institut, Universität Karlsruhe, Englerstraße 7, D-7500 Karlsruhe, Fed. Rep. of Germany
Mittermayer, E.,	Techn. Universität Berlin - Höhere Geodäsie, Straße des 17. Juni 135, D-1000 Berlin 12, Fed. Rep. of Germany
Murnane, B.,	Surveying Department, University of Melbourne, Grattan St. Parkville, 3052, Victoria, Australia
Mussio, L.,	Ist. di Topografia, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
Ninkov, T.,	Gradjevinski Fakultet, Institut za Geodezia, YU-11000 Beograd, Bulevar Revolucia 73/I, Yugoslavia
Pelzer, H.,	Geodätisches Institut, Universität Hannover, Nienburger Straße 1, D-3000 Hannover 1, Fed. Re. of Germany
Perelmuter, A.,	University Tel Aviv, Dep. of Geodesy, Ramat Aviv, Israel
Perović, G.,	Gradjevinski Fakultet, Institut za Geodezia, YU-11000 Beograd, Bulevar Revolucia 73/I, Yugoslavia
Poder, K.,	Geodætisk Institut, Gamlehave Alle 22, DK-2920 Charlottenlund, Denmark
Quee, H.,	Netherlands Railways (Is 1), Moreelsepark 1, NL-3500 GB Utrecht, The Netherlands
Remmer, O.,	Geodætisk Institut, Gamlehave Alle 22, DK-2920 Charlottenlund, Denmark
Richardus, P.,	Dept. Surveying and Photogramm., Agricult. University, P.O. Box 339, NL-6700 AH Wageningen, The Netherlands
Schaffrin, B.,	Institut für Theoretische Geodäsie, Nußallee 17, D-5300 Bonn 1, Fed. Rep. of Germany
Schmitt, G.,	Geodätisches Institut, Universität Karlsruhe, Englerstraße 7, D-7500 Karlsruhe, Fed. Rep. of Germany

Sjöberg, L.,	National Land Survey, S-80112 Gävle, Sweden
Unguendoli, M.,	University of Bologna, Viale Risorgimento 2, I-40136 Bologna, Italy
Weise, W.,	Geodätisches Institut, Universität Hannover, Nienburger Straße 1, D-3000 Hannover 1, Fed. Re. of Germany
Welsch, W.M.,	Geodätisches Institut, Hochschule der Bundeswehr München, D-8014 Neubiberg, Fed. Rep. of Germany
Weng, L.W.,	Geodætisk Institut, Gamlehave Alle 22, DK-2920 Charlottenlund, Denmark
Zlatanov, G.,	FIG Bureau, P.O. Box 1386, Sofia 1000, Bulgaria

TOWARDS TOTAL OPTIMIZATION OF SURVEYING AND MAPPING SYSTEMS

Edward G. ANDERSON
Calgary, Alberta, Canada

ABSTRACT

The concept of a "black box" surveying and mapping ***system*** is defined. The term ***total optimization*** is introduced to describe the process of applying optimization theory to the control of both "quality" and "efficiency" of the surveying and mapping system. Approaches to this problem are explored, and plans for one practical methodology currently being implemented are presented. Some successful applications of operations research technology to surveying and mapping system problems are discussed. The problem of developing a fully-integrated, simultaneous model for total optimization is defined and considered.

1. INTRODUCTION

The term ***total optimization*** is introduced in this paper to describe the process of applying "optimization theory" to the control of ***all*** aspects of the surveying and mapping system. Actually, the qualifying adjective "total" is apparently necessary only when the term "optimization" is used specifically in a surveying context. Unfortunately "optimization" in surveying has, by common usage, come to be understood as a relatively limited process, whereby the design of a survey is optimized with respect to its ultimate accuracy, precision, sensitivity, or reliability. Such "optimization" problems as zero-, first-, second-, or third-order design (terminology after GRAFAREND [1974]) are deliberately limited in scope to consideration of the "quality" of the survey - as embodied in the various covariance matrices - and all concern for the "efficiency" of the survey - in terms of manpower, equipment, time, and costs - is suspended. Recent studies [STOLIKER and ANDERSON, 1981; ANDERSON et al., 1982] have shown that there is considerable scope for the application of optimization theory to the problems of designing, planning, and executing "least cost" surveys; hence the need for the term "***total*** optimization" when both least accuracy ***and*** least cost are sought.

2. SURVEYING AND MAPPING SYSTEMS

At the outset it will be necessary to consider the concept of "a surveying and mapping system", since it is this which determines one of the most difficult aspects of the total optimization problem: formulation of the mathematical model. The term ***system*** is deliberately employed here in the usual engineering sense, to imply an integral body of knowledge, equipment, and methods (a "technology") designed to produce "information" of a specified kind and/or perform "work" in a specified manner. To achieve this goal the system comprises an indefinitely large number of intricately interrelated components, each with a varying degree of "control" available to the user of the system and each manifesting a variable effect on the system output. Despite this complexity, it is possible to depict the surveying and mapping system as a "black box", in the common engineering way, with inputs and outputs, but unspecified contents. Thus, in figure 1, the surveying and mapping system is shown, with the usual input of resources and output of results.

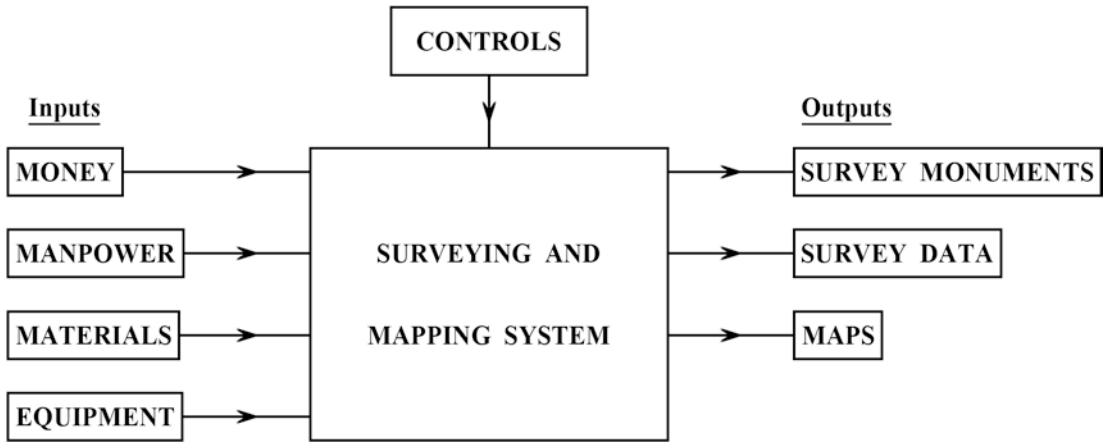


FIGURE 1 : A "BLACK BOX" SURVEYING AND MAPPING SYSTEM

2.1. Scope of the System

For the purposes of the study reported here, it has been assumed that the scope of the surveying and mapping system (i.e. the contents of the black box) should be as broad as possible. The boundaries of a system are almost always blurred because there is usually a certain amount of overlap and interaction with other systems: for example, the surveying system overlaps with the electronic engineering system, the computing system, the economic analysis system, and the business management system, to mention just a few of many. In this study we define the contents of the system in terms of the usual sequence of tasks which have a reasonably direct impact on its inputs and outputs. With this approach, the "processes" set out in table 1 are recognizably within the system.

3. OPTIMIZATION: THE STATE OF THE ART

Defining the state of the art of optimization theory and practice is a necessary prerequisite to an investigation of its applicability to surveying and mapping systems. At the outset it is necessary to try and clarify some terminology and basic definitions (section 3.1). In the following section (3.2) an attempt is made to review, albeit briefly, the existing body of knowledge usually grouped under the heading of "optimization".

3.1 Terminology and Basic Definitions

Optimization: Briefly, the theory of optimization may be defined by stating that: ***"It studies how to describe and attain what is Best, once one knows how to measure and alter what is Good or Bad."*** [BEIGHTLER et

SURVEYING AND MAPPING SYSTEM	
"MAINLINE" PROCESSES	ANCILLARY PROCESSES
Job or Project Requisition Design of Specifications Survey Design & Preanalysis Operations Planning Reconnaissance Observing, Data Testing & Remeasurement Data Adjustment & Analysis Publication & Data Banking	Executive Management (Policy and Planning) Research and Development Marketing Operations Management

Table 1 : PROCESSES WITHIN THE SURVEYING AND MAPPING SYSTEM

al., 1979]. The layman's word "best" may mean either "maximum" or "minimum" depending on the circumstances, but both are mathematically definable. The study of optimization, taken in isolation, may be regarded as a branch of applied mathematics, and thus is seen to have its foundations in various areas of pure mathematics. But, for the purposes of this investigation, it is necessary to see optimization in a broader context as just one building block in a larger structure (see figure 2).

Operations Research and Management Science: As WAGNER [1975, §1.1] has pointed out, "operations research" (the British prefer "operational research") is an unfortunate, undescriptive and ambiguous term which is nevertheless thoroughly entrenched. It means, approximately, the application of scientific principles (based largely on the applied mathematics branches: optimization theory, decision analysis, and game theory - see figure 2) to systems design and analysis, and decision-making problems. The more American term "management science" is almost, but not always, synonymous.

The basic process of operations research encompasses three steps:

(1) Define accurately and quantitatively the interactions of the system variables.

(2) Define a ***single*** measure of system effectiveness in terms of the system variables.

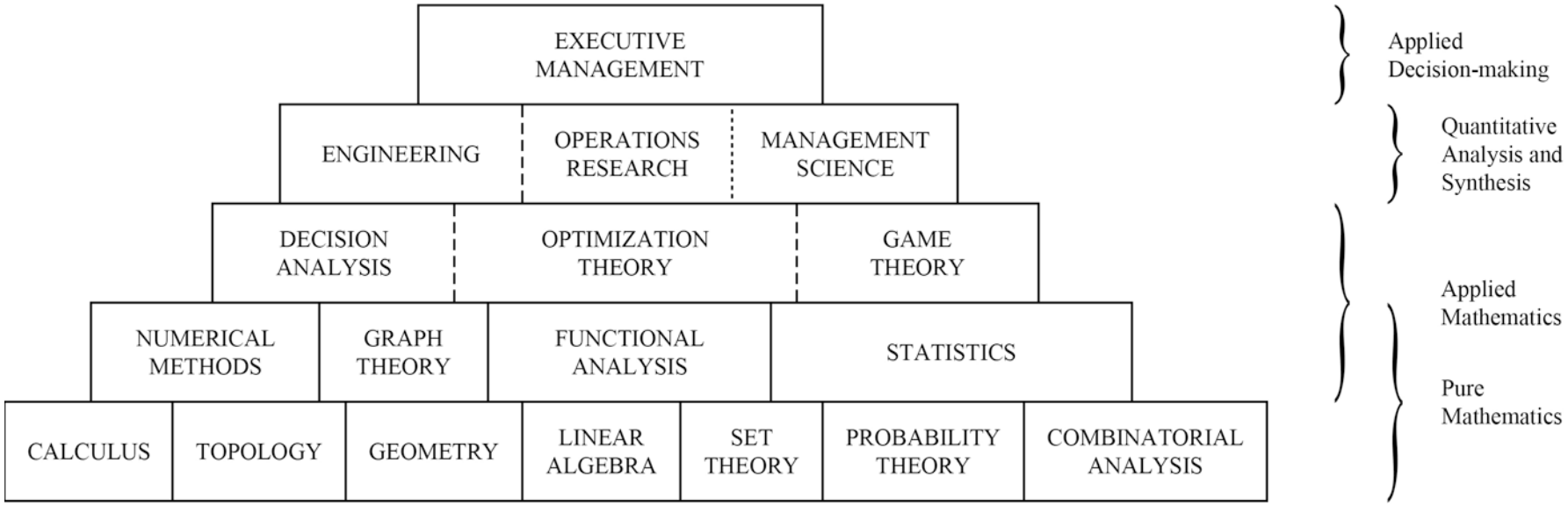


FIGURE 2 : THE STRUCTURAL ENVIRONMENT OF OPERATIONS RESEARCH

(3) Determine values of the system variables which yield optimum effectiveness.

To be more explicit, these stages can be expanded into a number of steps which can best be visualized in flowchart form, as in figure 3.

Problem formulation: Every parameter to be used as a ***variable*** in the mathematical model must be identified and a means of measuring it determined. Whether the values of a parameter can be controlled by the manager or not will determine which are ***decision variables*** - values of which will form the model information output - and which will make up the data input. The nature of the problem and its environment will impose ***constraints*** on the values of variables: these will form the basis of the constraint equations in the model. A single ***effectiveness criterion*** must be chosen so as to provide a means of defining a good or improved solution. This criterion forms the basis of the objective function in the model. Introduction of more than one criterion will usually lead to a conflict situation and thereby greatly complicate the problem. ***Time and space horizons*** must be identified so as to define the limits of the analysis. The consequent model will usually vary for different time horizons.

Model Building: Data input and ***information outputs*** are merely values of the model variables. However, measuring values of the input data may be a major part of the process. Determining the ***interactions of the variables*** and describing them mathematically is often the most difficult aspect of the process and is not always possible of feasible. The ***objective function*** describes mathematically the relationship between the effectiveness criterion and the decision variables. ***Constraint equations*** mathematically express the limits on the variables and their interaction.

Solution: (For a discussion of methods of solution see section 3.2).

Post-Analysis: Sensitivity testing is the study of how the optimal solution is affected by changes in the values of the model variables or by changes in the model itself. This analysis can provide information just as valuable as (or more valuable than) the actual solution. It can, for instance, show how the accuracy (or inaccuracy) of the input data is affecting the optimal solution, and thereby determine which are critical variables in the decision.

Implementation and Model Updating: Implementation the optimal solution can prove to be a major component of the operations research process. Here, human and other factors may intrude and perhaps force a revision of the whole analysis. The necessity for ***model updating*** may also arise from within the problem environment itself, since few problems are truly static.

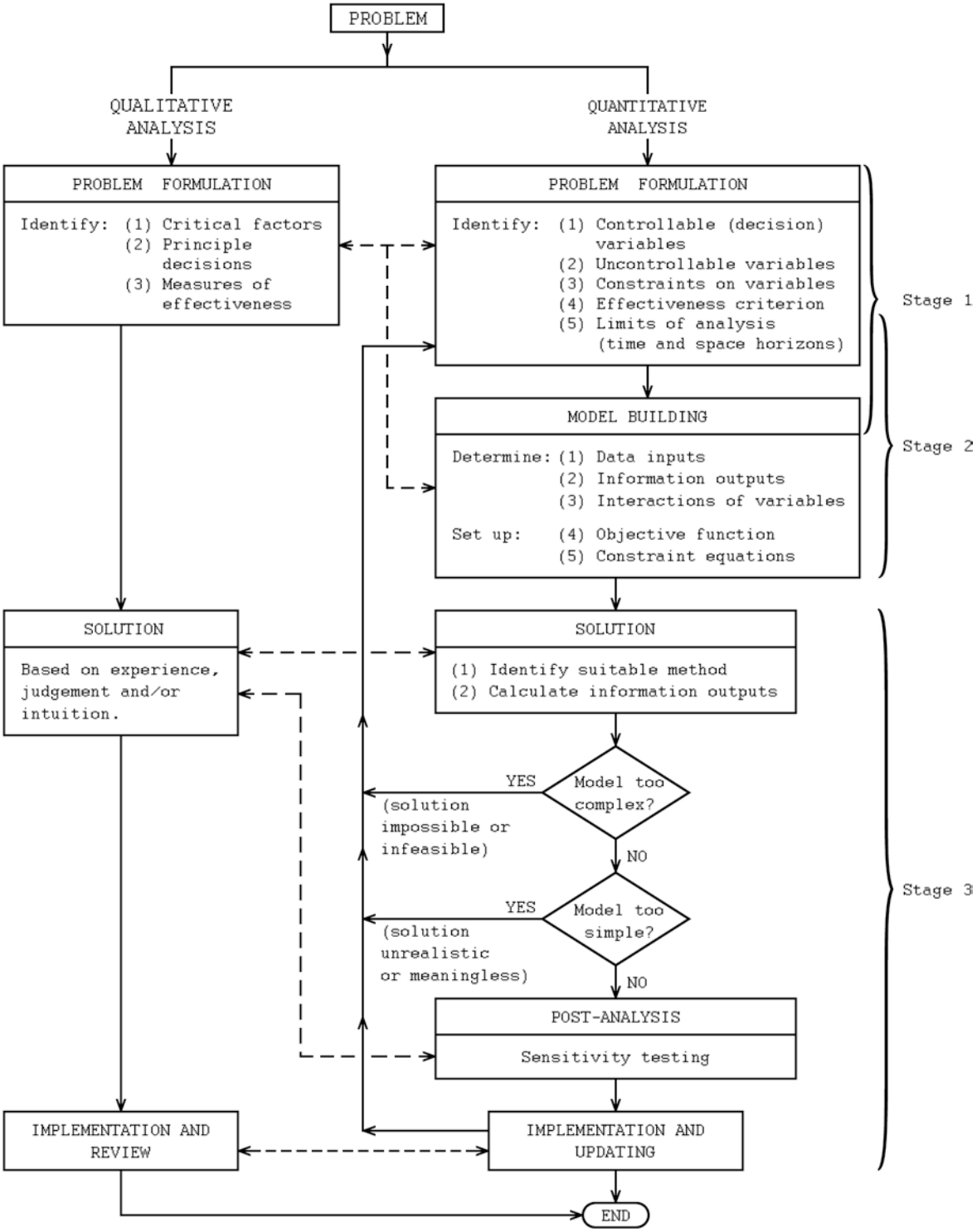


Figure 3 : THE OPERATIONS RESEARCH PROCESS

3.2 Review of Optimization Theory

A substantial literature in the form of both serial publications (technical journals) and textbooks exists in the areas of optimization theory and operations research or management science. For example, LUENBERGER [1969], in a rather specialized text on optimization theory, cites 159 references. And in operations research WAGNER [1975] lists 562 selected references. Clearly any form of generalized literature search would not be a feasible quite apparent to us that the body of literature is sufficiently extensive and diverse to support most specialized studies which might arise from, or be required by, surveying and mapping systems problems.

To the uninitiated, optimization can appear to be a very disjointed, disorganized collection of seemingly unrelated approaches, techniques, and "tricks". This image is partially illusory, and is largely due to the multiplicity and variety of the applications of the theory. For instance numerous quite different problems, drawn from different disciplines, will commonly exhibit similar, or identical, characteristics at the level of the mathematical model and can therefore be solved by a single method. Simplification, therefore, resides in the ability to recognize common underlying characteristics in the problem formulation. This process may be aided by some form of classification of optimization ***problems***. One such scheme is presented below. But there remains the sometimes bewildering array of ***methods***, all lumped under the heading of "optimization". These too can be classified, but this task is not so simple. There is a perplexing degree of interaction and overlap between the identifiable optimization methods and sometimes what appears to be a method (e.g. the branch-and-bound technique) is in reality little more than a basic concept or approach which may underly (often imperceptably) a number of evidently different methods. It may be of some comfort to realize that there is indeed a unifying principle behind almost all optimization methods, though it exists only at the level of mathematical representation. LUENBERGER [1969] states as the primary objective of his text, the unification of a large part of optimization methodology through recognition of a few "***geometric principles of linear vector space theory***". That this should be possible need not be surprising; a few of the fundamental principles are easily appreciated, for instance:

(i) ***The projection theorem***, which visualizes the minimum case of optimization as the shortest distance between a point and a plane

(ii) The ***duality theorem***, which recognizes the existence of a "dual" of the situation in (i), wherein the shortest distance is representable as a maximum under appropriate geometric circumstances. Thus the projection theorem is of equal utility when the optimum sought is a maximum.

(iii) ***The concept of differentials:*** A familiar technique for finding the maximum or minimum of a function (i.e. the optimum) uses the differential calculus. Thus the result of setting the derivative of the objective function to zero is the optimum. This concept pervades and unifies a great deal of optimization theory.

Classification of problems and methods: A basic categorization of optimization problem types has been presented by LUENBERGER [1969]. The categories are as follows:

(i) ***Allocation:*** Such problems are typically concerned with optimal distribution of resources. The resources may include money, 'manpower, equipment, or materials. An example would be the problem of allocating personnel with various experience and abilities to a number of concurrent projects, with the objective of achieving maximum profit or perhaps maximum efficiency or minimum time.

(ii) ***Planning:*** This category includes all those problems concerned with determining an optimal procedure to attain stated objectives. Examples could include: scheduling of activities within a project to achieve minimum project duration, sequencing observations in a doppler strategy or spaces inventory over some time horizon to maximize profit.

(iii) ***Control or Guidance:*** Usually refers to problems of determining the inputs to a dynamic system (i.e. one changing in time) to optimize its behaviour. "Filtering" techniques, such as those used in an inertial survey system, exemplify this category. However, the "system" may be the human resources involved in a project where personal factors cause continuous time changes.

(iv) ***Approximation:*** These problems involve the approximation of a known mathematical entity such as a function by some usually simpler form. For example, fourier representation of a specified regular waveform, or polynomial approximation of a given function would fit in this category. The objective normally is to minimize the error in the representation.

(v) ***Estimation:*** These are the problems, familiar to a surveyor, of estimating some quantity by means of imperfect observations. The objective, as in approximation, is to minimize the error in the adopted estimate. Least squares estimation (or "adjustment") is only one case within this category.

(vi) ***Games:*** This is the term often applied to problems involving a competitive element. The competitive element is normally represented in the objective function so that its value is jointly dependent on two variables with an antagonistic relationship. Thus the objective may be to simultaneously minimize cost and maximize accuracy of some observations; both cannot be fully realized concurrently, but an optimum comprise can be determined according to some criterion. Examples of problems amenable to games theory techniques often arise when two problems from the categories above are intermixed in a conflicting manner. For instance, an allocation-of-personnel problem may conflict with a scheduling-of-project-activities problem from the planning category. Sometimes the two problems may come from the same category but conflict can still arise; for example, simultaneous allocation of funds and personnel.

It is not easy to classify the available optimization methods. No matter what scheme is chosen, considerable overlap and ambiguity seems to arise. However, it is possible to simply list the more important methods, though it should be understood that within the term used for a "method" may reside a whole class of techniques with some unifying property (which may not be particularly relevant to the types of problems for which the method is suitable). A non-exhaustive list of methods and some associated algorithms follows (e.g. see WAGNER, [1975]):

(i) Linear programming
- . simplex algorithm
- . graphical algorithm
- . ellipsoidal algorithm

(ii) Quadratic programming
- . quadratic simplex algorithm

(iii) Network modelling
- . critical path methods (CPM/PERT)
- . decision tree algorithms
- . graph theory

. travelling salesman algorithm
. branch-and-bound technique

(iv) Dynamic programming
. recursive algorithm
. backwards induction algorithm

(v) Integer programming
. branch-and-bound algorithm
. combinatorial modelling
. travelling salesman algorithm
. zero-one technique

(vi) Non-linear programming
. trial-and-error techniques
- adaptive search
- Fibonacci search
. minimax methods

(vii) Stochastic modelling
. Bayesian inference
. probabilistic dynamic programming
. queuing theory
. simulation techniques

The question inevitably arises as to how to choose the best method to solve a particular problem. There is no straightforward answer. Occasionally considerable experimentation may be necessary to discover which of a number of available methods will most efficiently produce a solution. And relatively minor changes in the problem may alter such a decision. Much of this complexity arises because ***there is no one-to-one mapping of the set of problems onto the set of methods or vice-versa***. This situation is illustrated by figure 4, where just a few of the matchings of methods to problems and problems to methods are shown. Note that any one method may be suitable for solving different categories of problems and a particular problem may be solvable by more than one method.

3.3 Some Applications in Surveying and Mapping

It seems that most of optimization theory, and particularly operations research techniques, have hardly touched surveying engineering. This is in strong contrast with the other branches of engineering. Of course, surveyors have extensively employed least squares estimation methods to

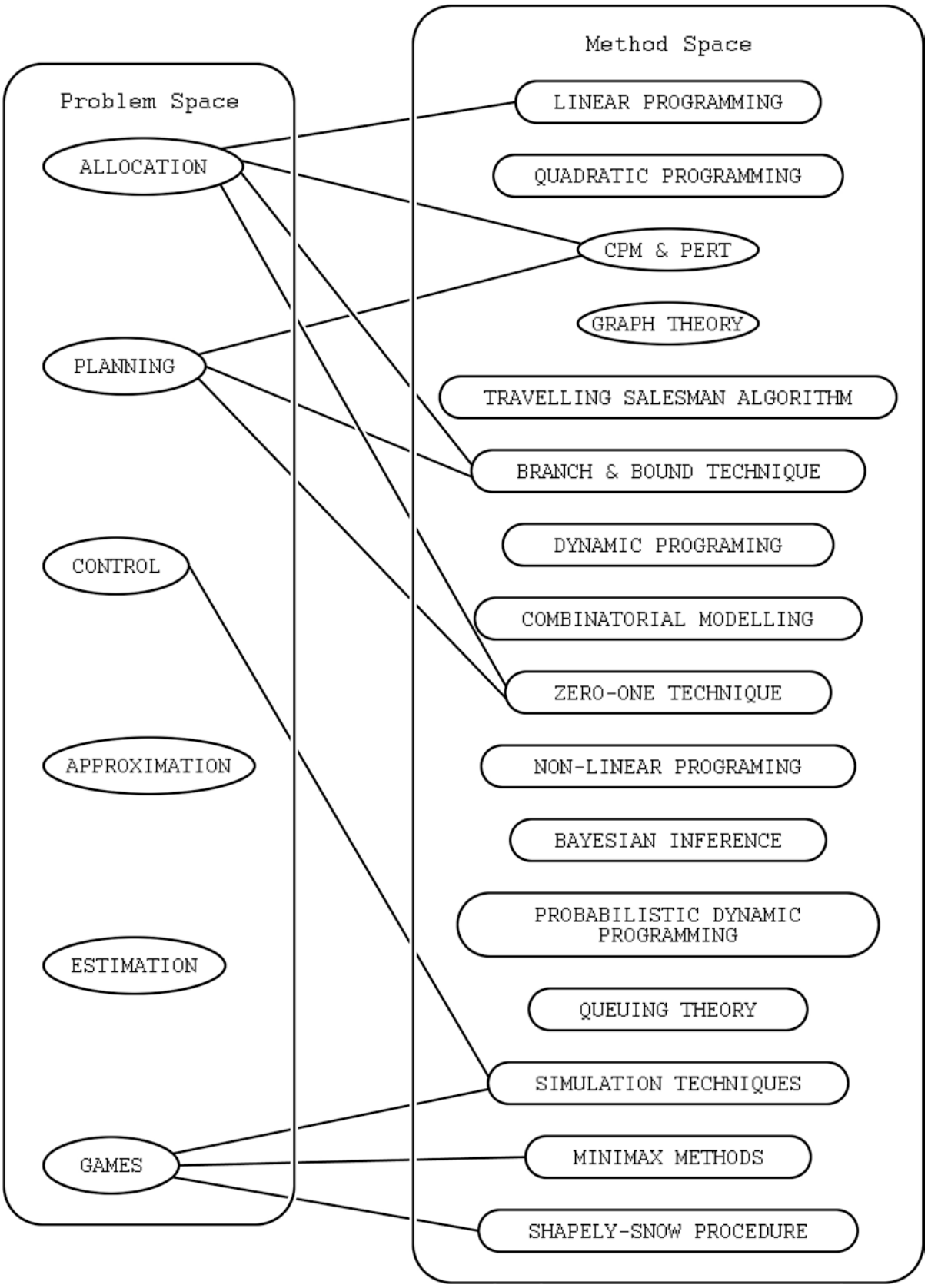


Figure 4 : MATCHING OPTIMIZATION METHODS TO PROBLEMS

optimize accuracy - indeed to such an extent that the process of least squares preanalysis has become totally synonymous with "optimization" in surveyors vocabulary - but we are more concerned here with the optimization of surveying and mapping costs and logistics. If we preclude this conventional application to the first and second order network design problems, the remaining applications of optimization theory to other aspects of surveying and mapping systems are so few that it is feasible to enumerate them in a few lines. Some examples are as follows:

(i) Applications of the critical path method to planning of map production and field work operations [LYONS, 1975].

(ii) Determination by linear programming, of a least-cost shipment plan for resources in planning geodetic field work [BAZHANOV, 1963].

(iii) Determination and correction of incorrect observations in field work by Bayesian Inference [BOSSLER, 1972].

(iv) An attempt to determine the optimum design of a levelling network using linear programming [CROSS and THAPA, 1979].

(v) Optimization of the number of observations required in a horizontal control network using non-linear programming [MILBERT, 1979].

(vi) An example of the application of Graph Theory to the observation of geodetic networks [TSOUROS, 1980].

Of these (iv) and (v) are, strictly, concerned with aspects of the second order design problem and accuracy rather than cost is the key objective parameter. It may be fairly concluded that applications in surveying engineering are, relative to other areas, at a very early stage of development.

4. TOTAL OPTIMIZATION: A PRACTICAL APPROACH

In this section we treat the dual task of optimization the accuracy and the logistics. To illustrate the procedure involved we have taken the specific problem of establishing a horizontal network in which (a) the accuracy of the coordinates are examined and (b) the logistics in collecting the observations in the field is determined.

In section 4.1 we give the basic concepts behind the engineering approach adopted, called the sequential method, and compare it to the simultaneous method - the method of the mathematical purist. A practical example is presented in section 4.2.

4.1 The Sequential Method

Basically, there are two approaches to optimizing accuracy and logistics - the simultaneous method and the sequential method. In the simultaneous method, the objective function is written explicitly and simultaneously in terms of accuracy and logistics. The solution yields an optimum answer for both sets of quantities but because of the complex interrelationship between these two rather non-heterogeneous sets of quantities, it may not be possible to relate the solution to the real world.

Because of these difficulties a more pragmatic engineering approach has been adopted called the sequential method. As the name implies, the operation is performed sequentially. First the accuracy is optimized (maximized) and then after certain information is passed onto the logistics stage, the logistics, i.e. travel time to collect the observations, is minimized. If a problem is encountered at the logistics stage, then it may be necessary to return to the accuracy stage and modify the design as illustrated in figure 5.

4.2 Practical Example

In this example a simple trilateration network is checked for accuracy and logistics. The network is shown in figure 6 and consists of four new stations to be established off two existing fixed stations. The 95% confidence region of the new stations is not to exceed 0.15 metres and it is planned to use trilateration as the measuring technique. The accuracy analysis of the network showed that using an E.D.M. with an accuracy rating of 5 mm + 5 ppm the largest semi-major axis at 95% was 0.130 metres, well within specifications.

The logistics of observing the distances in the network has been classified as a type-2 problem [STOLIKER and ANDERSON, 1981], in that two stations must be occupied simultaneously to measure the distances. An algorithm to find the least-cost ordering of observations in such a situation has been developed and implemented at The University of Calgary in a programme called "SURF" [ANDERSON et al., 1982]. The least-cost ordering problem is solved by SURF using a zero-one programming method based on the branch-and-bound technique, but with only feasible solutions evaluated to improve efficiency. The model and solution is set up as follows:

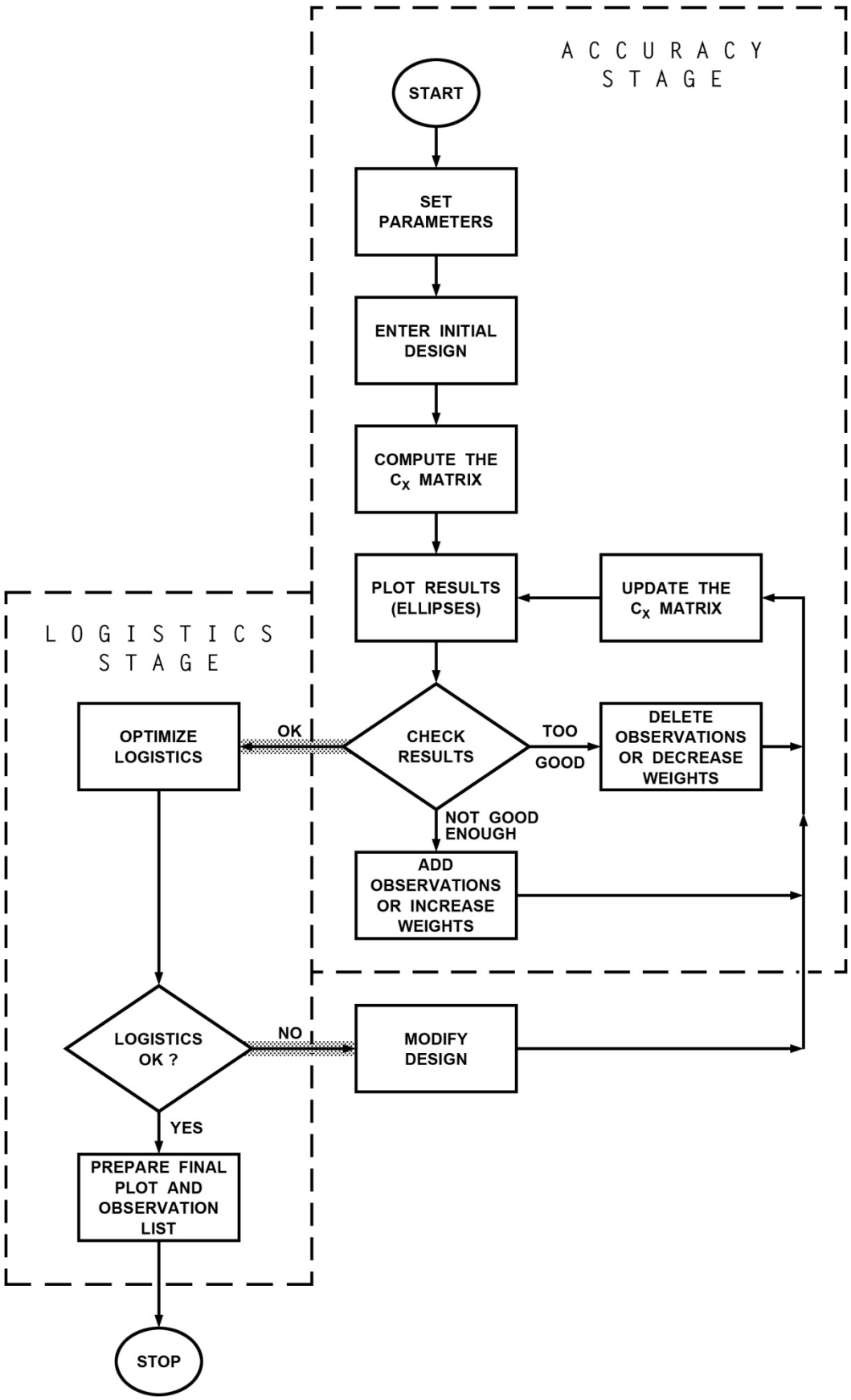


Figure 5 : SEQUENTIAL METHOD - INTERPLAY OF ACCURACY AND LOGISTICS STAGES

Consider a network of 'n' stations requiring simultaneous occupation of 'nobs', $1 \leq nobs \leq \binom{n}{2}$, pairs of stations; for example observations of distances, simultaneous azimuths, or doppler translocations.

Define U = (-nobs, -(nobs-1), ..., -1,0,1,2,3, ..., nobs)
W = (-nobs, -(nobs-1), ..., -1,1,2,3, ..., nobs)
Z = (1,2,3, ..., nobs)

Given
- a set of states $(t_i : i \in U)$,
- a symmetric matrix $(C)_{ij}$ describing the cost of transition from t_i to t_j $(i,j : i \in W \text{ and } j \in W)$,
- a vector $(S)_i$ describing the cost of transition from t_0 to t_j $(j : j \in W)$

Define the least cost ordering of states as $(h_0,h, ..., h_{nobs}, h_0)$.

Choose h_k such that:

(1) $h_0 = t_0$

(2) $\forall\ k \in Z\ \exists j \in W$ such that $h_k = t_i$

(3) $h_k = t_i \Rightarrow \sim [\exists j$ (different from k) $\in Z$ such that $h_j = t_i$ or $h_j = t_{-i}]$

Analysis of the logistics of the network in figure 6 using SURF showed that a minimum of 92 km of travel would be required to make the observations (neglecting travel to the initial setup and back from the final setup). When the distance between stations 1 and 3 was removed the largest semi-major axis at 95% was 0.131 metres, an insignificant change of 1 mm. The logistics analysis of this network showed the minimum internal travel to be 80 km, a significant decrease of 13% in travel time alone. Further savings would also result from the decrease in the number of distances to be measured and the elimination of the possible necessity of renting a long range E.D.M. to measure the 26 km distance between stations 1 and 3.

5. CONCLUSIONS

A few critical problem areas have become more clearly discernable through the results of this study. These problems should be treated as a basis for further work. They include:

(1) As prerequisite to improved modelling, the "contents" of the surveying and mapping system must be defined in greater detail (see table 1). As it is not clear at this stage which of the processes within the system are more important with respect to potential for savings due to improvements in efficiency, it may be necessary to assess priorities before proceeding further. In

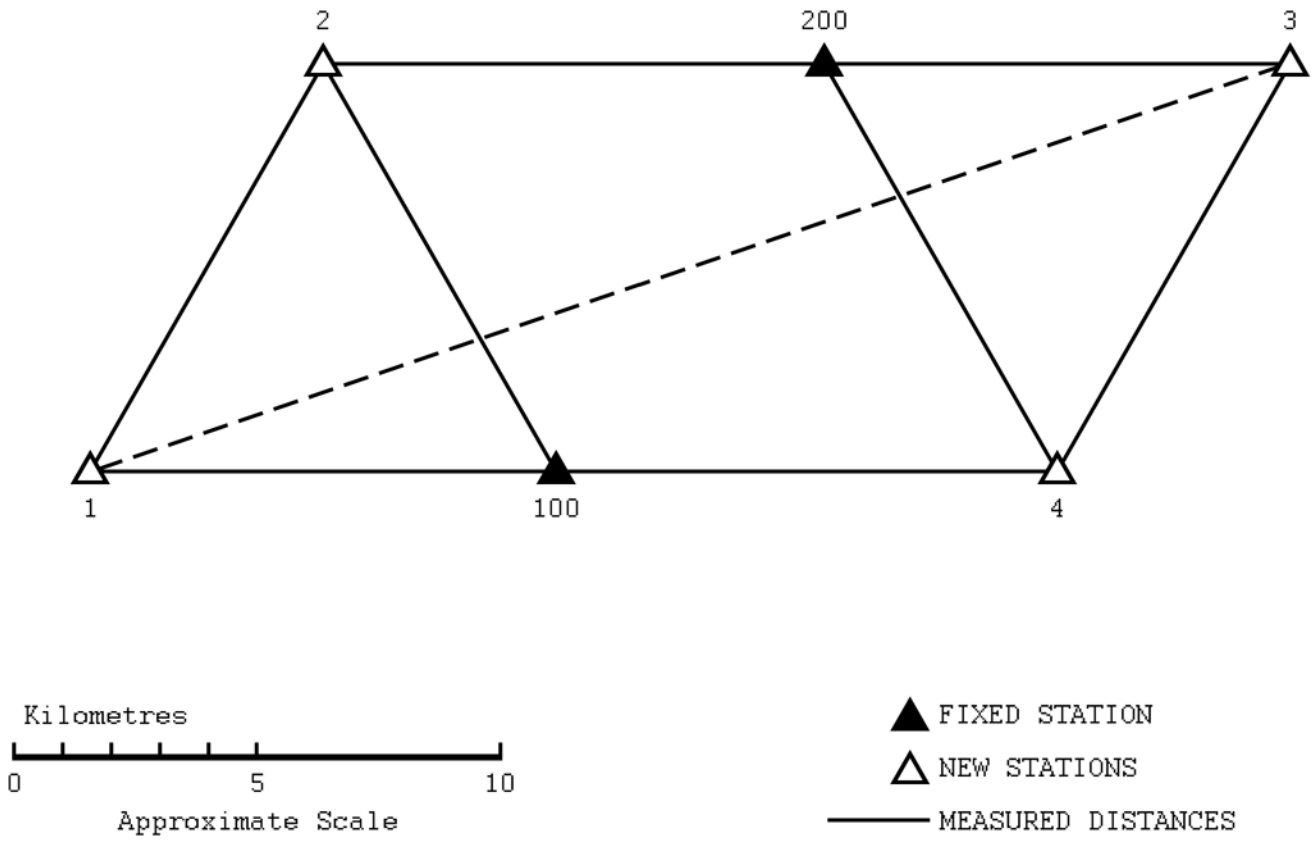


Figure 6 : TRILATERATION NETWORK FOR EXAMPLE IN § 4.2

any case, the key requirement is a preliminary "problem formulation" stage aimed specifically at the system processes listed in table 1; the questions being: (a) What are the variables? (b) Which are "controllable"? (c) How can they be quantified? (d) What are the constraints? (e) How do the variables interact? - particularly with regard to "quality" and "efficiency"? Most of the information required to answer these questions must come from the "real world" surveying and mapping system itself, where it is currently locked up in the qualitative analysis side of the system (see figure 3).

(2) Once the components of the system are defined in terms of specific optimization problems of significance, the basic problem of "model building" can be tackled. The key commodity required here is a combination of ***experience*** and ***knowledge***, relating to both the surveying and mapping system and operations research. The full range of operations research techniques should be explored.

(3) Concomitant with (2) is the unresolved difficulty - so far encountered on a few occasions in survey network design examples - of obtaining physically infeasible solutions to combined quality and efficiency problems, which are, nevertheless, mathematically valid. Such instances suggest an inadequacy in the modelling of the ***constraints*** rather than an inappropriate objective function.

(4) The stock of tools (i.e. computer routines) must be built and expanded, and, where necessary, modified to suit the specifications called for by surveying and mapping system problems. The software repertoire - in terms of installed routines capable of handling the types of requisite models in surveying and mapping - is meagre indeed.

(5) The problem of CPU efficiency of some computer routines may become limiting. Prediction of software performance in particular cases will always be complicated and prone to error, but there is considerable scope for investigations into the means of improving software efficiency for longer and more realistic examples. This is already a major concern in the area of software development for "type-2" problems.

It is recommended that the challenge of total optimization be met by efforts on two fronts:

(a) The problems of total optimization based on ***simultaneous*** models of precision and cost should be further investigated. An approach based on the principles of game theory may be beneficial. The means of

introduction adequate constraints to mitigate the effects of meaningless solutions requires further study.

(b) As a practical alternative, the ***sequential*** method is known to be reliable, even though it may not produce a perfectly optimum solution. However, this methodology has not yet been fully implemented; only the accuracy stage has been realized at The University of Calgary in a convenient package as programme CANDESN [MEPHAM and KRAKIWSKY, 1981]. Development and implementation of the "logistics", of efficiency, stage of this package should be given high priority. An outline of the package architecture, as currently envisaged, is shown in figure 7. In the first phase a travelling salesman algorithm, the type-2 problem routine SURF, some other basic network modelling routines such as a shortest-route algorithm, and a linear programming simplex algorithm are being included in the efficiency package.

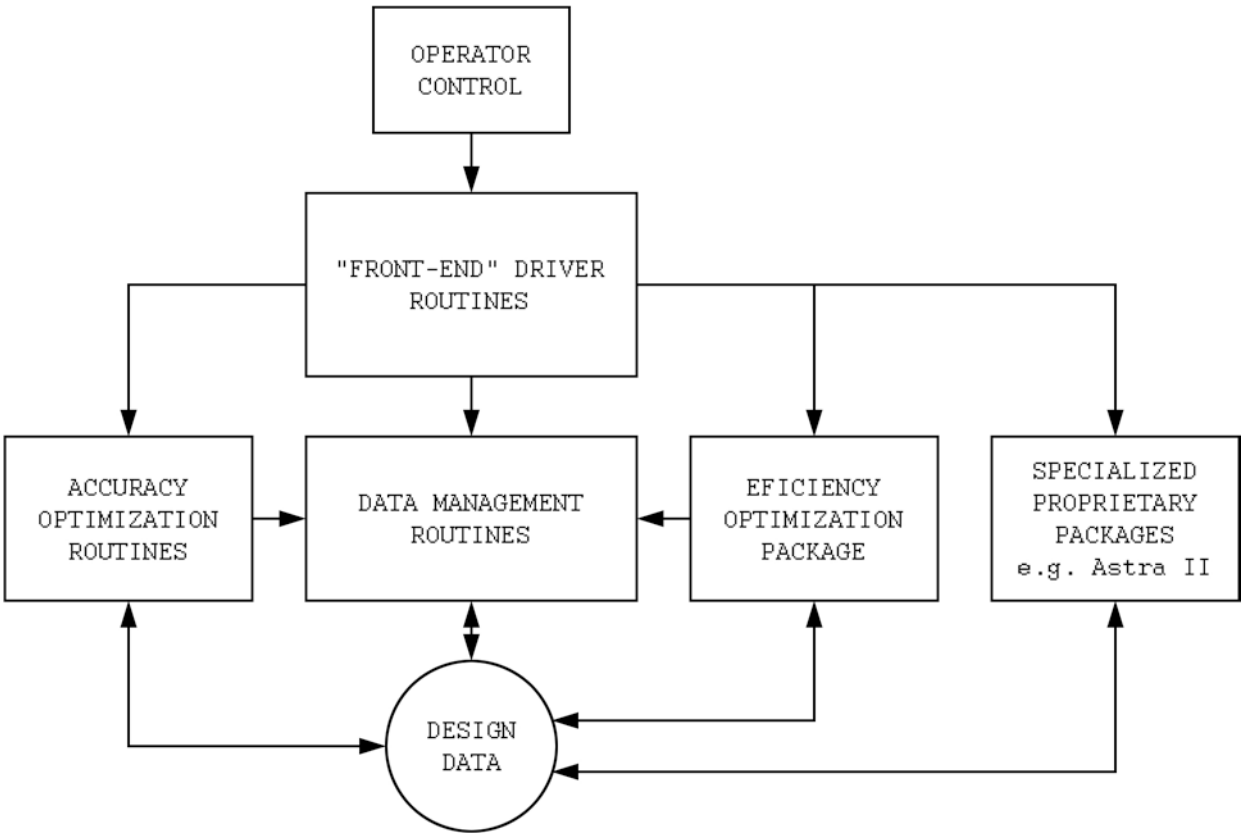


Figure 7 : OUTLINE OF EXTENDED CANDESN ARCHITECTURE

6. ACKNOWLEDGEMENTS

The author would like to gratefully acknowledge the input to section 4 by Professor Edward Krakiwsky. Computer programme CANDESN is being developed by Mr. M. Mepham, who also provided assistance with the example in § 4.2. Mr. P. Stoliker wrote programme SURF. Research work on total optimization has been partially sponsored by SHELTECH CANADA and the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

ANDERSON, E.G., P.C. STOLIKER, and E.J. KRAKIWSKY, 1982: Optimization of Surveying and Mapping Systems. Unpubl. Contract Research Report to SHELTECH CANADA; March 1982: 144 pp.

BAZHANOV, K.V., 1963: The Use of Linear Programming Methods on the Organization of Geodetic Field Work. Geodesy, Mapping & Photogrammetry, 5 (3): pp. 169-173.

BEIGHTLER, C.S., D.T. PHILLIPS, and D.J. WILDE, 1979: Foundations of Optimization, 2nd ed. ISBN: 0-13-330332, Prentice Hall Inc., Englewood Cliffs, N.J., 07632, U.S.A.

BOSSLER, JOHN D., 1972: Bayesian Inference in Geodesy, Ph.D. thesis, The Ohio State University.

CROSS, P.A. and K. THAPA, 1979: The Optimal Design of Levelling Networks. Survey Rev. XXV (192): pp. 68-79.

GRAFAREND, E.W., 1974: Optimization of Geodetic Networks. The Canadian Surveyor, 28 (5) (December 1974): pp. 716-723.

LUENBERGER, DAVID G., 1969: Optimization by Vector Space Methods. (SBN 471 55359x) John Wiley & Sons Inc., New York.

LYONS, K.J., 1975: CPM - A Management Tool for the Surveying and Mapping Industry. The Australian Surveyor, 27 (4): pp. 180-211.

MEPHAM, M. and E.J. KRAKIWSKY, 1981: Interactive Network Design and Analysis. Proc. Sympos. on Geodetic Networks and Computations, (VIth Int. Sympos. on Geodetic Computations), Munich, August 31 - September 5, 1981.

MILBERT, O.G., 1979: Optimization of Horizontal Control Networks by Nonlinear Programming. NOAA Tech. Rep. NOS79NGS12, U.S. Dept. Commerce, Wash. D.C., U.S.A.

STOLIKER, P.C. and E.G. ANDERSON, 1981: Optimization of Observing Logistics in Geodetic Networks. Proc. Sympos. on Geodetic Networks and Computations, (VIth Int. Sympos. on Geodetic Computations), Munich, August 31 - September 5, 1981.

TSOUROS, C., 1980: An example of Application of Graph Theory to Geodetic Networks. Quarternions Geodaesiae (Τεωδαιτικα Τετραδια) 2/1980: pp. 79-84.

WAGNER, H.M., 1975: Principles of Operations Research. 2nd ed. ISBN: 0-13-709592-9, Prentice Hall Inc., Englewood Cliffs, N.J., 07632, U.S.A.

COMPUTATION AND DATA-MANAGEMENT OF BLOCKS OF MEDIUM SIZE GEODETIC NETWORKS

Olaf ANDERSSON
Karsten ENGSAGER
Willy L. WENG

København, Danmark

ABSTRACT

The ongoing revision of the Danish densification network using EDM and EDP is the topic of this paper. Described are:

- the entry and management facilities for geodetic data,
- the adjustment of blocks of 2000 - 3000 stations using in-house developed software,
- the computer-assisted network analysis producing lists of measurements wanted from the field parties.

Contents:

I. Introduction
II. Data-management
  1. Registers
  2. Coordinateregisters
  3. Observationregisters
  4. Auxiliary programs
  5. Data-entry
III. Adjustment
IV. Network-analysis
V. Conclusions
VI. References
Appendix A, B & C

I. Introduction.

The background for this paper is the ongoing revision of the densification network of the Danish Geodetic Institute (the network will be used as control for further densification by local surveyors). The revision include additional measurements, readjustment and editing of recovery and control diagrams (see appendix A). For administrative reasons this revision is done in blocks of 2500-3500 square kilometers. The results are published yearly. At present half of the work is done and the revision is scheduled to be completed in 1989 (see fig. 1). The techniques described here have been in use without major changes since 1976. As they are developed in-house it has been possible to modify and include new facilities in the software when needed. The purpose of the revision is a complete readjustment of the network using both existing and new observations and utilising EDM and EDP. The principles of the revision are:

- only 1. order fixed (spacing 40 - 50 km)
- station spacing 2 km
- homogeneous quality
- all taped traverses (outside town areas) are remeasured using EDM

The spacing and the block size aimed at results in adjustments of 2000 - 3000 stations, and as 4 - 6 of these blocks are being processed in parallel it has been necessary to develop software to handle the data. This software interfaces with the adjustment system. Data entry, adjustment and maintenance of files are mainly done by field and office personal under supervision of the data-manager. It is as important to have a good and reliable data-management system as to be able to solve the normal equations of the adjustment. In the following we will treat:

- the database facilities, auxiliary programs and data entry procedures
- the adjustment
- the present primitive network analysis tools.

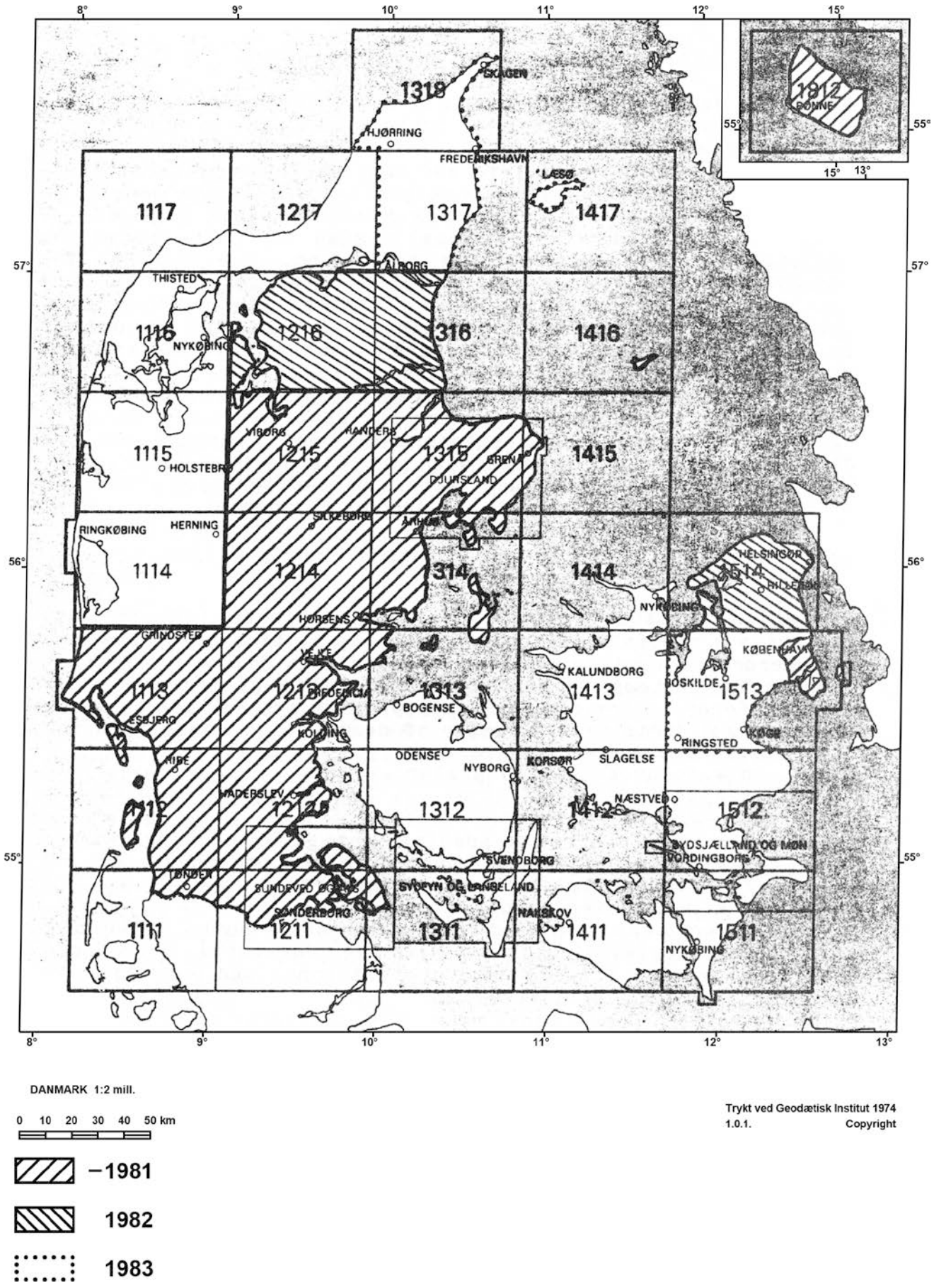


Figure 1: Status of the Revision

II. Data-management.

II.1 The Registers.

The database consists of separate binary indexed sequential files with the station number as common key. The manufacturers isq-software (ref. (3)) is used. The files are divided in two categories, station registers and observation registers (described in section II.2 and II.3 respectively). Both types have in principle only one key, but using scanning other keys may be simulated. This is feasible as a scan of the observation registers at the present size takes app. one minute. The observation registers contains 73500 sets/distances and the station register contains 38000 stations with 85000 pair of coordinates in 12 different coordinate systems. The adjustment software will access these registers directly, which makes input to an adjustment reliable and fast.

Of less importance here is the third category of registers, the description of how to recover the stations.

The above mentioned registers are not a truely unified database, but are accessed through a multitude of user programs. The possible superstructure with cross-references, unified access etc. does not seem to be needed at present.

II.2 Station registers.

A station register (also called a coordinate register) is basically a sequence of records with the station number as key. Each record may contain a number of coordinate pairs in different coordinate systems (national system, UTM, geographic etc.), cross-references for the dual point-number system (see appendix B), survey district and a number of description codes of which the following are used at present:

p = monumented with pillar
t = lost
u = not maintained
s = to be plotted on special large scale network diagram
n = unaccessible intersection point (chimney etc.)
f = fundamental station (first order)
r = restricted point.

The concept has been used for 8 years and a number of entry, editing and retrieval functions have been developed. Coordinate entry can be done in two modes: unconditional and conditional. Unconditional mode will enter both new coordinates and change existing values. Conditional mode will only enter new values. In both cases differences between existing and (attempted) values will be reported. Editing of auxiliary information is done by retrieval/clear, edit, re-enter. In retrieval specifications station-names, areas (by coordinates of boundary polygon or coordinate-free by survey district) and description codes may be used. Some of the output formats are suitable for publication.

Due to the important fact that a versatile coordinate transformation system is available the retrieved coordinates may be delivered in virtually any reasonable system.

II.3 Observation-registers.

The observation-registers are (except for various datamatic information) sequences of generalised sets. Each register only contains one type of observations:

- geometric geodetic
- unblock
- monocomparator
- stereocomparator.

As only geometric geodetic observations (i.e. directions and distances) are used in the present densification network only that type will be treated here.

With our indexed sequential system a few extremely long records will ruin the effective backing storage usage. This is avoided by having these long records (less than 0.5 per thousand) stored in a special register. This is, however kept transparent to the user by a link between the two registers.

Retrieval from the registers can be done by key or by area (the location is known using a coordinate register). In the former mode (which is used by the adjustment system) the sub-key may be used: station name (of station occupied), kind of observation (direction or distance) and set reference number (which gives a connection to the original field notes). Editing in the registers is done by retrieve/clear, edit, re-enter. Some review-functions are available, e.g. the connections within the network of one or several points or the observations within a specified time-interval.

II.4 Auxiliary programs.

A central concept in the adjustment system of the Geodetic Institute is the jobdefinition, which specifies the stations to be fixed and those to be adjusted (see ref. (4)). An auxiliary program compiles a job definition from various input lists (e.g. jobdefinition-, coordinate-, observation-lists etc.). The program allows a flexible user specification of stations to be fixed, new or deleted. The resulting jobdefinition is sorted, which makes later manual search much easier.

Other programs will compile cross-references for the dual number system, make surveys of the register content, perform sum-check of the records in the register etc. The consistency of observations (either new being ready for entry to the registers or a register itself) and a coordinate register is checked by another program, which locates many of the defects before the more sensitive blunder detection of the adjustment system.

II.5 Data entry.

The data entry is done post season by the field parties using an interactive program running on the RC8000 of the Geodetic Institute (see ref. (5)). Both preliminary coordinates, cross-references and observations are entered. The consistency of these data provides the first rough check. The entry procedure consist of the following steps (see fig. 2):

1. Making the manuscript ready. As the field-notes are designed for easy data entry it is only checked that all elements are available.

2. All values of each set are added on a calculator with hard-copy log, which is used in the error location in next step.

3. Online key-punch, including check sum. This technique eliminates virtually all entry errors.

4. The coordinates and cross-references are entered to the register.

5. The consistency of observations and coordinates are checked using the program mentioned in II.4. To remove all detected errors, corrections in both coordinates, cross-references and observations may be necessary.

6. Observations are entered to the register.

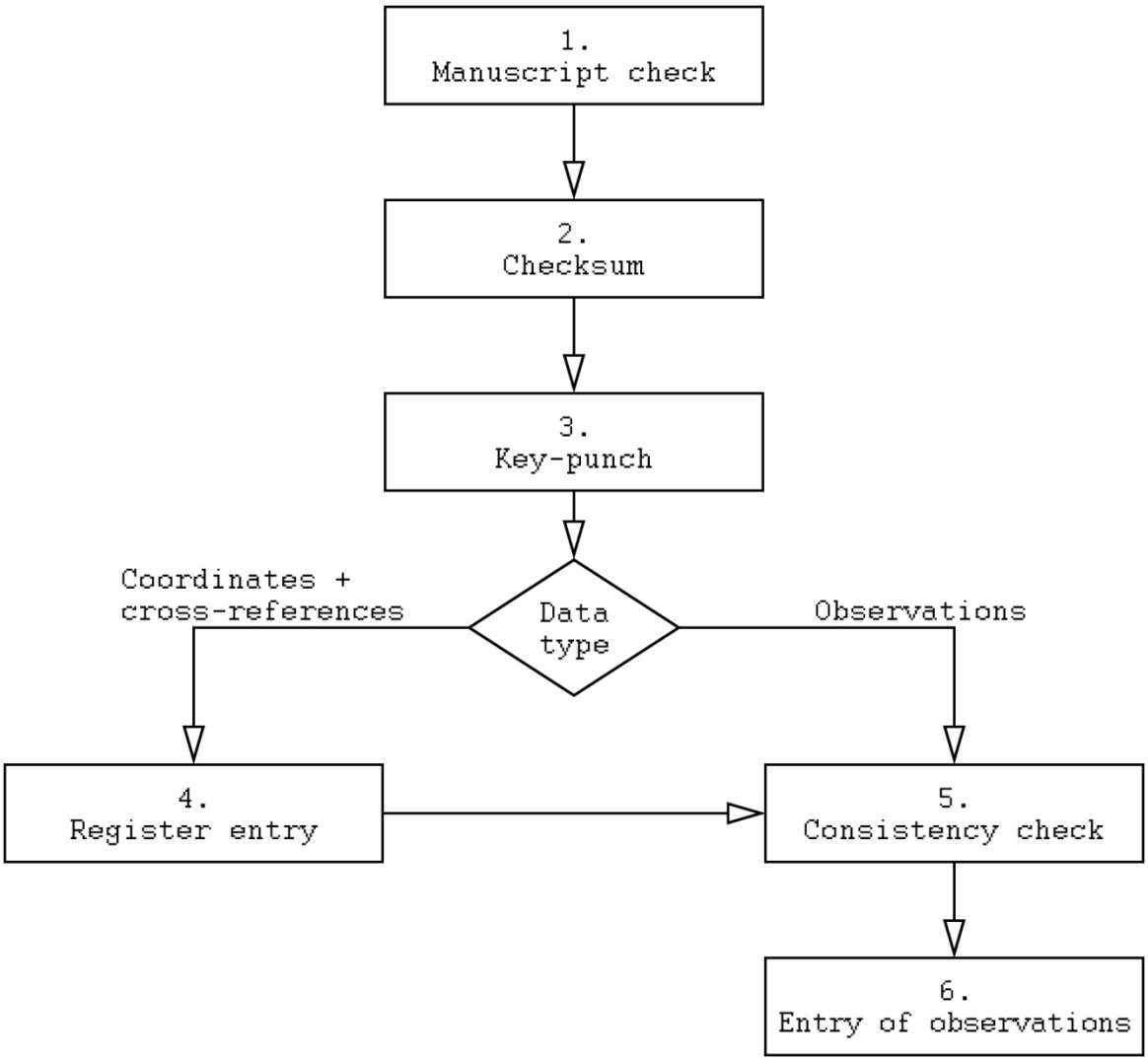


Figure 2: Data entry.

III. The adjustment.

Along with the algebra of least squares an adjustment contains a number of steps which we will try to list (see fig. 3). Basically the same sequence of steps are used in

- The preliminary adjustment prior to the field work, using existing observations.
- The verification adjustment after field work.
- The final adjustment.

The adjustment is central in the revision. The other steps are listed in appendix A. In appendix C some typical times for compilation of jobdefinition and for the adjustment are shown. The least squares program is described in ref (1).

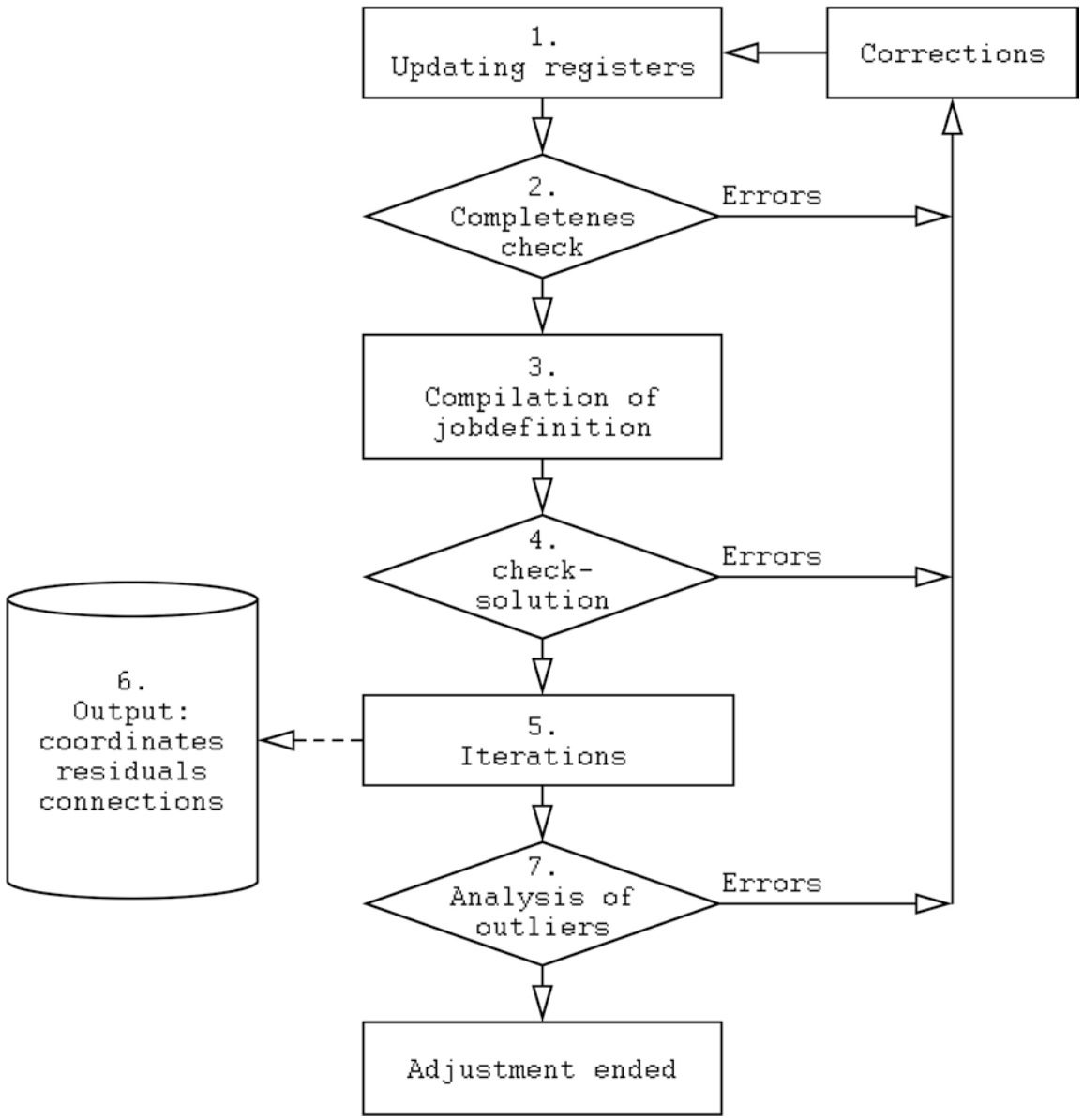


Figure 3: The adjustment.

1. The registers are brought up to date. All observations relevant to the readjustment are entered. This is in principle already done due to a nation-wide entry.

2. Completeness of coordinate registers are checked. Using survey districts all stations in the block are retrieved and those without coordinates are inspected and proper action taken, e.g.

- missing coordinates are entered,
- the point is u-coded, i.e. it will not be used,
- the point is added to the (machine readable) list of stations to be surveyed.

3. Compilation of jobdefinition. A number of station lists are created: (a) Using polygon-boundaries two areas are defined, one covering exactly the area under revision and another 10 kilometer wider in all directions. In both cases u-coded points are skipped. (b) To be used in the adjustments after field work a list of both readjusted and first order stations in the vicinity of the revision area is compiled.

These lists are used to produced the actual jobdefinitions. In the preliminary adjustment all external and first order stations are fixed. In the final adjustment first order and all previously readjusted are fixed and stations in the area itself and in the rest of the vicinity are adjustable. These compilations are done using the program mentioned in II.4.

4. Execution of the least squares program. The program reads the jobdefinition and access coordinate and observations registers and retrieve relevant information. Various defects may show up:

- missing coordinates. The program can not generate preliminary coordinates, but will transform preliminary coordinates to the correct coordinate system. Transformation of fixed values are not allowed.
- sets of un-oriented directions only gives information to the adjustment if at least two of the targets are used in the adjustment. When accessing the observation register the adjustment program is able to list those sets with only one target in use. It is checked manually whether any of the other targets in the set ought to be included in the jobdefinition.
- the re-ordering will pin-point stations without observations.
- after first solution singular stations will be listed. Although the adjustment program is able to continue the iterations the process is generally stopped.

The defects are resolved by e.g.

- u-coding of stations,
- adding stations to the list of stations to be surveyed,
- enter any additional observations that may exist,
- modify the jobdefinition.

If any correction occurred the adjustment continues at step 3 or 4, otherwise at step 5.

5. The iterations are continued until convergence, which is defined as a computational noise 1000 times smaller than the observation noise. In each pass both observation- and normal-equations are recomputed and the normals are solved by Cholesky.

6. The output from the adjustment program consist of (a) lists of fixed and adjusted coordinates, (b) used observations,

computed residuals and detected blunders (see step 7 below), (c) the so-called connection file to be used in the net analysis (see section IV).

7. The outliers located by the least-squares program are checked using:

- earlier computations
- connections
- the adjusted coordinates
- original field notes
- local adjustments
- various maps
- to a large extent common sense !

The detection of blunders uses an idea by T. Karup. In each pass the weight of observations, the residuals of which exceeds a user defined limit (say 3 times the standard error) are dynamically decreased as a function of the magnitude of the residual. This simple technique has proven to work and in most cases the masking effect is handled. It is highly important in adjustments of this size that some sort of blunder detection exist.

The first adjustment usually contains 1-2% outliers. Except approximately one per thousand very small "outliers" which are kept, the outliers are resolved or removed.

IV. Network-analysis.

The analysis of the network (see fig. 4) is done twice. Once prior to the field work and once after (as the result of a campaign is never what was expected, a simulation study is not enough). The analysis is at present manual, but based on four computer produced elements:

1. Displacement vectors, see fig. 5. The adjusted coordinates are compared with old values. Also without new measurements things may show up as most of the old coordinates are hand-computed in many small local densifications.
2. Reference and witness diagrams, see fig. 6. In order to assure that the network diagram isn't too crowed and that details can be perceived, locally connected groups of stations (say less than 400 m apart) are plotted on individual diagrams at large scale. This is in most cases a real station and a number of reference, recovery or witness points. The station with most non-local connections is called the main station and the non-local connection of this station is not plotted on the reference diagram, but on the network diagram.
3. Network diagram, see fig. 7. The entire network except intersection points and reference points are plotted with all connections shown. (A reference point is in this context defined as a point shown on a reference diagram, but not being a main-station).
4. Table of intersection points, see fig. 8. The table lists all directions to each station. Those with less than a user defined limit (say 5) are listed separately

and it is checked that the geometry is strong.

The four elements are checked manually. On a list of all stations in the block the decisions concerning each individual station are compiled. In most cases it is acceptance. Stations not checked in this way are then considered. Based on the analysis and the file of points to be surveyed (see III.2) the list of required measurements is made in co-operation with the field parties.

If the analysis does not indicate a need for additional measurements, the coordinates are ready for publication. Otherwise the process continues at step III.1 when the necessary observations are available.

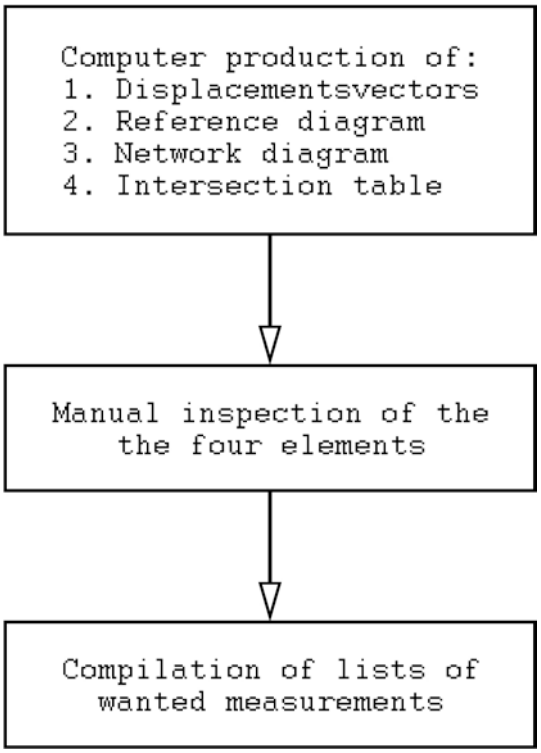


Figure 4: Network analysis.

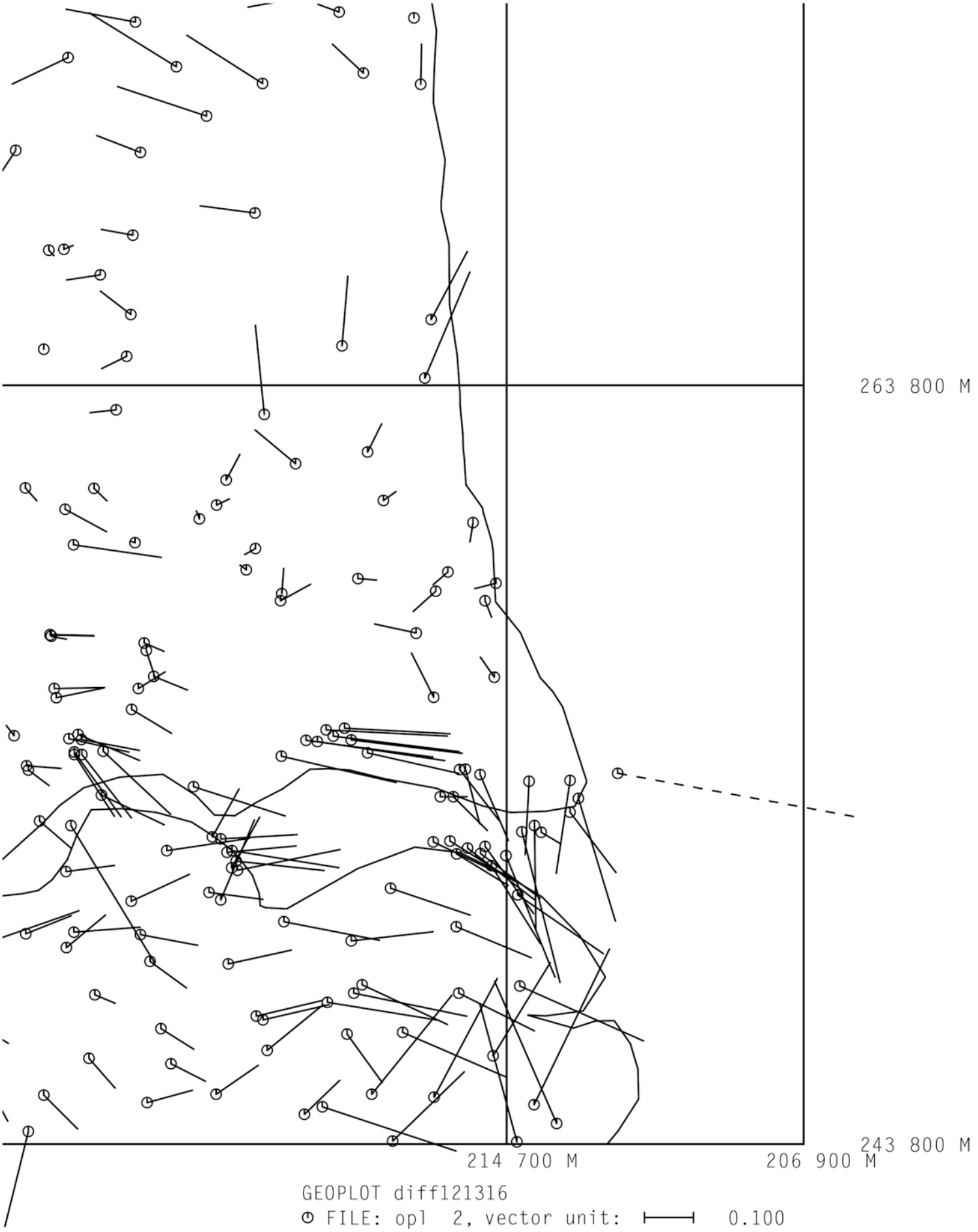


Figure 5: Section of displacement vector diagram.

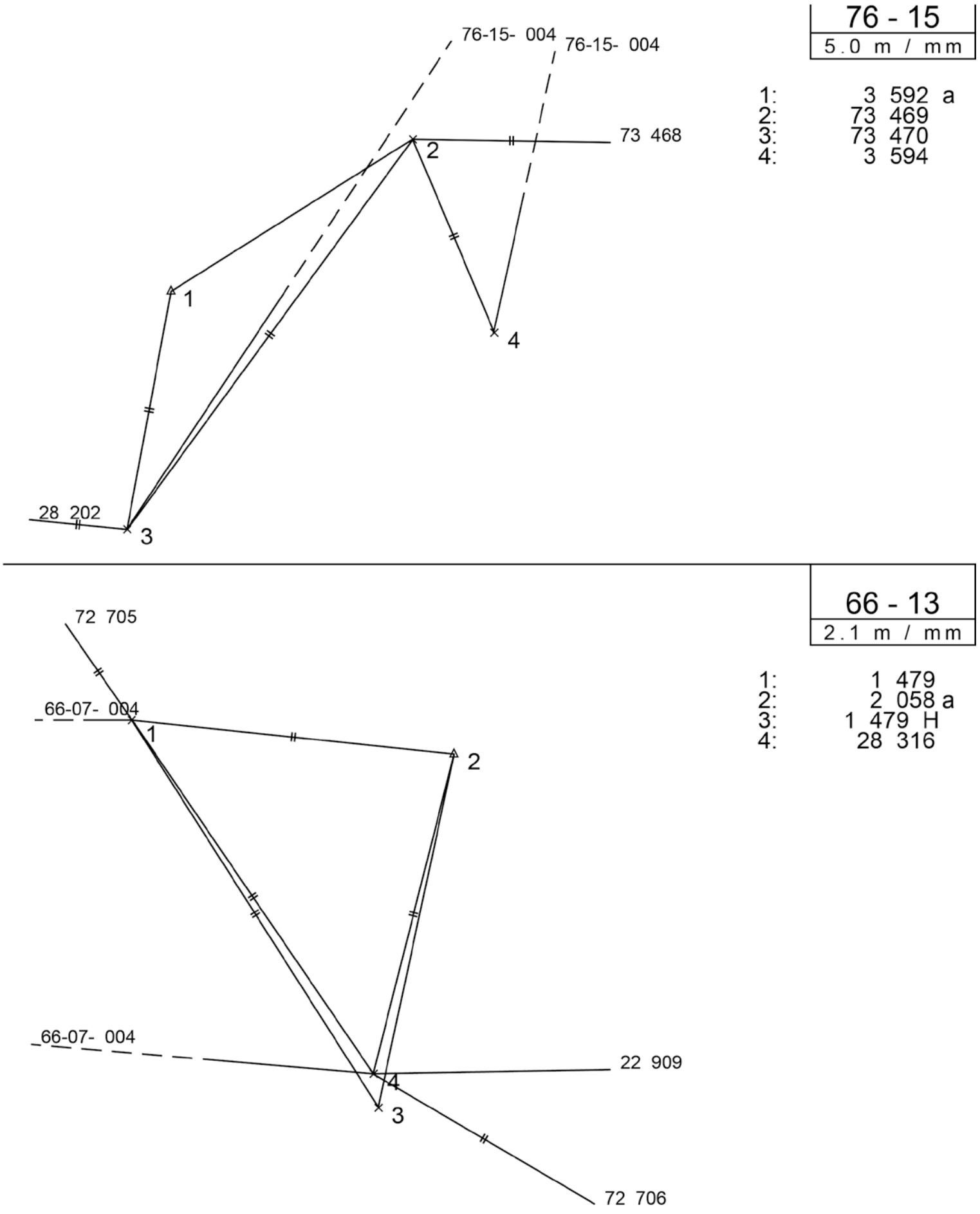


Figure 6: Reference diagrams.

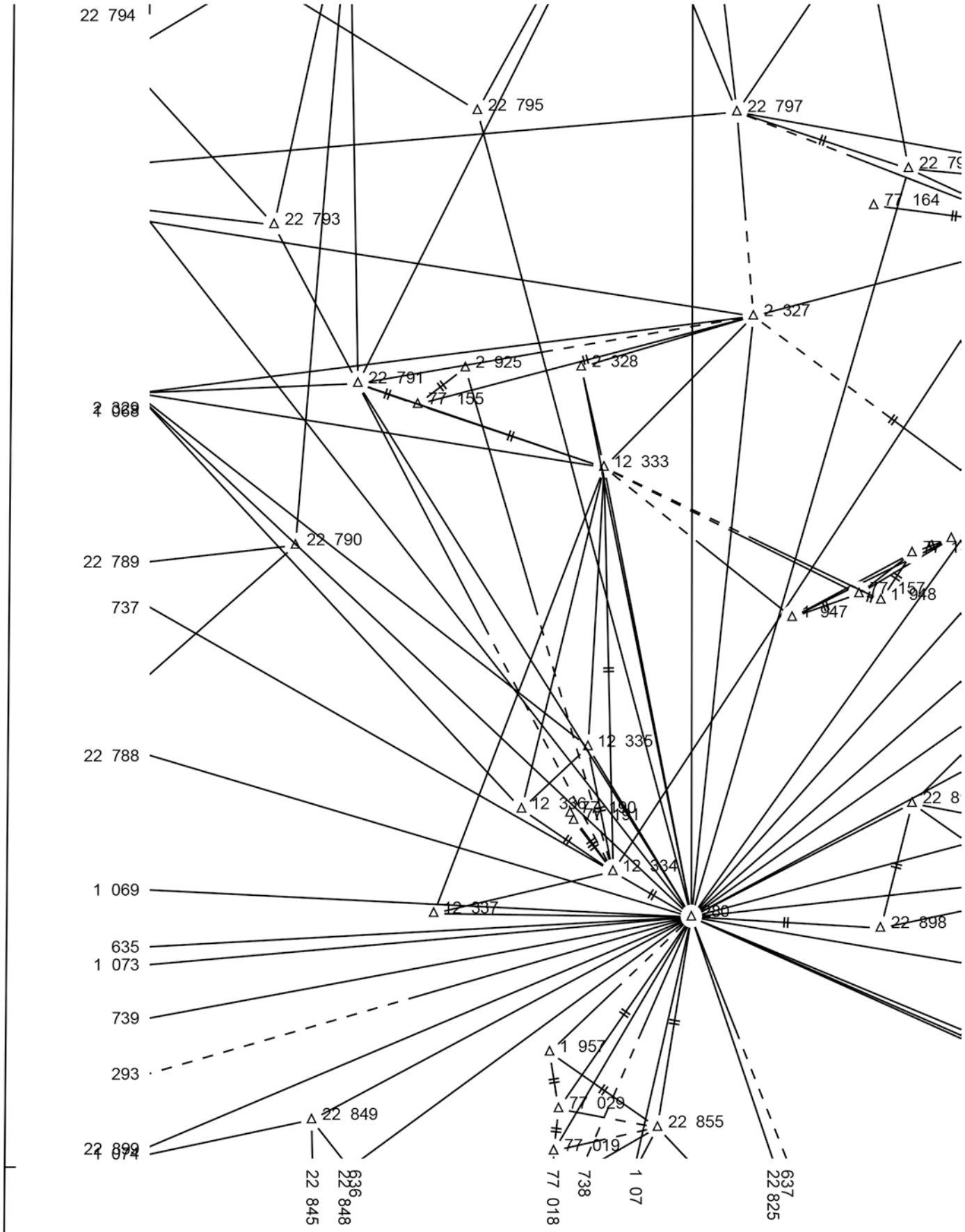


Figure 7: Section of network diagram.

* connections run dato 27 03 81, 5.54 side 19

master a connection	polar coordinate		direction to	direction from	dist	zdist to	zdist from	height diff
~~57-05- 006~~ a								
268	114 dg	7.0 km		f				
2 219	183 dg	0.4 km		f				
2 220	41 dg	8.2 km		f				
57-05- 007 a								
268	123 dg	7.6 km		f				
2 219	181 dg	1.7 km		f				
2 220	47 dg	7.3 km		f				
57-05- 816 a								
2 219	189 dg	0.2 km		f				
24 137	241 dg	1.6 km		f				
24 140	56 dg	3.0 km		f				
24 142	305 dg	3.1 km		f				
57-06- 829 a								
2 229	274 dg	0.4 km		f				
24 067	236 dg	3.2 km		f				
24 080	83 dg	2.2 km		f				
24 081	133 dg	2.4 km		f				
57-06- 842 a								
2 229	205 dg	0.6 km		f				
24 081	139 dg	3.0 km		f				

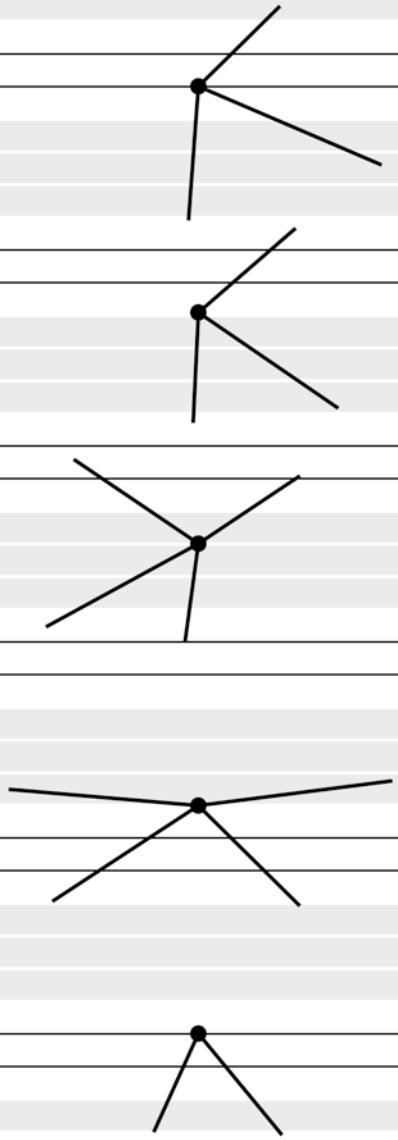


Figure 8: Table of directions to intersection points.

V. Conclusion.

The software described here is not only used for the revision of the Danish densification network, but also for surveys in Greenland and the Faroe Islands, and it is used for height- and three-dimensional adjustment (including photogrammetry combined with classical geodetic observations). This, and the active co-operation between the user and the developer (which is often the same person), has been of great importance for the development of the described software and techniques. Most of the development is done by or under the supervision of stategeodesist Knud Poder.

Current developments are within the areas of

- hardware development (parallel processor for computation of 70-bits product-sums),
- automatic Helmert blocking,
- network quality (both in terms of coordinate precision and local redundancy of observations,
- determination of systematic parameters.

VI. References.

(1) Poder, Knud and Frede Madsen: Status of the New Adjustment in Greenland and the Adjustment System of the Danish Geodetic Institute.
Proceedings of the Second International Symposium on Problems Related to the Redefinition of North American Geodetic Networks. pp 9 - 19.
US National Oceanic and Atmospheric Administration, 1978.

(2) Poder, Knud: Processing of Datamatically Blocked Normal Equations.
Ibid. pp 327 - 331.

(3) Borch, Inge: RC8000 Indexed Sequential Files.
March 1979, RCSL No. 31-D558.
A/S Regnecentralen, 1979.

(4) Poder, Knud: Geodetic Software Problems.
Symposium on Geodetic Networks and Computations.
Munich 1981.

(5) RC8000, System Architecture.
April 1979, RCSL No. 42-i1221.
A/S Regnecentralen, 1979.

Appendix A.

The entire revision.

1. Station and observation registers are brought up to date.
2. Compilation of station lists for field use. For each station the list contains the internal GI-sequence number, the official reference number, station name, status and map reference.
3. Copying of recovery diagrams.
4. First preliminary adjustment, used to check observations and registers and to produce the basis for the network analysis.
5. Network analysis producing lists of wanted observations.
6. Field work including revision of recovery diagrams, reconnaissance and observations. Replacements for lost points are established, taped traverses are remeasured and the observations indicated by the network-analysis are performed.
7. Entry of observations, preliminary coordinates and cross-references.
8. Combined adjustment of old and new observations.
9. Check of the adjustment. May indicate a few defects requiring additional observations. In that case the process is continued at step 6. Otherwise is continued at step 9.
10. The final network analysis with network diagram, reference diagrams and intersection station list.
11. The coordinates are ready for publication. Lists are delivered to the sales department, values are added to the hard-copy file and announcements are made in the relevant technical journals.

The complete revision of a block lasts about 2 year, due to the field work. The computations, including the network analysis, takes about 2-3 month involving 2-3 men.

The annual publication is approximately 2500 stations in both UTM and national system, i.e. 5000 sets of coordinates. The total amount of adjusted stations per year is approximately 10000 as many more stations are involved in preliminary adjustments and verifications.

Appendix B.

The dual number system.

Within the Geodetic Institute two parallel number systems are used. All coordinated and permanently marked points have official numbers, but only those established by the Institute gets internal numbers. (The National Cadastral Service has established some of the stations used by the Institute). Due to the cross-references in the station registers the two number systems can be mixed.

The official number has three sub numbers corresponding to the former administrative division of the country (see fig. 9). The sub-numbers are the county number, the parish number within the county and a sequence number within the parish. Town areas are numbered separately, indicated with the letter K instead of county number. The official system is used for publication purposes by the Institute and is used by the National Cadastral Service and local surveyors.

The internal number is a simple 5-digit sequence number. It is used during field work and computations and for archival purposes.

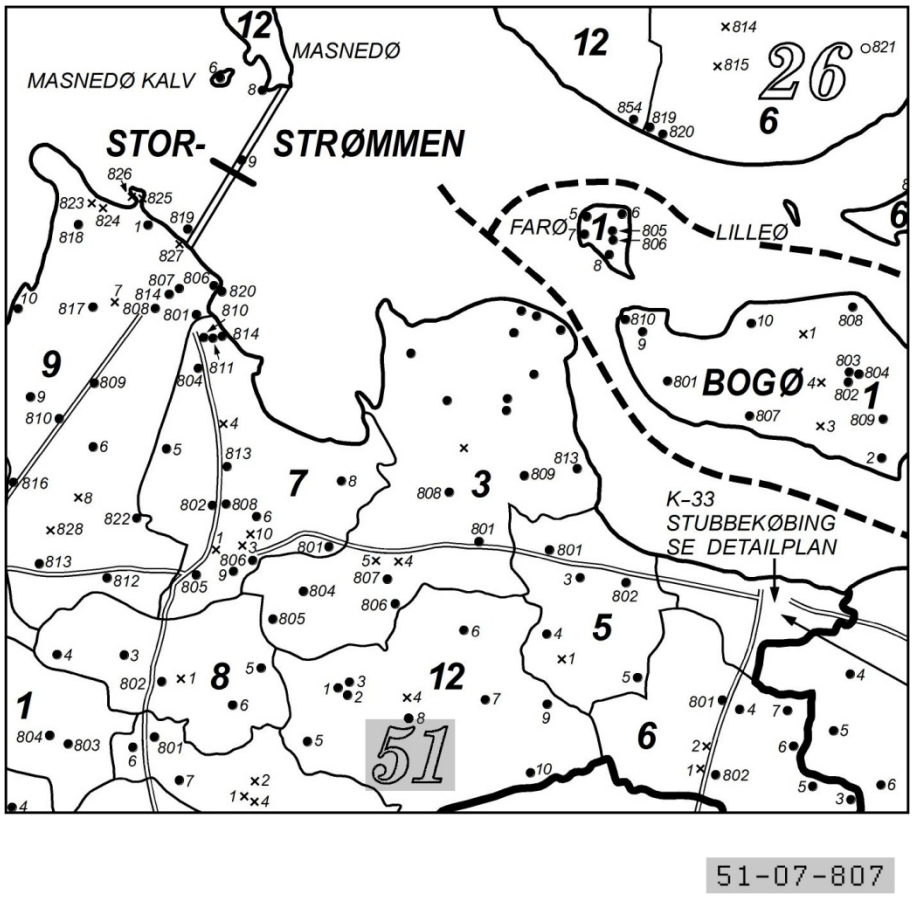


Figure 9: The number systems.

Appendix C.

Time studies.

Size of area	km^2	2500	1200	600	200
Coordinate retrieval of	polygon	400	250	300	150
	survey district	150	120	50	30
Compilation of jobdefinition		50	40	35	10
Input to adjustment-program of	jobdefinition	400	200	120	40
	coordinates	70	30	25	15
	observations	350	200	180	30
Size of adjustment	new stations	2100	1000	500	100
	fix stations	1350	1500	1000	200
	observations	19000	10400	5800	800
	redundancies	10800	5800	3300	380
	sets	5200	2500	1400	230
Execution times for one solution (unit = cpu sec.)		3500	2100	900	200

All times are in seconds. The programs are executed as batch jobs under usual machine load.

ACCURACY AND RELIABILITY MEASURES CONCERNING DESIGN AND QUALIFICATION OF DENSIFICATION NETWORKS

Wolfgang AUGATH
Hannover, Federal Republic of Germany

ABSTRACT

Preliminary a definition of the word "densification network" is given. Then some numerical analyses about measures of point accuracy and user's accuracy in typical densification control networks are made. In the same way the reliability is analysed. Proposals about the design of networks are derived.

1. Introduction

A report about criteria of quality for geodetic networks was given in another section of this symposium. At this place the application of general criteria of quality shall be discussed.

1.1 Definition of densification networks

A densification network differs from other networks by the position in the hierarchy of networks and by special aspects in the adjustment. In the establishment of networks in several steps it was usual, to takeout the primary network (1. order network) as bearer of the parameters of the reference system and to call the following networks (2. - 4. order networks, traverse network) densification network. Nowadays this is no absolute rule. First because it would be possible to calculate all points in one step. On the other part, a first order network could be calculated as densification network of some superior Doppler-points. On this point of view it is not possible, to make an assignment without doubt.

It remains the possibility to look for the differences in the adjustment. A densification network differs from other networks (p.e. free networks) in such a way, that the necessary parameters of the reference system as "orientation" or "scale" are leaded off from given points (fixed points).

A further restriction is not possible because of the various technical possibilities in our days. If we discuss criteria of quality for densification networks, all methods of measurement and ranges between the points must be included.

1.2 Restrictions by refections about the costs

Since electronic distance meters are available, densification networks can be established in various ways. The most important methods of measurement are put together in table 1, partitioned by the typical ranges between the points. She contains the traditional methods of angle measurement, the methods of measurement with microwave instruments, with electrooptical distance instruments and combined methods. From the combined methods, only the traverse networks are interesting, because they are especially cheap. The costs in DM only contain the expences for reconnaissance, signalization and measurement and they are coming from an internal calculation of the year 1975.

We notice the following principles:

1. The methods of horizontal angle measurement are too expensive, especially by greater ranges (expensive signalization, long times without possibilities to measure).
2. Distance networks can be recommended in all three ranges. In the two upper ranges the distance measurement with microwave instruments are extremely cheap (cheap signalization, no times without possibilities to measure).
3. In the lowest range between the points traverse networks should be preferred. In this method of measurement normally no signalization is necessary and a lot of additional points will be produced.

All together the following kinds of networks remain:

- traverse networks
- distance networks, measured with microwaves instruments
- distance networks, measured with electrooptical instruments

horizontal

measurement methods / distances between the points	angle measurement	distance measurement with micro-waves	electro-optical distance measurement	combinations
S > 30 km	δ_R = 0,1 mgon	$\frac{\delta_S}{S}$ = 1-2 · 10^{-6}	$\frac{\delta_S}{S}$ = 0,5-1 · 10^{-6}	–
S = 5 - 20 km	= 0,2 mgon	= 2-3 · 10^{-6}	= 1 · 10^{-6}	–
S = 2 - 5 km	= 0,3 - 0,5 mgon	–	δ_S = 10 mm	δ_R=0,5 mgon δ_S=5 mm
remarks				nets of traverses

Table 1a: Accuracy of measurement methods

measurement methods / distances between the points	angle measurement	distance measurement with micro-waves	electro-optical distance measurement	combinations
> 30 km	63.000 DM	7.000 DM	33.600 DM **) 21.600 DM***)	–
5 - 15 km	10.300 DM	2.600 DM	5.000 DM	–
2 - 5 km	2.000 DM	–	2.000 DM	1.000 DM - 2.000 DM
remarks	single-points			nets of traverses

Table 1b: Costs per point of the measurement methods

measurement methods / distances between the points	km per point	angle measurements	distance measurement with micro-waves	electro-optical distance measurement	combinations
> 30 km	2500 km	25 DM	2,8 DM	8,4 DM***) (13,5)**)	–
5 - 15 km	75 km	137 DM	35 DM	66 DM	–
2 - 5 km	5 km	400 DM	–	400 DM	200-400 DM
remarks		single points			nets of traverses

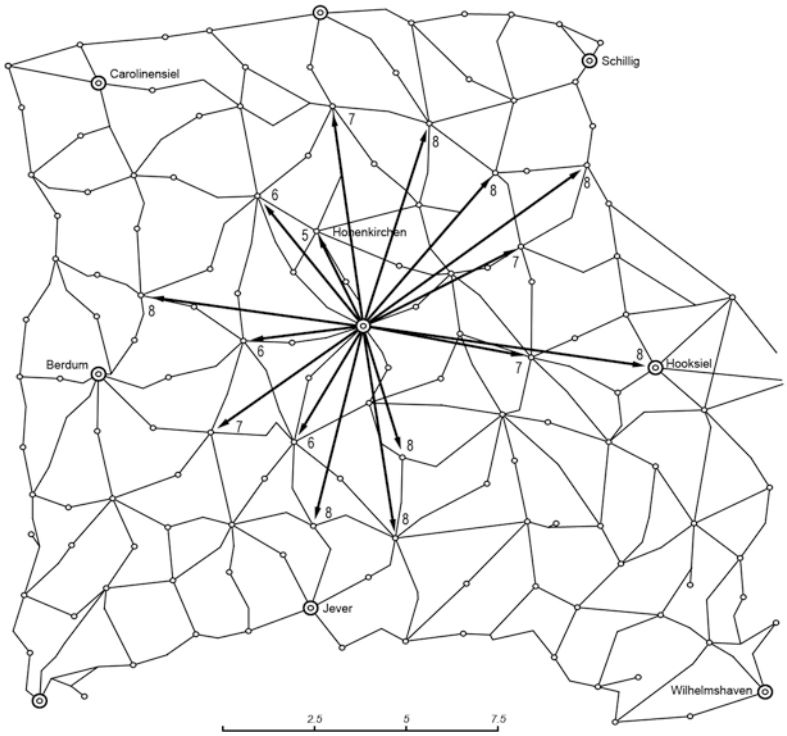
Table 1c: Relative costs of the measurement methods (costs per km^2)

---

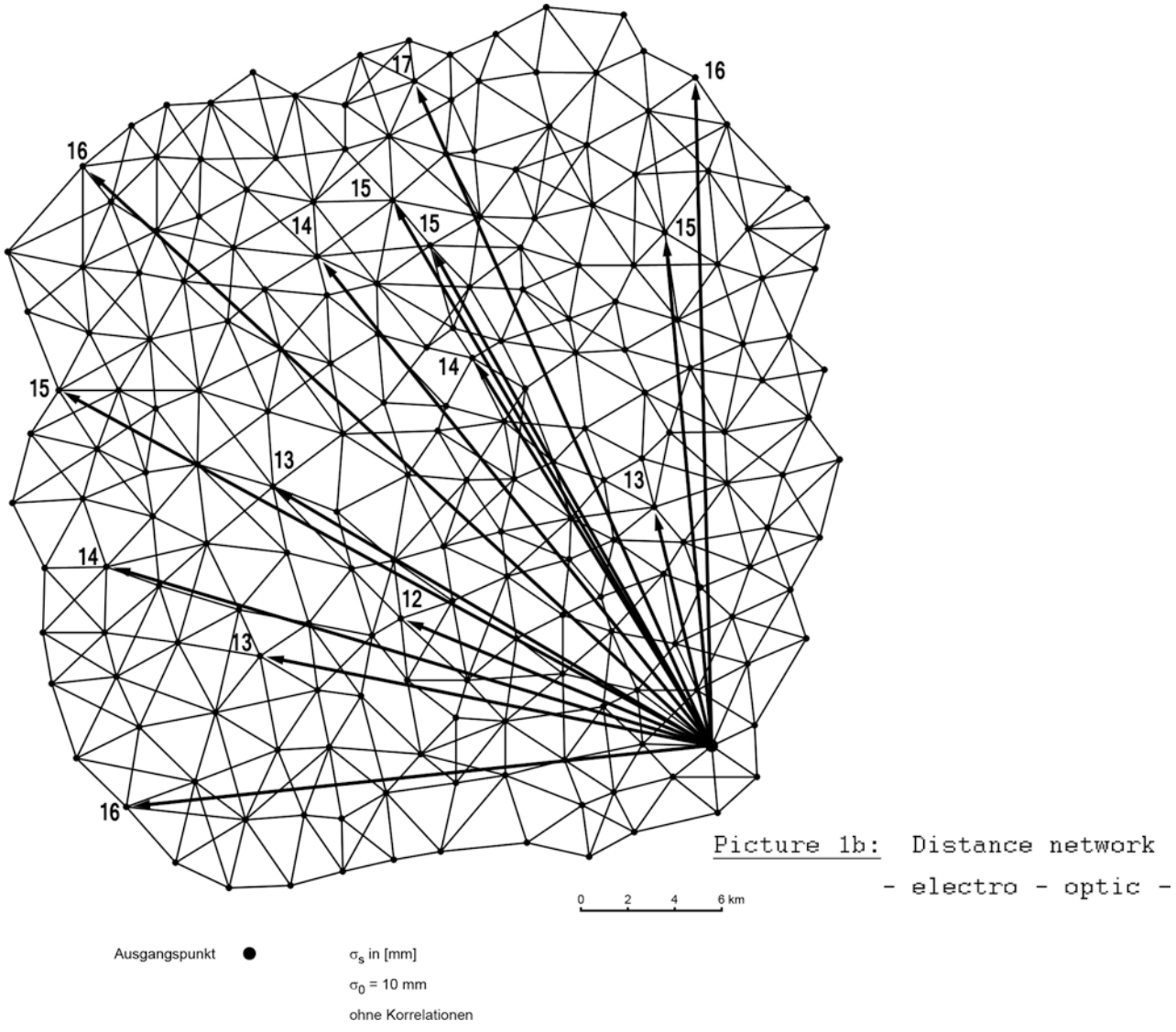
*) Definition of the measurement methods see AUGATH (1976)

**) Signalization with wooden towers

***) Signalization with iron towers

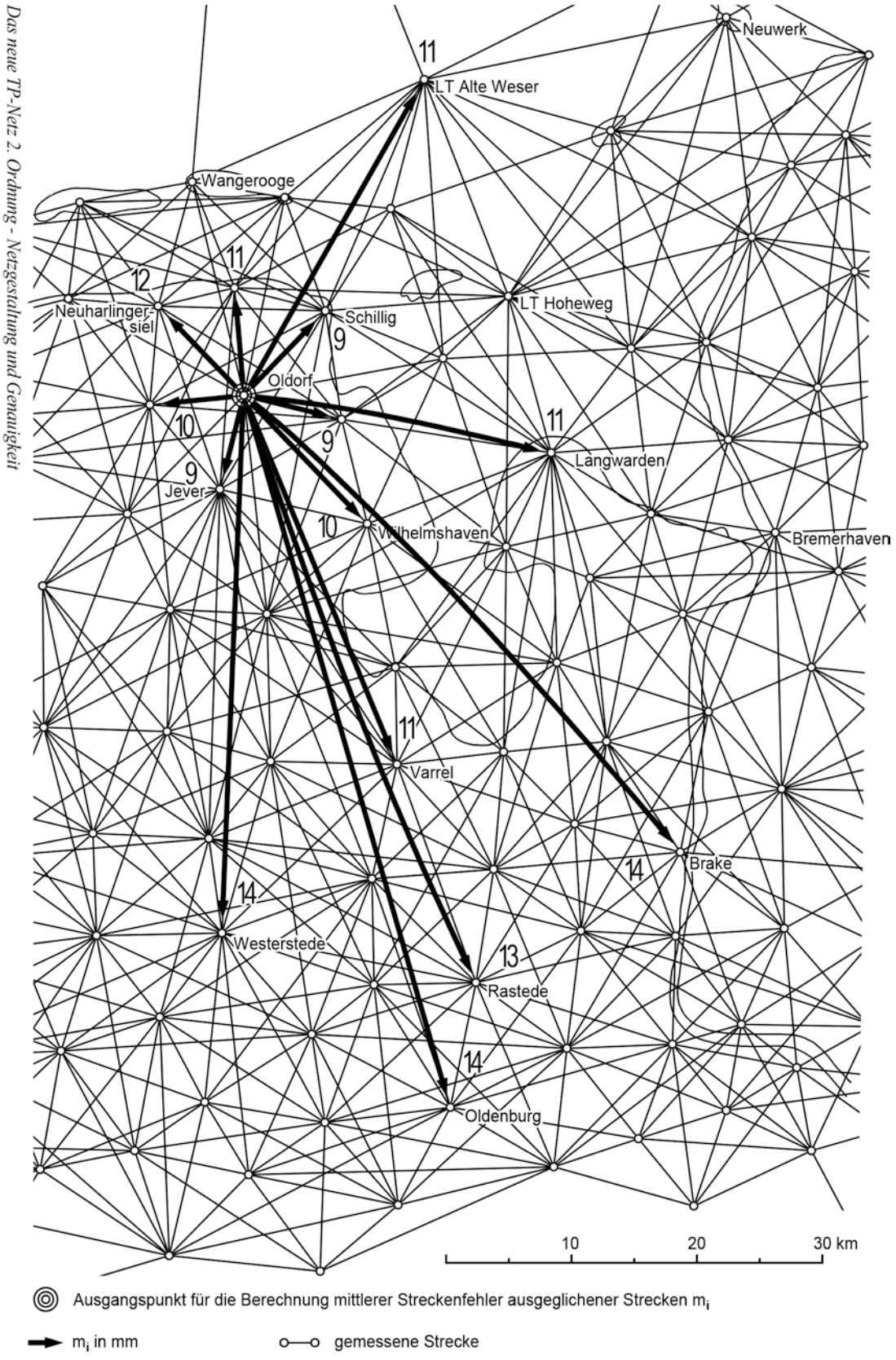


Picture 1a: Traverse network



Picture 1b: Distance network - electro - optic -

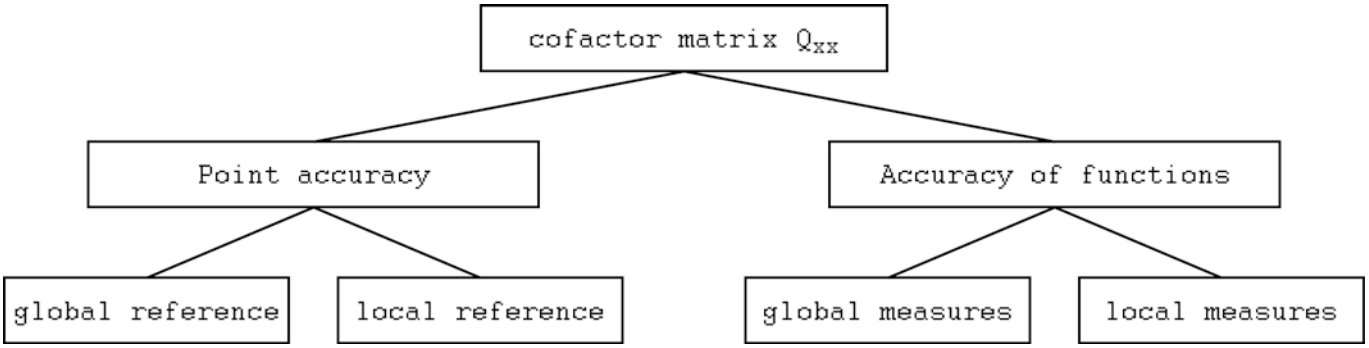
Picture 1: Typical densification networks



Picture 1c: Distance network - micro waves -

2. Accuracy measures of densification networks

Generally in densification networks the same accuracy measures may be used as in all other networks. But not all accuracy measures are efficient in the same way. All reflections start with the cofactor-matrix of the coordinate unknowns Q_{XX}. We can notice the possibilities of picture 2, which may be classified in measures of point accuracy and in measures of accuracy of functions (functions of the coordinate unknowns). We have to notice that the values of point accuracy depend on the reference parameters and we have to notice the dependence of the measures of accuracy of functions from the range between the points.



Picture 2: Measures of accuracy in densification networks.

2.1 Measures of point accuracy

The producers of horizontal control networks normally deliver values for the accuracy as point accuracy. We know

- standard deviations of the coordinate unknowns
- point error of HELMERT
- point error of WERKMEISTER
- error ellipses.

The theoretical advantages and disadvantages of the different measures disappear in the regularly networks of survey offices (AUGATH, 1976). Therefore it would be admissible to use the measure, which may be calculated in the simplest way (p.e. AdV, 1967: standard deviation of the coordinate unknowns) or to use means of measures for a whole network (MI, 1982). But we should always notice, that these measures of accuracy only describe the relative accuracy in respect of the fixed points. Normally these values are not important. We also have to notice a strong dependence of the values of the measures and the special references system

of the adjustment. By a later comparence of values, very often this fact is not considered enough.

To increase the possibilities of a comparison some authors propose to use measures of point accuracy with a local reference (SCHMIDT, 1968) (PELZER, 1980):

$$PA_{(local)} = S_o \cdot f\,(Q_{xx}) \qquad (1)$$

PA = measure of point accuracy

S_o = standard deviation of unit weight

In formula (1) the value of S_o must be taken from the total adjustment. The values of $-Q_{xx}$ have to be calculated in a special single point adjustment for the examined point with all his Observations.

2.2 Measures of the accuracy of functions (user's accuracy)

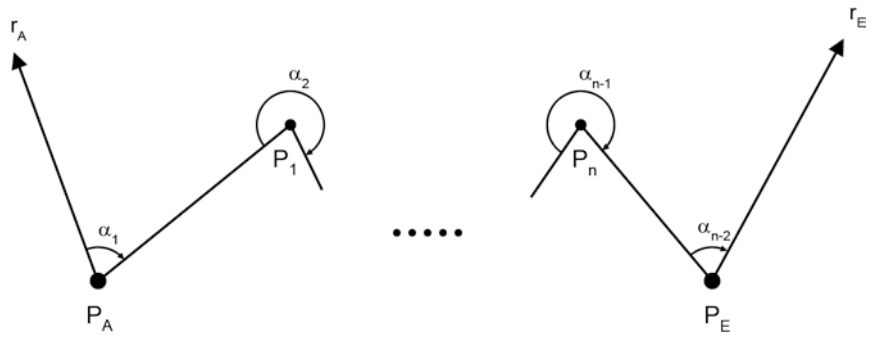
Normally the user of horizontal control networks is not interested in the accuracy of points. He always combines two points or more with his observations and therefore the accuracy of functions of the coordinates are most important. Such functions may be:

- differences of coordinates
- distances
- directions
- misclosures of traverses
- differences of double observations of polar points

.
.
.

a.s.o.

The choice of these functions depends on the special method of measurement of the user. In most cases the accuracy of the network will be important if the user has to discuss misclosures, which depend on the accuracy of the observations and on the accuracy of functions of the coordinates of the junction points.



Picture 3: angular closure f_β of traverses

The misclosure of the directions f_β of the traverse of picture 3 can be calculated with the formula

$$f_\beta = r_A - r_E + \sum(\alpha \pm 200^g) \qquad (2)$$

r_A, r_E = directions, calculated with coordinates

α = measured angles

We get the variance $S^2_{f_\beta}$ in respect to the law of the propagation of errors to

$$S^2_{f_\beta} = S^2_{r_A} + S^2_{r_E} + \sum S^2_\alpha \qquad (3)$$

The variances $S^2_{r_A}$, $S^2_{r_E}$ are functions of the coordinate unknowns and describe that part of $S^2_{f_\beta}$, which depends only on the network. It is very important for the user to know the largeness of this part.

The designer of a network cannot know in advance, which method of measurement will be applied. In this case it is useful to estimate the accuracy of functions in a general way, as done by PELZER (1979). The accuracy of a function f = f(x) is

$$S^2_f = S^2_0 \; \underline{f} \; \underline{Q}_{xx} \; \underline{f}^T \qquad (4)$$

S^2_0 = variance of the unit weight

With the inequality for the RALEIGH quotient

$$\lambda_{min} \leq \frac{\underline{f}\ \underline{Q}_{xx}\ \underline{f}^T}{\underline{f}\underline{f}^T} \leq \lambda_{max} \qquad (5)$$

λ_{min}, λ_{max} = minimum respec. maximum eigen - value of $\underline{Q}_{xx}$

we get:

$$S_f^2 \leq S_o^2\ \underline{f}\underline{f}^T\ \lambda_{max} \qquad (6)$$

A comparison shows, that geodetic functions always have lower values than the boundary, given by λ_{max} (AUGATH, 1976). We can notice the following facts:

1. With the condition $\lambda_{max} \rightarrow$ min ! we minimize among other things the accuracy of functions.
2. An analysis of the vector product $\underline{f}\underline{f}^T$ in formula (6) shows us, that the methods of measurement which are "neighbourship friendly" (p.e. orthogonal survey) have extremely lower variances of the misclosures which depend on the network than others (p.e. polar survey), which are not "neighbourship friendly" (proportion 2 : 6).
3. The variance of the functions "distance between adjacent points" may be estimated by the eigenvalue λ_{max} of a local matrix $\underline{Q}_{xx}$ as defined in formula (1).

For the future the German survey offices will give a local measure of the accuracy of functions in addition to the point accuracy (AdV, 1981). This function shall be the "distance between adjacent points".

3. Measures of reliability of densification networks

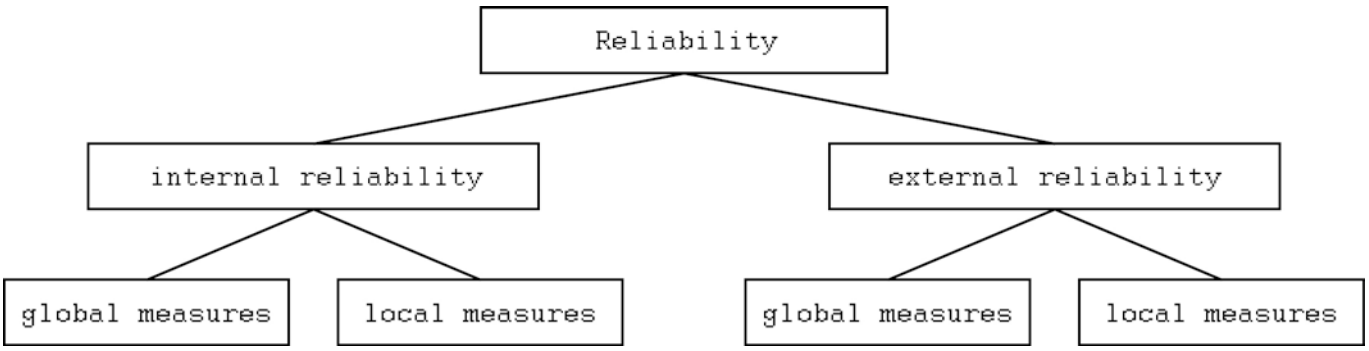
With the word "reliability" the old problem "how should be controlled our geodetic work in the best way?" has been discussed very often in the last years. If we analyse the work in geodetic networks, we have to recognize three parts:

1. First we need reliable observation staffs with a great practical knowledge and concentration. In this case we can be sure, that certain mistakes will not be possible or very seldom.

2. Second we need reliable methods of measurement. The conception of these methods of measurement must be constructed in such a way, that another part of errors will be automatically eliminated or be shown to the observer in the field.
   (p.e.: forward and backward measurements in levelling)

3. Last not least we need a reliable design of networks, which is built in such a way, that the observations which participate in the adjustment control themselves.

In the following pages only the mathematic part of reliability shall be discussed. There we have to point out model errors (errors between the reality and the stochastic and functional model of the adjustment) and to describe the consequences of possible model errors on the results of the adjustment. FÖRSTNER (1979) calls these two parts also "internal" and "external" reliability. For both parts exist global and local measures of reliability.



Picture 4: Measures of reliability

3.1 Internal reliability

If we divide the vector $\underline{l}$ of the observations in two parts in a way, that the part $\underline{\bar{l}}$ corresponds with the model of the adjustment while a second vector $\underline{\delta}$ includes the errors to the model

$$\underline{l} = \underline{\bar{l}} + \underline{\delta} \qquad (7)$$

so we get with PELZER (1980) the global measure of internal reliability λ. λ describes the variation of the variance of the unit weight caused by model errors:

$$\lambda = \frac{1}{f}\ \delta^T\ \underline{P}\ \underline{Q}_{vv}\ \underline{P}\ \delta \qquad (8)$$

f = number of redundant observations

P = matrix of the weights of the observations

$\underline{Q}_{VV}$ = matrix of the cofactors of the residuals

We can't use the formula (7), because the structure of the vector _ is unknown. But we can except, that the global measure of reliability λ will be as greater as the elements of the diagonal of the matrix $\underline{Q}_{VV}$ increases. For the design of the network we can notice the global condition

$$\sum \text{diag. } Q_{vv} \rightarrow \max ! \qquad (9)$$

If we look after the local measures of internal reliability defined by different authors, we get the same statement for single observations, p.e.:

PELZER, 1977 : $z = \sqrt{1-\frac{\delta_i^2}{\delta_i^2}}$ (10)

FÖRSTNER, 1979 : $r = z^2$ (11)

PELZER; 1980 : $n = \frac{\delta_i^2}{\delta_i^2}$ (12)

These measures say, that single a observation is as better controlled as greater will be the values of z, r, n. The German survey offices uses the measure r; boundary values of r are yet not fixed (AdV, 1981). Table 2 shows us mean values of measures of reliability in typical networks.

method of measurement	distance networks - micro waves -	distance networks - electro-optical -	traverse networks
repetition number of the observation	2 - 4	1	1
r	0,8	0,5	0,01-0,04 mean: 0,1
z	0,9	0,7	0,3
n	5,2	2,0	1,1

Table 2: Values of measures of reliability in typical networks

3.2 External reliability

The dimension of the measures of internal reliability fix, which model errors will probably be detected. Finally we are interested to describe the effects of model errors, which could probably not be detected. We get the global effects δ_x, δ_f for the coordinate unknowns $\underline{x}$ or functions $\underline{F}$:

$$\underline{\delta}_x = \underline{Q}_{xx}\ \underline{A}^T\ \underline{P}\ \underline{\delta} \qquad (13)$$

$$\underline{\delta}_f = \underline{F}\ \underline{Q}_{xx}\ \underline{A}^T\ \underline{P}\ \underline{\delta} \qquad (14)$$

A = design matrix

The global reliability of a network depends on the length of the vectors $\underline{\delta}_x$ or $\underline{\delta}_f$. PELZER (1976) has proved, that these lengths depends on the maximum eigenvalue of Q_{XX}. This leads once more to the global condition $\lambda_{max} \rightarrow$ min !

(see also chapter 2.2).

If we replace in formula (13) or (14) the vector $\underline{\delta}$ by a vector $\underline{\Delta}$ ($\Delta^T = 0, 0, \ldots, \Delta_i, \ldots, 0$), we are able to calculate the local effects of a single gross error Δ_i of an observation l_i on the coordinates ($\rightarrow \Delta_X$) or on functions of coordinates ($\rightarrow \Delta_F$).

$$\underline{\Delta}_x = \underline{Q}_{xx}\ \underline{A}^T\ \underline{P}\ \underline{\Delta} \qquad (15)$$

$$\underline{\Delta}_F = \underline{Q}_{xx}\ \underline{A}^T\ \underline{P}\ \underline{\Delta} \qquad (16)$$

In (15) or (16) Δ_i may be replaces by values of ∇_{l_i}.
(BAARDA, 1967) (FÖRSTNER, 1969).

It is possible to calculate the vectors $\underline{\Delta}_X$ or $\underline{\Delta}_F$ for each observation. In this way we get a lot of numbers, which must be reduced. The study group "TRIG-PROGRAM" of the German survey offices examines this problem. They have yet no solution, but probably it will be sufficient to calculate only the effects on coordinates of neighbour points and on the distances between neighbour points.

REFERENCES

AdV, 1967, Richtlinien für einen Erlaß zum trigonometrischen Festpunktfeld (TP-Erlaß). Niederschrift zur 40. Tagung der AdV in Bad Neuenahr am 07. - 09.11.1967 (unpublished).

AdV, 1981, Niederschrift zur 20. Tagung des Arbeitskreises Triangulation der AdV in München am 04. - 06.03.1981 (unpublished).

AUGATH, W., 1976, Untersuchungen zum Aufbau geodätischer Lagenetze. Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie an der Universität Hannover Nr. 72, Hannover (1976).

AUGATH, W., 1977, The design of reliable geodetic networks. Symposium on optimization of design of control networks, Sopron (1977).

AUGATH, W., 1980, Optimierungskriterien und deren Auswirkung in Netzen der Landesvermessung. Geodätische Netze in Landes- und Ingenieurvermessung, Wittwer-Verlag, Stuttgart (1980), pp. 479 - 498.

GRAFAREND, E., 1974, Optimization of geodetic network. Bollettino di Geodesia e Scienze affini (1974).

FÖRSTNER, W., 1979, Ausgleichung und Gütebeurteilung geodätischer Lagenetze. Zeitschrift für Vermessungswesen (1979), pp. 61 - 72.

MI, 1982, Erlaßentwurf über die geodätischen Festpunktfelder. Hannover (1982), (unpublished).

PELZER, H., 1976, Genauigkeit und Zuverlässigkeit geodätischer Netze. Tagung Oberwolfach (1976).

PELZER, H., 1977, Criteria for the reliability of geodetic networks. Symposium on optimization of design of control networks, Sopron (1977).

PELZER, H., 1980, Beurteilung der Genauigkeit und Zuverlässigkeit geodätischer Netze. Geodätische Netze in Landes- und Ingenieurvermessung, Wittwer-Verlag, Stuttgart (1980), pp. 273 - 304.

SCHMIDT, R., 1970, Die Genauigkeit der TP-Netze höherer Ordnung und das Problem der Netzerneuerung in NRW. Nachrichten aus dem öffentlichen Vermessungswesen NRW, Düsseldorf (1970), pp. 1 - 15.

A SPECIAL METHOD TO DERIVE A CRITERION MATRIX

B. BANOV
Sofia, Bulgaria

ABSTRACT

Ways of deriving criterion variance-covariance matrix functions are examined in the paper. First of all the initial project is designed, in which the initial variance-covariance matrix of measurements is given. After this the accuracy of the functions of the unknowns is determined, which will ensure the purpose of the network. The mean square errors of the unknowns, the bearings, the angles, the relative error of the distances, the two elements of error ellipses of two orthogonal functions and so on were determined. The criterion variance-covariance matrix is formed in the iterative process.

Introduction

Ensuring the necessary accuracy with minimal expenses and minimal period for realisation of the project is the basic purpose of optimal projection. The solution of this problem is connected with taking into consideration various conditions, such as relief, vegetation, hydrography, climate, type of instrument, qualification of the engineers, the basic purpose of the network, possibilities for its continuous use and others. With these conditions, step by step or at the same time, during projection answers must be obtained to the following questions:

1. What will be the density of the points in the network, where points should and could be placed.
2. In what area and in what form should be the functions that will be used after the construction of the network. What will be the variance-covariance matrix of these functions.
3. What types of measurements are necessary and what possible. In what areas, in what form and of what accuracy should the measurements be completed, so as to ensure the purpose of the network.

Conditions for designing determine the limits and possibilities during the preparation of the project, and the answers to the above basic questions form the concrete solution to the designer. If the optimal project will have to be designed through mathematical model, the conditions, as well as the basic questions should have to receive analytical representation. A passing glance at the conditions, however, shows that only their representation is a big series of problems. The analytical representation of all conditions and questions, and the finding of optimal solution is a complicated, and at this stage of development of science and computer technique - a labour consuming and economically inefficient problem [ALBERDA, 1974]. The present day methods and computer technique allow the problem of optimal designing to be solved during the reduced mathematical models of the conditions, by iterative or still better by interactive computer processes. The reductions of the model of the conditions however widens the domain of the permissible solutions and the necessary continuous analysis of the solution. That is why analytical solution with this method of worm must be examined as an analysis and specification of the designer's solution. The reasons considered necessitate the process of optimal designing to be divided into different optimizational problems [GRAFAREND et al., 1979]. During their solution, irrespective of the purpose of the project, one of the basic requirements is to ensure the necessary accuracy of the chosen functions. For characterizing the accuracy during the solution of the optimizational problems, two methods of approach have been applied:

a) The use of generalized function for the accuracy (generalized criterion for the accuracy), which is rotation and translation invariant from the elements of variance-covariance matrix $C_{\varphi\varphi}$ (φ is a vector of the analyzed functions, which ensure the purpose of the network).

b) The use of the whole function variance-covariance matrix $C_{\varphi\varphi}$.

The application of generalised criterion leads to easier solution than the use of criterion matrix, but meeting the requirements of generalised criterion will not ensure the necessary accuracy of all analyzed functions and necessitates the introduction of more conditions.

In this second method variance-covariance matrix $K_{\varphi\varphi}$ meets the differential requirements for the accuracy. This matrix is normally called criterion variance-covariance matrix, and therefore $C_{\varphi\varphi} \sim K_{\varphi\varphi}$, the solution of the respective optimizational problem will be optimal. Through the application of $K_{\varphi\varphi}$ the possibility of designing some of the covariance is formed, if necessary.

Iterative Method of Obtaining $K_{\varphi\varphi}$. [BAARDA, 1977] distinguishes two ways of obtaining criterion matrix:

1. The network project is designed with a view to the relief, vegetation, hydrography, climate, the type of instrument and the purpose of the network. The variance-covariance matrix $C_{\varphi\varphi}$ is calculated, the "weak places" in the network are determined, and with modification of $C_{\varphi\varphi}$, $K_{\varphi\varphi}$ is obtained.

2. Matrix $K_{\varphi\varphi}$ is calculated in accordance with the theoretical precondition for the accuracy of the network, without taking into consideration the possible measurements and their dispersion matrix $C_{\ell\ell}$, $\underset{n,1}{\ell}$ - vector of the measurements. In the paper the first method is applied. With the application of $C_{\varphi\varphi}$ the possibility of using always functions with unbiased estimates is obtained, irrespective of the kind of the network. Such solution gives series of problems. Substantiated preconditions during the selection of the type, number and the place of the functions do not always exist. Despite the determined problems during their application, it is the general case, since $C_{xx} \in C_{\varphi\varphi}$ / $\underset{m,1}{x}$ - vector of the unknowns/. Therefore further deductions will be

done for matrix $C_{\varphi\varphi}$, respectively $K_{\varphi\varphi}$.

For obtaining $K_{\varphi\varphi}$ we will make use of widely applied differential criterion for accuracy, which is used in geodetic practice. We will note part of them.

a) The distance relative error

/1/ $$\eta_{S_{ij}} = \frac{m_{S_{ij}}}{S_{ij}} = 1 : \left(S_{ij}/m_{S_{ij}}\right)$$

where $\eta_{S_{ij}}$ is the relation, $m_{S_{ij}}$ is the square error of the distanced S_{ij}. During designing $1 : \bar{\eta}_{S_{ij}}$, from where the dispersion of the function S_{ij} is obtained

/2/ $$\bar{m}^2_{S_{ij}} = \left(S_{ij}\ \bar{\eta}_{S_{ij}}\right)^2$$

/with the line above the values which are given or calculated from the given, are noted/.

b) The cross error q_{ij}

/3/ $$q_{ij} = \frac{m_{\alpha_{ij}}}{\varrho}\ S_{ij}$$

where ϱ is the value of one radian, α_{ij} is the bearing and $m_{\alpha_{ij}}$ is the square error of α_{ij}. If $\bar{q}_{ij}$ is given, for $\bar{m}^2_{\alpha_{ij}}$ is obtained:

/4/ $$\bar{m}^2_{\alpha_{ij}} = \left(\frac{\bar{q}_{ij}\ \varrho}{S_{ij}}\right)^2$$

c) The mean square error in the mutual position of two points m_{ij}

/5/ $$m_{ij} = \left(m^2_{S_{ij}} + q^2_{ij}\right)^{\frac{1}{2}}$$

If $\bar{m}_{ij}$ is given and the relation of angular and distant measurements is accepted $r_1 = m^2_{S_{ij}}/q^2_{ij}$, the dispersion of two functions is obtained

/6a/ $$q_{ij} = \frac{\bar{m}^2_{S_{ij}}}{1 + r_1}$$

/6b/ $$\bar{m}^2_{\alpha_{ij}} = \frac{1}{1 + r_1}\left(\frac{\bar{m}_{ij}\ \varrho}{S_{ij}}\right)^2$$

d) The relative error in the mutual position of two points

/7/ $$\eta_{ij} = \frac{m_{ij}}{S_{ij}}$$

With given $1/\bar{\eta}_{ij}$ and obtained $\bar{m}_{ij} = \bar{\eta}_{ij}\, S_{ij}$ by /6a/ and /6b/ dispersion of the functions is obtained again.

e) Mean square error of the angle β m_β - in this case dispersion of the angle $\bar{m}_\beta^2$ is obtained directly.

f) Mean square error in the position of point

/8/ $$m_{S_i} = \left(m_{x_i}^2 + m_{y_i}^2\right)^{\frac{1}{2}}$$

If given the relations $r_2 = \bar{m}_{x_i}^2 / \bar{m}_{y_i}^2$ and $\bar{m}_{S_i}$ from /8/ then

/9a/ $$\bar{m}_{y_i}^2 = \frac{\bar{m}_{S_i}^2}{1 + r_2}$$

/9b/ $$\bar{m}_{x_i}^2 = \frac{r_2}{1 + r_2}\ \bar{m}_{S_i}^2$$

g) Error ellipses

Let φ_1, and φ_2 be two orthogonal functions. Let $\begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$ be a cell from the diagonal of the weight matrix with coefficients $\underset{\kappa\kappa}{Q}_{\varphi\varphi} = \frac{1}{\mu^2} C_{\varphi\varphi}$ corresponding to φ_1 and φ_2. For the elements of the error ellipses for the two functions are in force:

/10/
$$R_{max}^2 = \frac{\mu^2}{2}\left(q_{11} + q_{22} + [(q_{11} - q_{22})^2 + 4q_{12}^2]^{\frac{1}{2}}\right)$$
$$R_{min}^2 = \frac{\mu^2}{2}\left(q_{11} + q_{22} - [(q_{11} - q_{22})^2 + 4q_{12}^2]^{\frac{1}{2}}\right)$$
$$\operatorname{tg} 2\psi = 2\, q_{12}/(q_{11} - q_{22})$$

where R_{max} and R_{min} are respectively big and small axis of the ellipse, ψ is the bearing of the bigger axis and μ^2 is the dispersion for unit weight. Within the three parameters of the error ellipses R_{max}, R_{min} and ψ, and the respective elements of the weight matrix $Q_{\varphi\varphi}$ or more correctly $C_{\varphi\varphi}$, exists simple correspondence. If $\bar{R}_{max}$, $\bar{R}_{min}$, and $\bar{\psi}$ are given, then $\bar{q}_{11}$, $\bar{q}_{22}$ and $\bar{q}_{12}$ respectively the corresponding elements of $K_{\varphi\varphi}$ can be determined. If we introduce

/11/
$$2a = q_{11} + q_{22} = \frac{1}{\mu^2}\left(R_{max}^2 + R_{min}^2\right)$$
$$2b = [(q_{11} - q_{22})^2 + 4q_{12}^2]^{\frac{1}{2}} = \frac{1}{\mu^2}\left(R_{max}^2 - R_{min}^2\right)$$

After substituting with $\bar{R}_{max}$, $\bar{R}_{min}$ and $\bar{\psi}$ in /11/ and taking into consideration the third formula from /10/, then

/12/
$$\bar{q}_{11} = \bar{a} \pm \bar{b}\cos 2\bar{\psi}$$
$$\bar{q}_{22} = \bar{a} \mp \bar{b}\cos 2\bar{\psi}$$
$$\bar{q}_{12} = \bar{b}\sin 2\bar{\psi}$$

Out of the two solutions, it is more advisable to use the one which satisfied $(q_{11} - q_{22})(\bar{q}_{11} - \bar{q}_{22}) > 0$, that is the one which maintains the direction of inequality between the diagonal coefficients in $C_{\varphi\varphi}$. Through the same way if the elements of error ellipses of the remaining functions from the network are given, the corresponding coefficients from $K_{\varphi\varphi}$ can be determined.

In the considered 7 cases so far are given formulas, with which the given values of the dispersion of the distances, bearings and co-ordinates can be determined. It can be pointed out however, that the simple functional co-ordinates are not invariational during the rotation of the co-ordinate system, although their values are determined from invariational quantities $\bar{R}_{max}$, $\bar{R}_{min}$ and $\bar{\psi}$. Within the considered 7 cases, only in /12/ the next diagonal elements can be determined, and in others only the respective diagonal elements from $K_{\varphi\varphi}$ can be determined. For the determination of non-diagonal elements of $K_{\varphi\varphi}$, the ideas of [VAGIN, 1979] for leveling network and [BANOV, 1980] for angular-distant network can be applied. Let the variant V_o of the project is designed with conditions for realising the purpose of the network. Let also the respective variance-covariance matrix is determined then

/13/ $$C_{\varphi\varphi}^{(o)} = B\, C_{\varphi\varphi}^{(o)}\, B^{*}$$

where $B_{k,m}$ contains coefficients during the expansion of the functions in the Tailor's series. If the diagonal elements of $C_{\varphi\varphi}^{(o)}$ symbolised with C_{ii} and in the matrix

/14/ $$M_{\varphi\varphi}^{(o)}{}_{\kappa,\kappa} = \operatorname{diag}\left(\sqrt{C_{11}}, \sqrt{C_{22}}, \ldots, \sqrt{C_{\kappa\kappa}}\right)$$

are arranged the determined mean square errors, multiplied from left and right with $\left(M_{\varphi\varphi}^{(o)}\right)^{-1}$ then matrix with coefficients of correlation between the functions can be determined:

/15/ $$R_{\varphi\varphi}^{(o)}{}_{\kappa,\kappa} = \left(M_{\varphi\varphi}^{(o)}\right)^{-1} C_{\varphi\varphi}^{(o)} \left(M_{\varphi\varphi}^{(o)}\right)^{-1}$$

where $r_{ij} = C_{ij}/\left(C_{ii}\, C_{jj}\right)^{\frac{1}{2}}$, $i,j = 1,\ldots,\kappa$ are elements of matrix $R_{\varphi\varphi}^{(o)}$.

Let the determined dispersions with /1/ ÷ /12/ be symbolised with $\hbar_{ii}$, $i = 1, \ldots, \kappa$ and calculated mean square errors were arranged in the matrix:

/16/ $$\underset{\kappa,\kappa}{D_{\varphi\varphi}} = \operatorname{diag}\left(\sqrt{\hbar_{11}}, \sqrt{\hbar_{22}}, \ldots, \sqrt{\hbar_{\kappa\kappa}}\right)$$

Then the first approximation of criterion matrix can be determined with the formula

/17/ $$K^{(1)}_{\varphi\varphi} = D_{\varphi\varphi}\, R^{(o)}_{\varphi\varphi}\, D_{\varphi\varphi}$$

If in /17/, /15/ is substituted, then it can be determined finally

/18/ $$\boxed{K^{(1)}_{\varphi\varphi} = D_{\varphi\varphi} \left(M^{(o)}_{\varphi\varphi}\right)^{-1} C^{(o)}_{\varphi\varphi} \left(M^{(o)}_{\varphi\varphi}\right)^{-1} D_{\varphi\varphi}}$$

which shows, that $K_{\varphi\varphi}$ determined from $C_{\varphi\varphi}$ after multiplying from left and right with the same positively determined matrix. These multiplications lead to elementary modifications [RAO, 1968], in which the structure and the rank of the initial matrix $C_{\varphi\varphi}$ is retained. Therefore $K_{\varphi\varphi}$ is symmetrical and $R\left(K^{(i-1)}_{\varphi\varphi}\right) = R\left(C^{(i-1)}_{\varphi\varphi}\right)$ is in force /$R(\cdot)$ - rank of the matrix/, which is the basic requirement in forming the criterion variance-covariance matrixes. Why is the matrical equality /18/ approximate? The diagonal elements of $K^{(1)}_{\varphi\varphi}$ correspond exactly to the given dispersions, when the nondiagonal elements is determined with the formula

/19/ $$\hbar_{ij} = r_{ij} \left(\hbar_{ii}\, \hbar_{jj}\right)^{\frac{1}{2}}, \qquad i, j = 1, \ldots, \kappa$$

Since the correlation coefficients r_{ij} are standartised variance-covariance moments, they do not depend on the accuracy of measurements if the measurements are with uniform accuracy. But when $C^{(o)}_{\ell\ell} \neq E_n$ / E_n - unit matrix of order $n \times n$ /, there is remaining influence on $C^{(o)}_{\ell\ell}$ which is involved in the calculation of $R^{(o)}_{\varphi\varphi}$. Therefore the determination of $K_{\varphi\varphi}$ has to be in iterative calculation process. After changing the indexes (o) with (i) and (1) with (i+1), in /13/ ÷ /18/ can be defined the respective iterative process. In [VAGIN, 1979] the congruity of the process of forming $K^{(i)}_{\varphi\varphi}$ in the levelling network

is proved, but under given conditions. Without further conditions in angular-distant network the formation process can be non-congruous. As a mark for congruity of the process in optimizational problem for the weight of the measurements, the inequality

/20/ $\Delta p_i^* \Delta p_i < \Delta p_{i-1}^* \Delta p_{i-1}$

can be used, where $\Delta p_i = \mu^2 \operatorname{vec}\left[\left(C_{\ell\ell}^{(i)}\right)^{-1} - \left(C_{\ell\ell}^{(i-1)}\right)^{-1}\right]$, $C_{\ell\ell}^{(i-1)}$ and $C_{\ell\ell}^{(i)}$ are variance-covariance matrixes of measurement in two arranged iterations. When the inequality /20/ is broken, it is the pointer of noncongruity, resulting from contradictions between the requirements in the project and the configuration of the network.

Therefore when in the network the elements of error ellipses are given, during the formation of criterion matrix the next diagonal elements $k_{2i-1,2i}, i = 1, \ldots, \frac{\kappa}{2}$ will receive different values - one with the third formula from /12/ multiplied by μ^2 and second times from /18/. If the values determined through /12/ are restored in after finding from /18/, this means complete consideration of the requirements and partial neglecting of the influence of the configuration. Often it is necessary to be given only 2 parameters of the error ellipses. In that case there can be more complete correspondence between the results from formula /12/ (multiplied with μ^2 and those from /18/ for the respective next diagonal coefficients. For this purpose is used correlation coefficients $r_{2i-1,2i}, i = 1, \ldots, \frac{\kappa}{2}$ determined from /15/ and only elements $k_{2i-1,2i-1}$, $k_{2i,2i}$ are determined. Let us examine concrete examples.

1) Given are $\bar{R}_{max}$, $\bar{R}_{min}$

For determining $\bar{q}_{11}$, $\bar{q}_{22}$, $\bar{q}_{12}$, respectively k_{11}, k_{22}, k_{12} with formula /12/ (the first two orthogonal functions in $C_{\varphi\varphi}$ are necessary $\bar{a}$, $\bar{b}$ and $\bar{\psi}$. Quantities $\bar{a}$ and $\bar{b}$ can be calculated directly from /11/. For getting ψ, with known correlation coefficient r_{12} from $R_{\varphi\varphi}^{(o)}$, we can use /19/ and /12/. After modifications

we have

/21/ $$\cos^2 2\overline{\psi} = \frac{\frac{\bar{a}^2}{\bar{b}^2} r_{12}^2 - 1}{r_{12}^2 - 1}$$

Which is in force only when $\frac{R_{max}^2}{R_{min}^2} > \frac{1 + |r_{12}|}{1 - |r_{12}|}$

2) Given are the bearing of a given direction $\overline{\Theta}$ and the admissible value of the mean square error in this direction $\overline{m}_{\overline{\Theta}}$. Here again k_{11} and k_{22} are being sought with determined matrix $R_{\varphi\varphi}^{(o)}$. This problem can be solved with changing proportionally the size of the error ellipse with given value of the dispersion for unit weight. From the first variant of the project for dispersion in direction $\overline{\Theta}$ can be determined

/22/ $$m_{\overline{\Theta}}^2 = R_{max}^2 \cos^2(\psi - \overline{\Theta}) + R_{min}^2 \sin^2(\psi - \overline{\Theta})$$

If the relation $\eta = \frac{\overline{m}_{\overline{\Theta}}^2}{m_{\overline{\Theta}}^2}$ is calculated, for k_{11} and k_{22} will be in force

/23/ $$k_{11} = \eta\ C_{11}$$
$$k_{22} = \eta\ C_{22}$$

The proof that /23/ ensures getting $\overline{m}_{\overline{\Theta}}$ comes from the immediate substituting /23/ in the first 2 equations of /10/ and considering /19/. The conventional symbols in 1), 2), /12/ are for the first two functions. Generalization comes immediately after exchanging index 1 with 2i-l and 2 with 2i.

<u>One Solution of Matrical Equation $B K_{xx} B^* = K_{\varphi\varphi}$</u>

After leaving $K_{\varphi\varphi}$ during the greater part of optimizational problem the determination of K_{xx} is necessary. Direct solution to the matrical equation

/24/ $$\underset{\kappa, m}{B}\ \underset{m, m}{K_{xx}}\ \underset{m, \kappa}{B^*} = \underset{\kappa, \kappa}{K_{\varphi\varphi}}$$

in relation with K_{xx} is possible only when $R(B) = m$. Therefore there exists $B^+ = (B^* B)^{-1} B^*$, so that

/25/ $\mathrm{K}_{xx} = \mathrm{B}^{+}\,\mathrm{K}_{\varphi\varphi}\,(\mathrm{B}^{+})^{*}$

When a given network has multipurpose, the above condition is favourable and can be satisfied. More complicated case is when the network has special purpose - when the exact determination number of functions must satisfy the given conditions and $\mathrm{R}(\mathrm{B}) < m$. The problem in this case has infinite solutions and simplicity can be received only with more conditions. If we use Kronecker product, /24/ can be written in this form

/26/ $$(\mathrm{B} \otimes \mathrm{B})\mathcal{y} = \underset{\mathrm{t}_1, n_1}{\mathrm{G}} \quad \underset{n_1, 1}{\mathcal{y}} = \underset{\mathrm{t}_1, 1}{\mathcal{q}}$$

where $\mathrm{t} = (\kappa^2 + \kappa)/2$, $n_1 = (m^2 + m)/2$, and for $\mathcal{y}$ and $\mathcal{q}$ are in force

/27/
$$\mathcal{y} = \mathrm{vec}(\mathrm{K}_{xx}) = \left[\mathcal{k}_{x_1x_1}, \mathcal{k}_{x_1x_2}, \ldots, \mathcal{k}_{x_1x_m}, \mathcal{k}_{x_2x_2}, \ldots, \mathcal{k}_{x_2x_m}, \ldots, \mathcal{k}_{x_mx_m}\right]$$
$$\mathcal{q} = \mathrm{vec}\left(\mathrm{K}_{\varphi\varphi}\right) = \left[\mathcal{k}_{\varphi_1\varphi_1}, \mathcal{k}_{\varphi_1\varphi_2}, \ldots, \mathcal{k}_{\varphi_1\varphi_\kappa}, \mathcal{k}_{\varphi_2\varphi_2}, \ldots, \mathcal{k}_{\varphi_2\varphi_\kappa}, \ldots, \mathcal{k}_{\varphi_\kappa\varphi_\kappa}\right]$$

Since $t_1 < n_1$ $/\kappa < m/$ for the solution of /26/, analogy with conditional adjustments can be done. If it is applied condition $\mathcal{y}^{*}\,\mathcal{y} = \min$, a solution for $\mathcal{y}$ is received. The specialized form of Kronecker product for symmetrical matrix ensures solution in which K_{xx} is symmetrical. The condition $\mathcal{y}^{*}\,\mathcal{y} = \min$ means that the solution to /26/ is being sought in which the module of vector with element from K_{xx} is minimized. Since no additional condition for positive determination of K_{xx} is given, there is no guarantee that $\det(\mathrm{K}_{xx}) > 0$. This solution is examined in [SCHMITT, 1980], when it is assumed that has no zero columns, that is in the functions are involved all the points in the network. Irrespective of the positive determinations, however it can be pointed out that "The criterion-matrix should be not-negative definite". This assertion in the common case is unfounded, since all the elements of K_{xx} were examined equally in the solution.

For the solution of /25/ it is applied here the following way: Let

in /25/ in place of K_{xx} and $K_{\varphi\varphi}$ are replaced variance-covariance matrix C_{xx} and $C_{\varphi\varphi}$. If we symbolise $\underset{n_1,1}{u} = \mathrm{vec}(C_{xx})$, $\underset{t_1,1}{s} = \mathrm{vec}(C_{\varphi\varphi})$ through analogy with /26/ it can be written:

/26a/ $G\,u = s$

Let us form the difference between /26/ and /26a/

/28/ $G\,(y - u) = q - s$

or

/28a/ $\underset{t_1,n_1}{G}\ \underset{n_1,1}{z} = \underset{t_1,1}{w}$

where $z = y - u$ contains digital values of the changing, which must be in C_{xx} to be able to get K_{xx}, and $w = q - s$ is the known vector of the exchange between $C_{\varphi\varphi}$ and $K_{\varphi\varphi}$. If the condition $z^* z = \min$ is applied, it is sought in this way solution in which the exchange in C_{xx} is minimized. Therefore without limits for the rank of G will be in force

/29/ $z = G^{+}\,w\,,$

and when $R\,(G) = t_1$

/29a/ $z = G^{*}\,(G\,G^{*})^{-1}\,w$

where with G^{+} symbolises the generalized rnatrix inverses. After getting z is determined

$$y = z + u\ ,\quad K_{xx} = \begin{bmatrix} y_1, & y_2, & \dots, & y_m \\ y_2, & y_{m+1}, & \dots, & y_{2m-1} \\ \vdots & \vdots & \vdots & \vdots \\ y_m, & y_{2m-1}, & \dots, & y_{n_1} \end{bmatrix}$$

Despite the solution, the direct solution /26/ is more stable under the condition $y^* y = \min$, under greater changes in $C_{\varphi\varphi}$ for finding $K_{\varphi\varphi}$ is possible getting the negative determined matrix K_{xx}, which is the pointer of the contradictions between the requirements of the accuracy of the project and the given configuration.

Examples. Example 1: In fig. l is given angular-distant network, in which all directions /$m_{\mathrm{R}} = 10^{\mathrm{cc}}$/, all distances /$m_{s_{ij}} = 2s_m$/ are measured. Points with numbers 10, 20, 30 and 40 were given. The criterion matrix was formed under $\mathrm{B} \equiv \mathrm{E}_m$ for the mean square errors of the unknowns, so that the maximum and minimum radius of the point error ellipses will be very near $-\mathrm{R}_{\max_i} = 28,3$ mm / $m_{s_i} = \left(\mathrm{R}^2_{\max_i} + \mathrm{R}^2_{\min_i}\right)^{\frac{1}{2}} =$ 40 mm /. Under the inequality $\frac{\mathrm{R}^2_{\max}}{\mathrm{R}^2_{\min}} \geq \frac{1+|r_{i\,i+1}|}{1-|r_{i\,i+1}|}$ were determined respectively:
for points 1 and 3 $-|r_{1\,2}| = |r_{5\,6}| =$ 0.028 $\rightarrow \mathrm{R}_{\min} =$ 27.5;
for points 2 and 5 $-|r_{3\,4}| = |r_{9\,10}| =$ 0.116 $\rightarrow \mathrm{R}_{\min} =$ 25.1;
and for point 4 $-|r_{7\,8}| =$ 0 $\rightarrow \mathrm{R}_{\max} = \mathrm{R}_{\min}$. For comparing the initially determined matrix $\mathrm{Q}^{(o)}_{xx}$ and $\frac{1}{\mu^2}\,\mathrm{K}^{(1)}_{xx}$, the elements of $\mathrm{Q}^{(o)}_{xx}$ were multiplied by the quality $\frac{1}{\mu^2}\,\mathrm{tr}\left(\mathrm{K}^{(1)}_{xx}\right)/\mathrm{tr}\left(\mathrm{Q}^{(o)}_{xx}\right)$, where $\mathrm{tr}(.)$ is the trace of the matrix. In table 1 were given the results of the solution, while only the elements above the diagonal were written. For every element of the matrix $\mathrm{Q}^{(o)}_{xx}$ $\mathrm{K}^{(1)}_{xx}$ were given 2 values - the first is the dimensional changing of $\mathrm{Q}^{(o)}_{xx}$ and the second - the respective element from $\frac{1}{\mu^2}\,\mathrm{K}^{(1)}_{xx}$.

Example 2: Let $\mathrm{C}_{xx} = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$, $\mathrm{B} = [1 \quad -1]$, then $\mathrm{C}_{\varphi\varphi} = \mathrm{B}\,\mathrm{C}_{xx}\,\mathrm{B}^* = [6]$.
For determining K_{xx} let $\mathrm{K}_{\varphi\varphi} = q = [3$ is given. Taking into consideration /26/ will be in force $\mathrm{G} = (\mathrm{B} \otimes \mathrm{B}) = [1 \quad -2 \quad 1]$, respective equation will be $[1 \quad -2 \quad 1]\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = [3]$. After the application of condition $y^* y = \min$ were determined $y_1 =$ 0.5; $y_2 =$ -1; $y_3 =$ 0.5, from where for K_{xx} we have $\mathrm{K}'_{xx} = \begin{bmatrix} 0.5 & -1 \\ -1 & 0.5 \end{bmatrix}$ and $\det(\mathrm{K}'_{xx}) = -0.75 < 0$, which is impossible.
Now let us use /28/. There will be in force $\mathrm{u}^* = \mathrm{vec}\,(\mathrm{C}_{xx})^* = [4 \quad 1 \quad 4]$, $\mathrm{w} = \mathrm{vec}\left(\mathrm{K}_{\varphi\varphi}\right) - \mathrm{vec}\left(\mathrm{C}_{\varphi\varphi}\right) = q - \mathrm{s} = [-3]$ and for /28a/ we have $[1 \quad -2 \quad 1]\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = [-3]$.
After applying $\mathrm{z}^*\mathrm{z} = \min$ $z_1 =$ -0.5; $z_2 =$ 1; $z_3 =$ -0.5 from where $y = \mathrm{z} + \mathrm{u} = \begin{bmatrix} 3.5 \\ 2.0 \\ 3.5 \end{bmatrix}$, $\mathrm{K}''_{xx} = \begin{bmatrix} 3.5 & 2 \\ 2 & 3.5 \end{bmatrix}$ and $det\left(K''_{xx}\right) = 33/4$.
It is clear that $\mathrm{B}\,\mathrm{K}'_{xx}\,\mathrm{B}^* = \mathrm{B}\,\mathrm{K}''_{xx}\,\mathrm{B}^* = [3]$ and $\mathrm{R}(\mathrm{C}_{xx}) = \mathrm{R}(\mathrm{K}'_{xx}) = \mathrm{R}(\mathrm{K}''_{xx})$, but K'_{xx} cannot be examined as criterion matrix.
If the condition $\mathrm{K}_{xx} \geq [36]$ is applied in the project, then the

9.34	-0.214	4,60	1.32	5.52	0.787	6.81	-0.216	3.09	-1.19
7.83	-0.218	4.31	1.31	4.62	0.800	5.27	-0.207	2.90	-1.17
	6.27	0.101	1.89	-0.787	2.72	-0.617	2.60	-0.589	1.45
	7.74	0.114	2.26	-0.800	3.36	-0.580	3.03	0.670	1.73
		7.02	0.752	3.09	0.589	4.01	-0.385	1.83	-0.781
		7.36	0.828	2.90	0.670	4.17	-0.413	1.92	-0.859
			6.02	1.19	1.45	1.51	2.81	0.781	0.916
			6.95	1.17	1.73	1.38	3.16	0.859	1.06
				9.35	0.214	6.81	0.216	4.60	-1.32
				7.83	0.218	5.27	0.207	4.31	-1.31
					6.27	0.617	2.60	-0.101	1.89
					7.74	0.580	3.03	-0.114	2.26
						11.16	0	4.81	-1.51
						8.00	0	4.17	-1.38
							7.29	0.384	2.81
							8.00	0.413	3.16
								7.02	-0.752
								7.36	-0.828
									6.02
									6.95

$Q_{xx}^{(o)}$

$\frac{1}{\mu^2} K_{xx}^{(1)}$

$r_{12} =$ -0.028

$R_{max} =$ 28.3 mm

$R_{min} =$ 27.5 mm

$m_{s_1} =$ 39.5 mm

$r_{34} =$ 0.116

$R_{max} =$ 28.3 mm

$R_{min} =$ 25.1 mm

$m_{s_2} =$ 37.8 mm

$r_{56} =$ 0.028

$R_{max} =$ 28.3 mm

$R_{min} =$ 27.5 mm

$m_{s_3} =$ 39.5 mm

$r_{78} =$ 0

$R_{max} =$ 28.3 mm

$R_{min} =$ 28.3 mm

$m_{s_4} =$ 40.0 mm

$r_{9\,10} =$ -0.116

$m_{s_5} =$ 37.8 mm

TABLE 1

two solutions lead to $\det(K_{xx}) < 0$. Since B cannot be changed, the single possibility for getting a solution is changing the configuration, respectively matrix C_{xx}. For this the check for the sign of $\det(K_{xx})$ is indispensable.

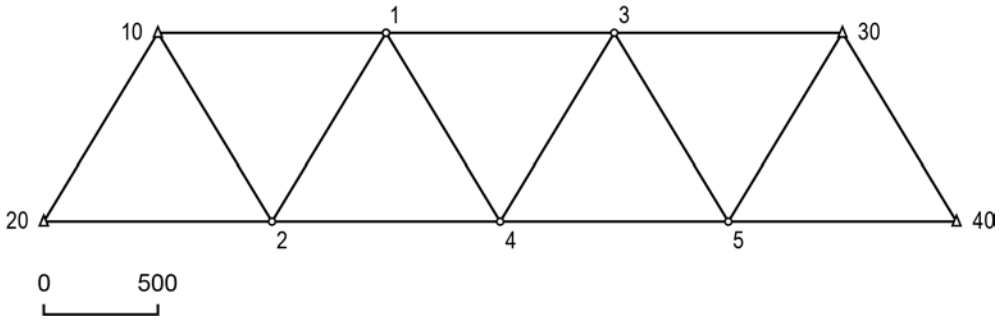


Figure 1

REFERENCES

ALBERDA, S.E. , 1973, Planning and Optimization of Networks. Same General Considerations, IAG Symposium, Oxford.

BAARDA, W., 1977, Measures for the Accuracy of Geodetic Networks, IAG Symposium, Delft.

BANOV, B., 1980, Optimization of the Weights of Observations in Geodetic Networks, FIG Symposium, Varna.

GRAFAREND et al., 1979, Optimierung Geodätischer Messoperationen, Karlsruhe.

RAO, C.R., 1968, Linear Statistical Inference and its Application, Nauka, Moskva.

SCHMITT, G., 1980, Second Order Design of a Free Distance Network, Considering Different Types of Criterion Matrices - Bull. geodésique, Vol. 54, No. 4, pp. 531-541.

VAGIN, V.A., 1979, Algorithm of Deriving Variance-Covariance Matrices of Levelling Networks, Izvestya VUZ, Geodesy & Aerial Photography, No. 1, p. 37-42.

ANALYSIS OF SOME DENSIFICATION NETWORKS IN ITALY

Maurizio BARBARELLA and Marco UNGUENDOLI
Bologna, Italy

ABSTRACT

This work considers the problem of densification networks with a view to large- and very large-scale map making.

Starting from a consideration of the standards imposed by the customers and on the basis of the results of work already undertaken, various problems relating to the measurements and to their analysis are investigated with the aim of deriving useful information for future work.

1. Introduction

In Italy at the moment, one of the greatest problems facing surveyors, geodesists, and cartographers is that raised by densification networks. Many regional authorities have in fact planned new large-scale technical map (1 :5,000) and in the case of city centres proposals have been made for the development of cartography on an even larger scale (1:1,000-1 :500).

It is a really sizable mass of work involving numerous private firms working in the fields of land survey and aerophotogrammetry. As regards the densification networks, one only has to remember that current standards relating to the 1:1,000 scale envisage the establishment of one point to every 1/2 kilometre square.

The need to employ several firms, even if on the one hand it speeds up work, on the other raises problems concerning the homogeneity of measurements that are taken with different instruments and working methods (traverses, triangulation, mixed networks) and often without adequate analysis both in the network planning phase and afterwards too, in relation to the existence of gross and systematic errors.

To these problems should be added others such as those relating to the adjustment of the networks, to the analysis of already existing networks, and to the incorporation of local networks in national ones. All these questions are wide open since the regulations laid down by the contractors leave the contracting firms a great deal of room to manoeuvre.

Before considering the problems mentioned, some regulations laid down for 1:1,000 cartography by the Emilia Romagna Regional Authority are cited. These may be taken as the springboard for several observations which we consider important and the confirmation of what has been argued above.

2. Regulations for the planimetric framing and densification network (Emilia Romagna Authority 1980)

As far as the planimetric framing network is concerned, the choice is a very straightforward one.

"The Geometric framing of the map will be derived from the national geodetic nets ..... It will consist of all the points of the 1st, 2nd and 3rd order of the I.G.M.I.'s (Istituto Geografico Militare Italiano) national network, of those of the 2nd, 3rd and 4th order of the I.I.M.(Istituto Idrografico Marina) and of the points of the cadastral network."

The general regulations for the densification nets may be summarized as follows (only the most important passages are quoted):

2-1) "The network must be planned as an intrinsically determined independent structure with surplus measurements"

2-2) "The scaling, the positioning and the orientation of the network must be carried out using the points of the national net ..... Not less than four points must be used, arranged in such a way as to enclose the entire survey ..... the density of the points must be in the order of 1 point to every half kilometre square."

2-3) "The measurements for densification networks must be performed using modern methods and instruments. The net must contain an excess number of measurements ..... Such excess measurements must make it possible to carry out valid statistical checks."

2-4) "The measurements will be acceptable if they satisfy the following tolerances:

Closed traverses

Linear misclosure $= \Delta = \sqrt{(\delta E)^2+(\delta N)^2} = \frac{\sqrt{\sum l^2}}{10{,}000} \cdot n$

when l is the length of the sides in metres and n is the number of the independent closed traverses.

Triangulation and mixed networks

The angular misclosure:

$\Delta\alpha \leq 20^{cc} \sqrt{n}$ when n is the number of the angles.

The maximum difference for two independent determinations of one side must be:

$\frac{\Delta l}{l} \leq 1/15{,}0000 = 0.07\%$

In the adjustment by minimum constraints of the networks, the mean square error on the coordinates must not exceed 0.4 me t res and 90% of them must be under 0.15 m."

2-5) "The calculations of the densification networks must be carried out in two stages. First, an adjustment by minimum constraints must be made assuming an arbitrary reference system; then must be carried out a two dimensional linear conformal transformation from the local to the national system. For both calculations the least squares method can be used. The densification network's scale variations must not exceed 0.5%."

As may be seen, the regulations proposed leave the contracting firms plenty of room for discretion both as regards the kind of net and instruments to be used, and also as regards adjustment methods, preferably with the exclusive use of least squares.

There are also certain inconsistencies such as the tolerances imposed on the coordinates of the network points which do not take into account the size of the net itself and which in practice do not indicate what origin should be assumed for the adjustment with minimum constraints. Further inconsistencies arise from the consideration of the existing network's points as sufficiently precise, i.e. as being invariant and therefore as not being in need in any case of ad hoc measurement checks. *

On the basis of what has been outlined above and without going into the prior planning and optimization of the nets, which would not correspond to actual working reality, both because a lot of the field-work has already been carried out, and because land surveying firms follow their own criteria and have difficulty accepting even slightly complex mathematic criteria, it seems to us that under the present circumstances there are three problems that must and can be come to grips with concretely.

The first problem has to do with a rigorous analysis of existing networks in order to evaluate their capability to be assumed as a basis for local medium-to-high accuracy nets. The second concerns the analysis of the nets that have been set up, to evaluate whether the tolerances imposed have been respected and to gain some idea of the quality of the measurements, of the extent of gross errors, and of the homogeneity of the networks carried out by a variety of firms and according to different working patterns etc. The third problem concerns the incorporation of the local networks in the national ones.

* On the basis of the experience and the studies carried out, the above specification will be revised in the near future.

3. Checking the existing networks

As already mentioned there are three networks that may be used for the framing of the local nets: I.G.M.'s national one (points of the 1st, 2nd, and 3rd order), the cadastral one, and the I.I.M. one (points of the 2nd, 3rd and 4th order). It seems to us that the correct approach involves evaluating each of the various networks separately and then seeing the degree of homogeneity and consistency existing among them.

3-1. The I.G.M. national network

As is well known, the I.G.M.'s Italian network consists of four orders of points of decreasing accuracy. As has already been said, the points of the 4th order are not taken into consideration whereas those of the other three orders are considered valid. In the case of the first order net, a new adjustment has been made and the coordinates of the points (E.D.79) have been recalculated, but unfortunately the coordinates that have been determined thereby have not yet been made available to the public. However, the coordinates obtained with the previous adjustment should also be held valid, as has often been verified, even on a personal basis, with distance measuring or the repetition of angular measurements. As far as the points of the 2nd and 3rd orders are concerned on the other hand, some doubts still linger, not so much as to the accuracy of the measurements, as to the semi-empirical methods employed in the adjustment made for extremely small blocks. Luckily, almost all the data relating to angular measurements taken towards the end of the century are kept in the I.G.M. archives and a new adjustment for large blocks would therefore be possible and is very much to be desired.

Studies have already been carried out on such networks (ARCA, 1979) and the results obtained seem to indicate the suitability of the measurements to provide results that might be used in the framing of local nets.

Within the context of a new adjustment, it would also be necessary to clean the measurements taken, which are far in excess of the numbers required, by using techniques based on statistical tests such as the "data snooping" one.

While waiting for this new adjustment and in any case until it is made,

we think that only the first order points should be regarded as fixed points, whereas in the case of the points of other orders, at the very least a check using ad hoc measurements ought to be carried out.

3-2 Cadastral Networks

In Italy, the Cadastre has constructed triangulation nets and traverses, using the points of the I.G.M.'s 1st, 2nd and 3rd orders for the drawing up of its Cassini-Soldner cartography. These nets therefore suffer, right from the start, from the errors of the already cited nets, to which must be added the same remarks concerning the empirical adjustment methods adopted. It should also be noted that in the transformation of the coordinates from the Cassini-Soldner system to the I.G.M. one of Gauss-Boaga, a phase that is nearing completion, approximate formulas are being used. In this case too however, there is an almost complete set of field-data which means that, providing the data themselves have been cleared from gross errors, a new adjustment in large blocks should be possible: and this adjustment ought to follow on from that of the I.G.M. nets, the results of which should be taken as the starting point.

These conclusions were reached following the analysis carried out by the present authors, as well by other researchers, of a lot of cadastral evidence regarding sufficiently extensive nets for which there exists archival data.

In one of these studies (MONTI, MUSSIO, 1979) the entire network of an Italian province (Parma) has been adjusted in block, taking in more that 300 points including 27 that belong to the I.G.M. nets and are considered as fixed. A lot of information was derived from this adjustment (BARBARELLA, FOLLONI, 1979), of which we shall cite that piece concerning the mean square error of the position of point $\sqrt{\sigma_N^2+\sigma_E^2}$, the distribution of which (fig. 1a) has a mode of 0.75 m and a mean of 0.90 m with about 4% of its values greater than twice the mean value. Another interesting piece of information emerges from the differences between the point positions obtained with the overall adjustment and those of the catalogue, obtained using, as has already been stated,

semi-empirical calculation methods. These differences (fig. 1b) have a mode of about 0.35 m and a mean of 0.80 m with more than 19% of values over 1.5 m.

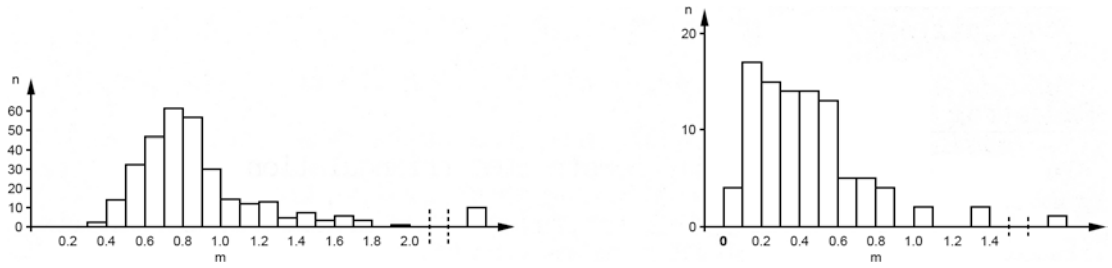


Fig. 1 - Cadastral Network (A) of the province of Parma. a) Distribution $\sqrt{\sigma_x^2+\sigma_y^2}$. b) Distribution of the differences in the position of the points between the new adjustment and the Catalogue.

In a subsequent study (BARBARELLA, 1981) the problem was taken up again with the analysis of another block of the cadastral triangulation net, including more than 100 net and sub-net points, six of which belong to the I.G.M.. Of this block, above all the measurement data were analysed by applying the Data Snooping method with the prior assumption of a reasonable m.s.e. for angle measurements (10^{cc}) on the basis of the instruments employed.

This analysis yielded good data consistency which made it advisable not to eliminate any of the data. To a certain extent this was predictable, in so far as the analysis didn't involve the entire set of field-data but only that part that had been used for the original calculation and which had already been, albeit in an empirical way, cleared from gross errors.

The data that had been checked in this way, have been analysed giving interesting results.

On the basis of these results the fundamental networks (fig. 2) was reexamined, while carrying out two adjustments.

In the first of these adjustments with the coordinates of I.G.M.I. points as constraints the old coordinates of the said points obtained through the empirical adjustment were introduced; in the second, on the other hand, the coordinates obtained by a block adjustment carried out recently by the I.G.M. itself were introduced.

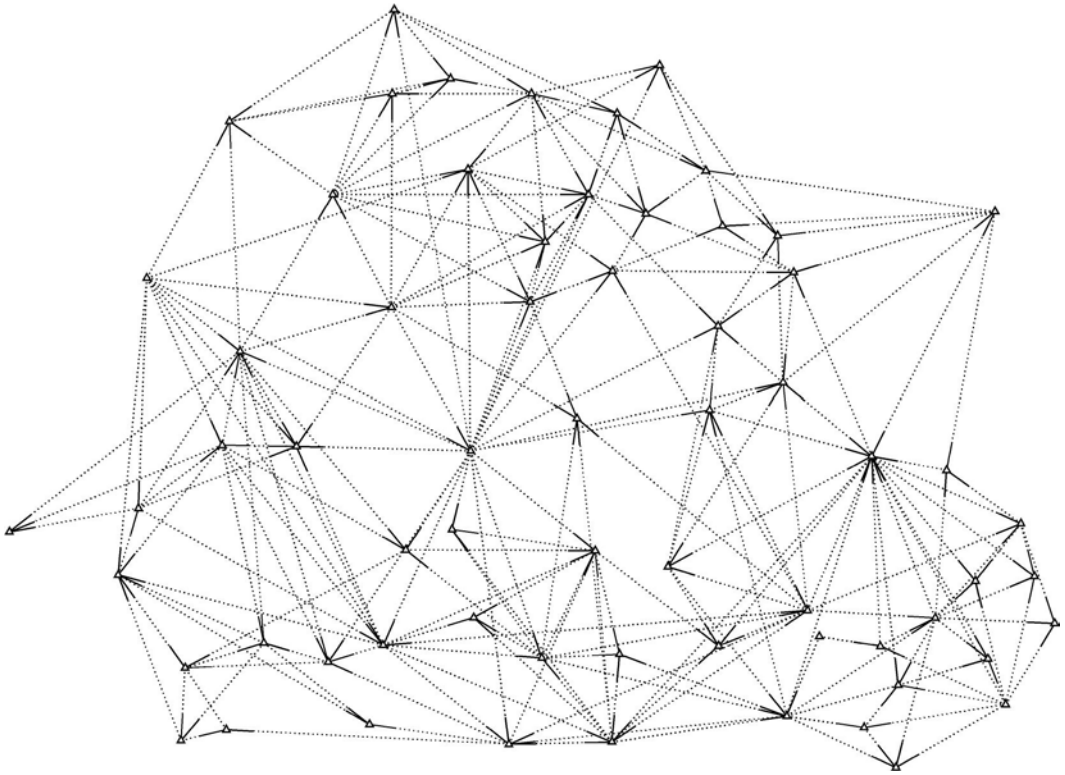


Fig. 2 - Cadastral Network (B) in the province of Bologna.

In our opinion the most interesting piece of information obtained is the clear-cut difference between the m.s.e. of unit weight which in the first case is 2.6 and in the second, 1.8. Furthermore, as may be observed from fig. 3, the values of the major semi-axes of the Standard ellipses differ significantly.

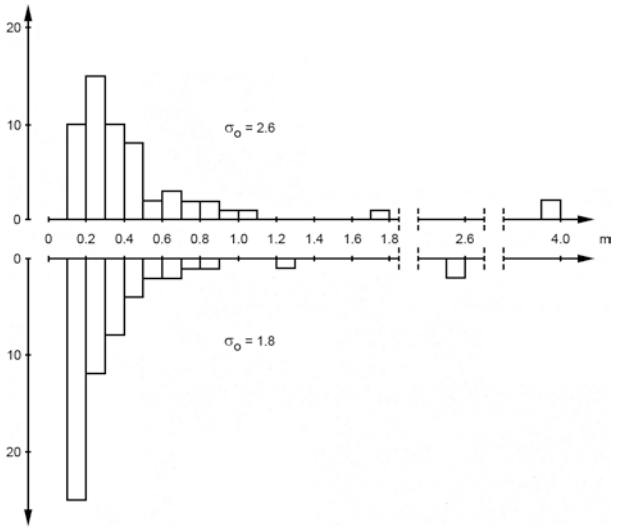


Fig. 3 - Distribution of the values of the major semi-axes of the standard ellipses for the two adjustments of net (B).

Another interesting result is represented by the fact that the introduction of the constraints represented by the newly determined I.G.M. points does not significantly alter the m.s.e. of unit weight which may be obtained using minimum constraints (from 1.7 to 1.8). In our opinion this means that the measurements of the two authorities are mutually consistent and of comparable accuracy.

This result backs up the idea that has already been expressed, i.e. that a new overall cascade adjustment needs to be made; first of all I.G.M. should readjust its 2nd and 3rd order networks and then the Cadastre should do the same to its network, while basing itself on the previously obtained points.

4. The I.I.M. Network

For its own institutional purposes, the I.I.M. has in the post-War period carried out a complete triangulation of a strip of coast, basing itself on the 1st order points of the I.G.M. national network, and laying down fairly precise working regulations. In fact, the I.I.M. envisages use of the Wild T3 with 12 observations for connections with the I.G.M. network, Wild T3 with 6 observations for the principal chain made up of quadrilaterals with sides of about 10 km, and Wild T2 with 3 observations for secondary triangulations. The tolerances accepted for triangle closure error are as follows: 3" for the first measurements, 6" for the second, and 12" for the final ones (I.I.M. 1966). Given these standards and the fact that the I.I.M. net covers a very thin strip of coast, and also because the reliance only on first series I.G.M. points affords a good guarantee for the said network, through analyses have not been made.

5. New nets

As has already been said, densification networks have two precise uses: for the construction of technical maps on the scale of 1:10,000 - 1:5,000 and the construction of special maps on the scale of 1:2,000 - 1:500.

For the first type of network, normal practice basically involves using mixed networks (angles and distances) formed by small adjacent blocks which may be linked up to form nets of quite considerable dimensions, given that

technical cartography is projected on a regional basis. In the case of the second type of map on the other hand, in which, at least for the time being, the survey of particular zones is involved, traverses are often used.

As an example of this kind of work, five blocks of a mixed network forming a single net (fig. 4) and two traverse networks, one of which is rather extensive (figs. 5 and 6) have been analyzed. The five blocks and the two networks of traverses have been made by seven different companies.

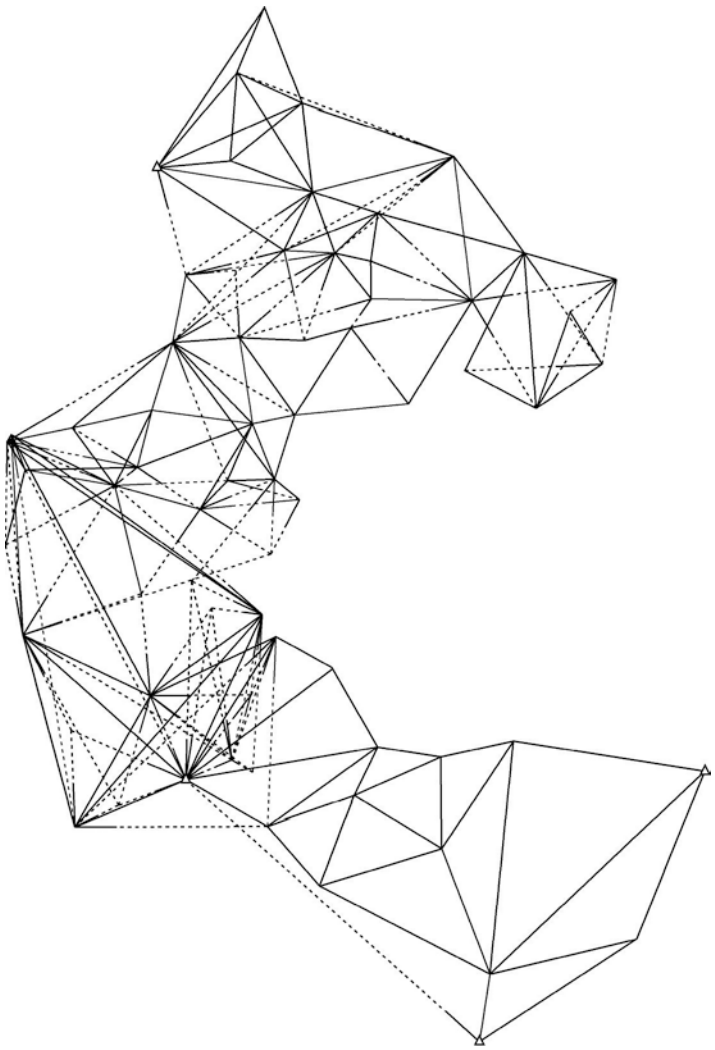


Fig. 4 - Network (C) for 1:5,000 cartography.

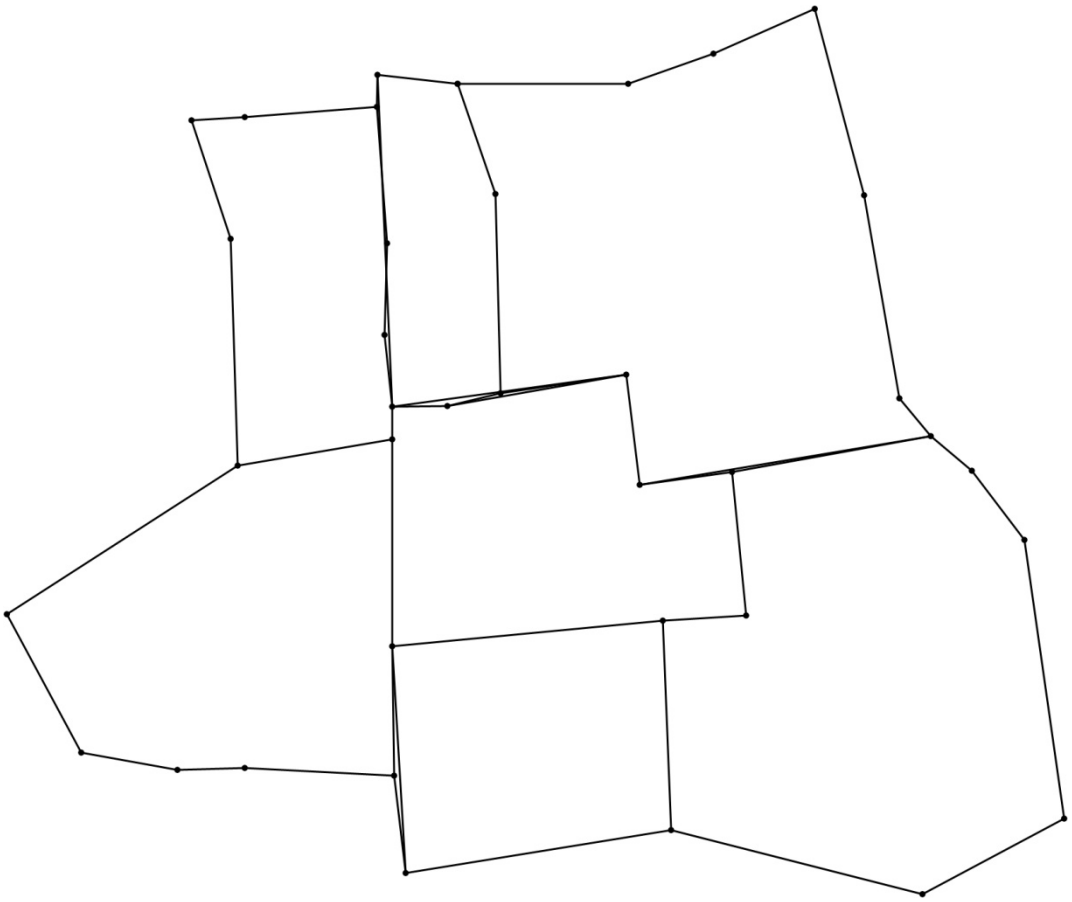


Fig. 5 - Traverse (D) of an urban area for 1:5,000 cartography.

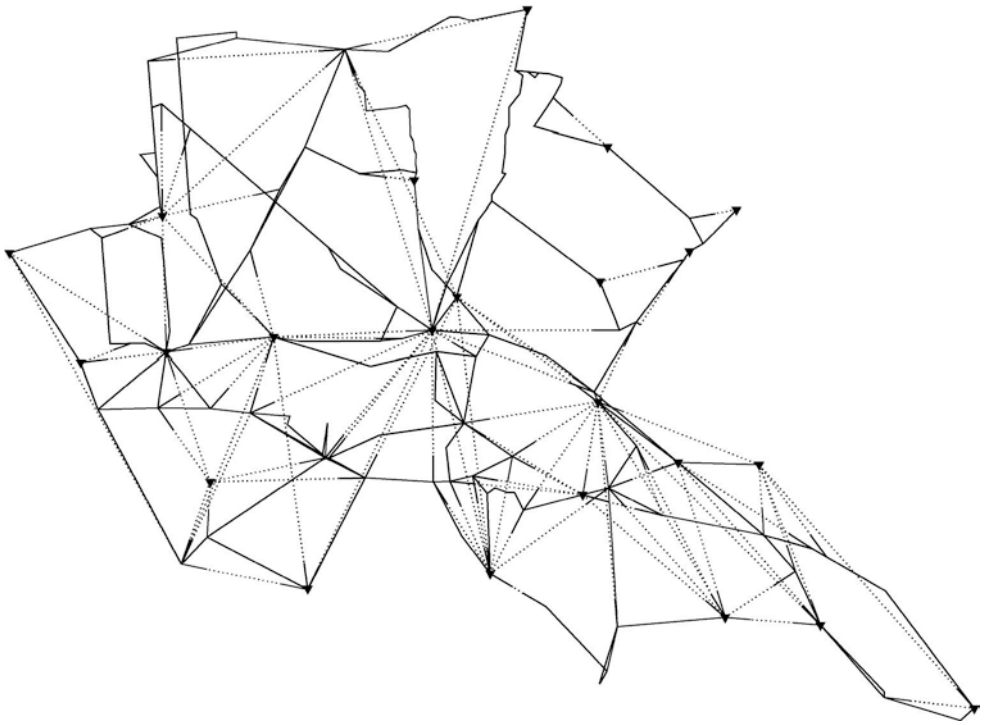


Fig. 6 - Traverse (E) for 1:2,000 cartography.

5a. Mixed nets

The analysis of the blocks was carried out following the classical Data Snooping method and using the free net adjustment. The data relating to the blocks and to the results obtained are set forth in the following table I.

TAB. I

Block		I	II	III	IV	V
Number of points	n.	14	19	22	21	8
Number of observations	Angles	37	73	86	65	28
	Distances	28	21	33	29	12
	N = A + D	65	94	119	94	40
	N / (2n - 3)	2.6	2.7	2.9	2.4	3.1
Length of the sides (average) km		7	9	4	5	5
Instrumentation	Theodolites	Wild T2, Kern DKM 2A, ecc. ...				
	E.D.M.	CA1000	CA1000 AGA 14	CA1000	AGA 14	HP3820 AGA 14
A priori M.S.E.	σ_α	6^{cc}				
	σ_l	(10 + 5.1) mm				
Preliminary free-net adjustment σ_o		0.9	2.3	4.9	7.4	3.1
Data Snooping $\lvert w\rvert \geq 4$						
Rejected measurements	Angles	0	4	13	5	1
	Distances	0	0	7	3	0
Free-net adjustment	σ_o	0.9	1.9	2.0	1.7	1.3
Data Snooping $\lvert w\rvert \geq 3.29$						
Rejected measurements (Total number)	Angles	0	11	18	9	2
	Distances	0	1	11	3	0
Free-net adjustment	σ_o	0.9	1.4	1.5	1.4	1.1
$\sigma_P = \sqrt{\frac{1}{n}\sum(\sigma_x^2+\sigma_y^2)}$ cm		4	9	13	19	3

As may be seen, the different blocks are fairly varied both in terms of precision and in terms of the number of outliers present. In fact, the σ_0 obtained by taking all the measurements into consideration, has a variation range from 0.9 to 7.4, whereas the percentage of measurements rejected using the data snooping technique varies between zero and 24 per cent. In this way the blocks become much more homogeneous and in fact the σ_0 varies between 0.9 and 1.5. There remains nevertheless a wide range in the average positional standard error σ_P from 3 to 19 cm.

An adjustment was subsequently made of the entire network, taking into account linking measurements between the various blocks. In this way a net containing 70 points with 292 angles and 114 distances with a redundancy of 2.9 is obtained. As regards the relative weights to be attributed to the measurements made by the various companies, it was decided to consider the σ_0 values obtained in the adjustment of single blocks. Using the data snooping method and still with $|w| \geq 3.29$, 10 further angles and 4 distances were eliminated, and the final adjustment yields $\sigma_0 = 1.3$ and $\sigma_P = 17$ cm for the entire network, whereas the distribution of the major semi-axes of the standard ellipses (fig. 7) has a mean of 14 cm and the 70% of the values lower than 15 cm.

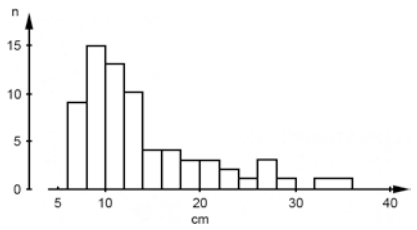


Fig. 7 - Distribution of the major semi-axes of the standard ellipses for the network (C).

5b. Traverses

As regards the traverses, as has been stated already, two very different ones, in terms of their purpose, size, and the length of their sides, were examined. The first was performed in a small built-up area for a 1:500 scale mapping and it followed centre lines; the second, used for a 1:2,000 scale mapping, covers on the other hand an area of about 15 x 20 km with many links, even at a long distance.

The data relating to the traverses and the results of the analysis have been set out in table II.

TAB. II

TRAVERSE		D	E
Number of points		39	195
Number of observations	Angles	71	385
	Distances	54	230
	N = A + D	125	615
	N / (2n - 3)	1.7	1.5
Length of the sides (average) m.		270	960
Instrumentation	Theodolites	Kern DKM 2A	Wild T2
	E.D.M.	Kern DM500	AGA 12-14
A priori M.S.E.	σ_α	15^{cc}	15^{cc}
	σ_l	10 mm	(10 + 5.1) mm
Data Snooping $\|w\| \geq 3.29$			
Rejected measurements	Angles	1	22
	Distances	1	4
Free net adjustment	σ_o	1.2	1.2
	σ_P cm	1.5	5.4

As can be seen, the results are rather good and certainly adequate for the purpose the traverses were designed for. Fig. 8 shows the distribution of the values of the major semi-axes of error ellipses for the 195 points of the larger traverse and this reveals the network's good consistency. The mean value is about 4.5 cm and only 7 points have a major semi-axis that is more than double the mean value.

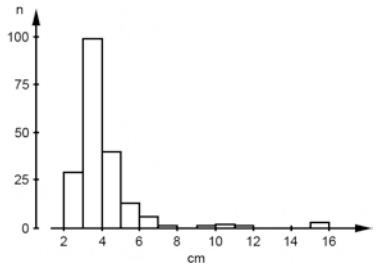


Fig. 8 - Distribution of the major semi-axes of the standard ellipses for the network (E).

6. Connection of local net with national ones

There are basically two methods to be followed for the connection of local networks with national ones.

With reference once again to our region (Emilia-Romagna) the regulation stipulate that the surveys for the Technical Map on the scale of 1:5,000 should be adjusted directly on the 1st, 2nd, and 3rd order points of I.G.M. net and on the cadastral net's points present in the same zone, envisaging, according to the judgement of the tester, the checking of 20% of the points with direct measurements which ought to yield discrepancies less than 30 cm.

In the case of the larger scales on the other hand, the regulations envisage the adjustment of the new networks together with the trigonometric points of the said orders in a single minimum constraints block and subsequently a similarity transformation taking into account all the trigonometric existing points

The first method, i.e. that of considering the trigonometric points fixed and without errors might usefully be employed should there be a reasonable certainty that the pre-existing networks possess intrinsic precision comparable with, if not greater than, that of local nets. This however, as has been pointed out in section 3 is not always the case and it does not seem to us that checking 20% of the points would get around this drawback.

On the other hand it would be unthinkable to saddle the contracting firms with the task of carrying out an overall check of the said networks. Therefore, as has already been said, it is necessary to perform an overall readjustment of the I.G.M. and the cadastral networks from which it will be possible to derive the necessary parameters for an evaluation of the suitability or unsuitability of assuming the various points as fixed points for the incorporation of the local nets.

The decision to incorporate the local nets for very large-scale mapping by means of a similarity transformation is justified by three objective facts:

a) the precision required for these networks is necessarily higher than that of national networks.

b) Normally these surveys concern very limited areas and so are rarely linked to one another and hence there is no need for a rigid connection with the national networks.

c) The method proposed makes it possible, without involving direct checking, to examine, on the basis of the residues, the consistency of the control points with the possibility of eliminating those that fail to give adequate guarantee.

In so far as rigid incorporation is concerned, there is no point repeating what has already been said.

In the case of the similarity transformation on the other hand, the traverse consisting of 195 points including 22 belonging to national networks, has been reexamined. After a first similarity transformation it was observed that several points had undergone a variation in position in the order of 50-60 cm whereas, as has already been said, the σ_P of the networks is 4.5 cm. The operation was then repeated without taking into account four points and a clear improvement in the results was observed. In

fact, the largest shifts in position do not exceed 35 cm, while the scale factor falls from 1.000020 to 1.000008. The slight congruence of the national network points is confirmed by an adjustment constrained to all the 22 points which yield σ_0 = 4.3 whereas as was seen, in the case of the free adjustment σ_0 = 1.2 and this demonstrates clearly enough that the net's intrinsic precision is decidedly superior to that of the national nets. On the other hand, taking only 18 points as fixed, one obtains σ_0 = 3.0 which confirms us in the decision taken to eliminate the 4 points even if there is still a worsening in comparison to the free adjustment.

7. Conclusions

The analyses carried out prompt several comments and the formulation of a few suggestions that we consider are not altogether lacking in interest.

The first point is that the measurements of the existing I.G.M.I. and Cadastral networks as well as the new one appear rather good, and that the lack of precision encountered on occasion with the coordinates of the points is due to insufficient adjustment.

A rigorous block adjustment already begun by the I.G.M.I. of its own high order nets will make it possible to obtain a sufficiently precise datum, we believe, to allow for the incorporation of the densification networks of the medium scale 1:5,000, 1:10,000 of the technical cartography, without introducing unbearable deformations into the said nets.

In the case of larger scales, 1:2,000 and 1:500, incorporation can not be carried out while keeping the trigonometric points fixed: the similarity transformation appears to be the simplest method, and able to provide some kind of check on the accuracy of the coordinates of the points of the framing net.

The readjustment of national networks also supplies point position errors, which might therefore enable one to use more refined fitting techniques in which the errors of the coordinates of both the local net (which in any case must be adjusted with minimum constraints) and the primary one, would be taken into consideration.

In the case of less precise nets that already been constructed and incorporated into existing national networks, it may be impossible or unsuitable to carry out a new adjustment based on the new coordinates; perhaps it would be advisable to consider linear or non-linear interpolation methods so as to obtain the coordinates of the new nets.

As regards the new networks, greater care would be welcomed in the planning phase as well as a centralization, if not of the field work, then of the phase involving the adjustment and fitting local networks in the national ones, with greater care taken also over the linking measurements between one block and another.

In any case it would be advisable that the regulations prescribe the analysis of the accuracy of the new nets on the basis of relative error ellipses instead of the absolute one which ore obvious - significantly effected by the reference system assumed.

Given the good degree of accuracy of the new networks, a classification and cataloguing of their components should be considered indispensable so as not to allow such an amount of work to be dispersed, and also with a view of a revision of the cartography and to enable different uses.

Appendix

The following gives some information about the programs used in the analysis of the networks.

The programs are written in FORTRAN IV, the computer used were a Digital Vax 780 and a CDC Cyber 76; the graphs were made on a Calcomp 936.

The procedure followed for the adjustment and the analysis of the networks was organized in the following stages:

a) The detection of outliers is performed using the Data Snooping technique, considering the weight matrix to be diagonal. The test is therefore carried out on the variable $w_i = \frac{v_i}{\sigma v_i}$.

If $|w_i| > F^{1/2}_{1-\alpha\ ;1,\infty}$, the i-th observation is eliminated and the adjustment is repeated. In the case of small networks, the interactive version of the procedure is preferred.

b) The adjustment is performed with the method of indirect observations. For the free net we can use the bordering technique (MEISSL, 1969) or the expression of $N^{+} = N(NN)^{-1} N(NN)^{-1} N$ (MITTERMAYER, 1972) or the singular value decomposition

$$N = U S V^{T}, \quad N^{+} = V^{T} S U$$

For the inversion of the regular matrices, the modified Cholesky method is used.

c) The analysis of the nets is carried out by considering the absolute and the relative error ellipses and the error on the distance and on the bearing between pairs of points. The ellipses may be plotted automatically using an interactive program.

d) The similarity transformation is actually carried out without taking into account errors in the coordinates of the two nets.

e) Other programs were used for the handling of the data.

REFERENCES

ARCA, S., 1979: La rete trigonametrica dell'I.G.M. e la sua utilizzazione nell'allestimento delle carte tecniche - Bollettino SIFET n. 4, 1979.

BARBARELI.A, M., 1981: Compensazione ed inserimento di una rete di poligonali catastali - Rivista del Catasto n. 2, 1981.

BARBARELIA, M., FOLLONI, G., 1979: Lo stato delle reti geodetiche italiane - Bollettino SIFET n. 4, 1979.

I.I.M. - Attivita geodetica per l'anno 1965, february, 1966.

MEISSL, P., 1969: Zusammenfassung und Ausbau der inneren Fehlertheorie eines Punkthaufens - Deutsche Geodätische Kommission, Reihe A, n. 61, 1969.

MIITERMAYER, E., 1972: A generalization of the least-squares method for the adjustment of free networks - Bulletin Geodésique, n. 1041, 1972.

MONTI, C., MUSSIO, L., 1979: Compensazione planimetrica di una rete geodetica con numerosi vertici sul piano conforme di Gauss - Bollettino SIFET, n. 4, 1979.

REGIONE EMILIA-ROMAGNA, 1980: Carte generali a grande scala 1 : 100 ÷ 1 : 2000 - July 1980.

A REMARK ON THE APPLICATION OF THE GPS PROCEDURE TO VERY IRREGULAR GRAPH

B. BENCIOLINI, B. BETTI, L. MUSSIO
Milano, Italy

ABSTRACT

A remark on the application of GPS procedure to very irregular graph, frequently arising in geodetic networks and photogrammetric blocks, is developed and a little modification to it is proposed.

1. Introduction

The use of reordering algorithms for reducing the bandwidth and the profile of a sparse matrix is now well known by geodesists and photogrammetrists.

The object of this paper is a remark about the GIBBS, POOLE and STOCKMEYER procedure and the proposal of a little modification to it. This modification allows a little gain in the bandwidth and profile reduction when the concerned matrix is very irregular as a consequence of the structure of the corresponding geometrical problem.

The proposed modification is certainly of theoretical and conceptual interest in spite of the smallness of the obtained gain.

2. The original procedure

The original reordering procedure of GIBBS, POOLE and STOCKMEYER is composed by three separate algorithms which work on the associated graph of a symmetric matrix, and which must be used sequentially.

The first algorithm produces two rooted level structures (L_v,L_u) the roots of which (v,u) are the endpoints of a pseudodiameter.

The second algorithm, which makes use of the results of the first one, partitions the original graph in a set of connected components, called C_0, C_1, C_2 ... C_n. The component C_0 is the set of vertices (including v and u) which belong to the same level in L_v and in reverse-L_u; C_0 has generally the structure of a chain. The set obtained by removing C_0 from the graph is composed by disjoint sub-sets which are ordered according to their numerosity so that the number of vertices in C_i is larger than that in C_{i+1}. These components can be considered as branches of the chain. Subsequently the algorithm assigns each vertex to a level of

a new, generally non-rooted, level structure following the criterion of minimizing the width of the structure.

The third algorithm actually produces the new numeration of the vertices of the graph, level by level:

- number one corresponds to the vertex v;
- the vertices adjacent to vertices of the previous level (selected with increasing number) are numbered at first;
- the vertices adjacent to already numbered vertices of the same level (selected with increasing number) are numbered later;
- the remaining vertices are numbered the last.

Inside each described step the vertices are numbered in order of increasing degree (the degree ρ of a vertex is the number of vertices adjacent to it).

3. The proposed modification

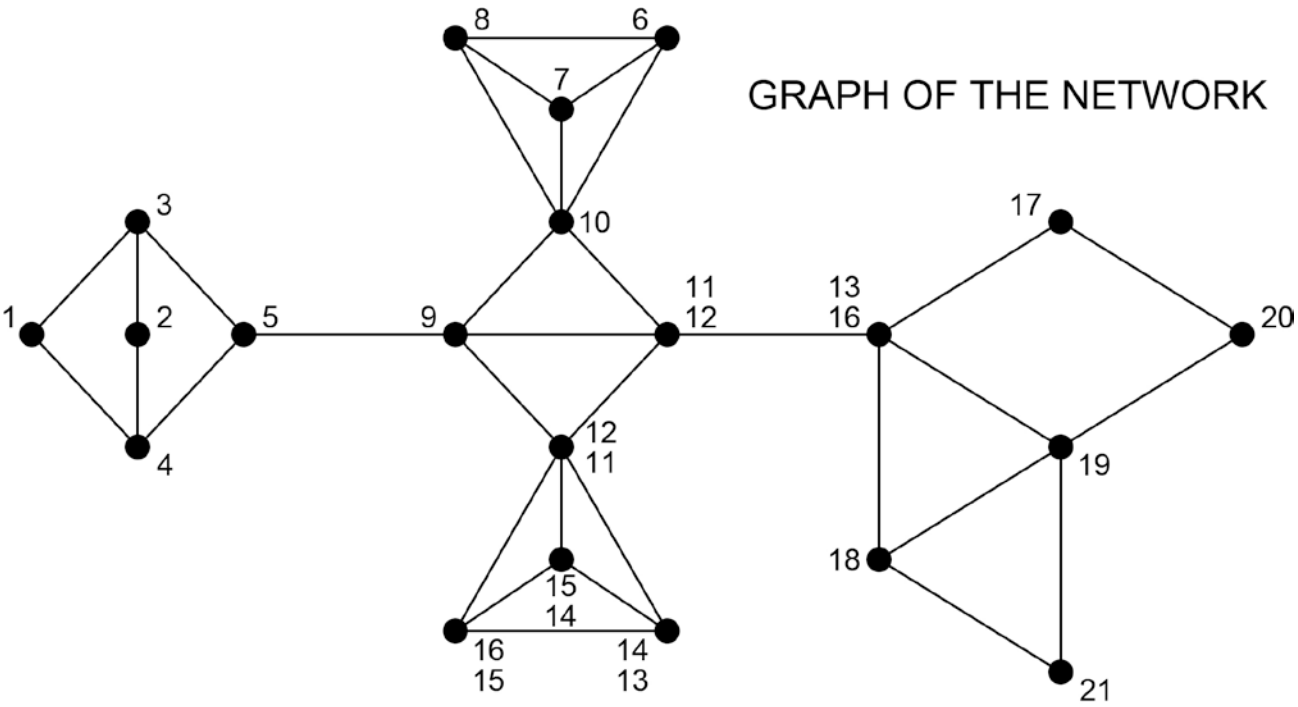
Let's begin considering the simulated examples the graph of which is represented in fig. 1.

The GPS procedure generates the component partitioning illustrated in fig. 2: the corresponding order is shown by the upper numeration of the vertices in fig. 1. The corresponding structure of the normal matrix is shown in fig. 3.

The bandwidth (=6) is determined by the vertices number 13 and 19. The vertex 13 gets its number being adjacent to the vertex number 11. The numbers 11 and 12 get their numbers being adjacent to the vertex number 9 in order of increasing degree ($\rho_{11} = 4$, $\rho_{12} = 5$). Let's note that the vertices 11 and 12 belong to different components ($11 \in C_0$; $12 \in C_2$). A better numeration (bandwidth = 4) is easily obtained if vertices numbered 11 and 12 are exchanged and the algorithm is then restarted. The profile decreases, by this exchange, from 76 to 69, too. The modified order is shown in the lower numeration of fig. 1, while the corresponding structure of the normal matrix is shown in fig. 4.

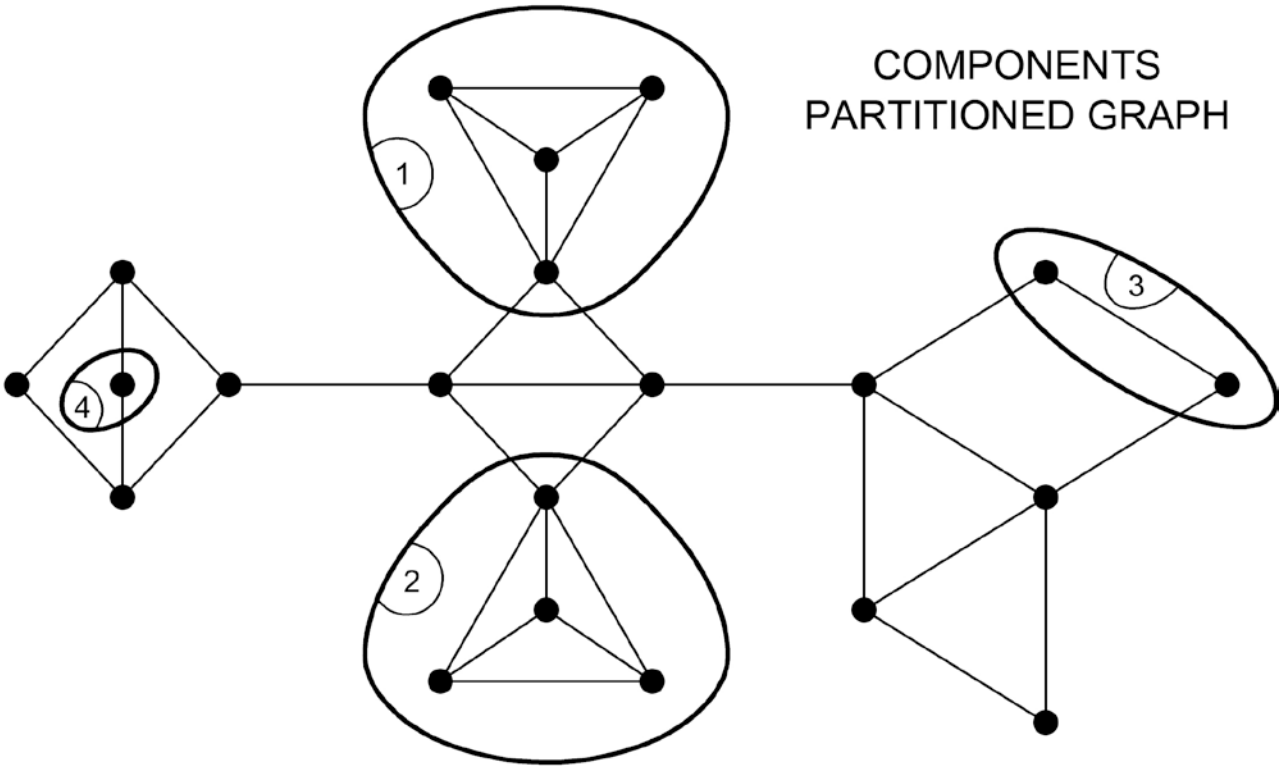
The presented example suggests the following question: is it possible to replace the "increasing degree order of numbering" mentioned at the end of previous paragraph with another more efficient order?

The aim of the use of the "increasing degree order of numbering" is the minimization of the "loca1 dispersion" of the normal matrix, which is usually produced by high degree vertices.



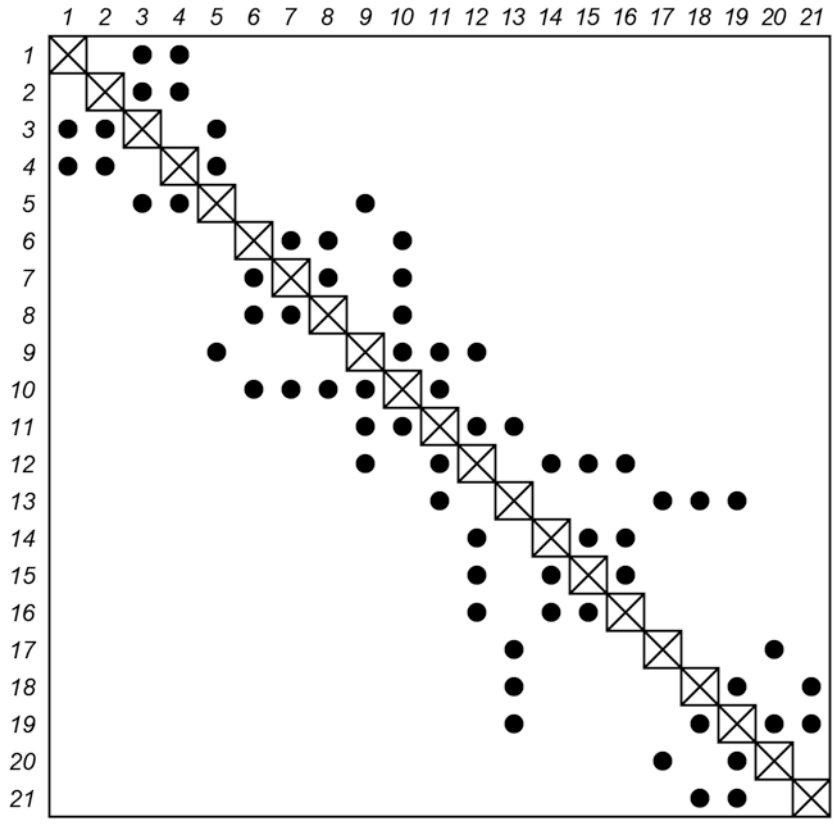
high: numeration (GPS)
low: numeration (mGPS)

fig. 1



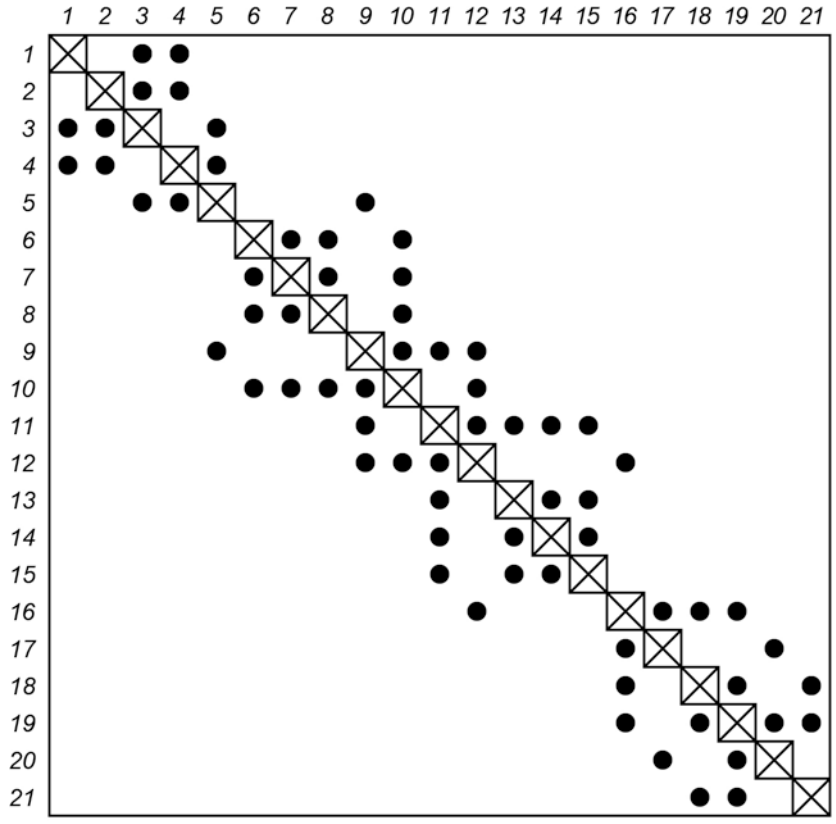
the non surrounded vertices belong to C_0

fig. 2



NORMAL MATRIX (GPS)

fig. 3



NORMAL MATRIX (mGPS)

fig. 4

The third GPS algorithm uses the level structure generated by the second algorithm, while the information about the partition in components of the graph, also supplied by the second algorithm, is not yet used.

The third modified GPS algorithm (m GPS), that is proposed here, combines the minimum degree information with the component partitioning information.

The erdering parameter (ϕ) used instead of the degree is a linear combination of the degree itself and the number, in reversed order, of the component (C_0 is a component, too) of the vertex under consideration:

$$\phi = K \frac{j}{n} + (100 - K) \frac{\rho}{\rho_{max}}$$

K = arbitrary constant ($0 \leq K \leq 100$)
j = reverse number of the component ($j = n-i$)
i = number of the component ($0 \leq i \leq n$)
Note that ϕ is not an integer number.

The use of ϕ as ordering parameter is sometime more efficient that the use of ρ in bandwidth and profile minimization. The component partitioning information is indeed not so much local as the degree, and the "zonal dispersion" in the normal matrix caused by the vertices belonging to large component is minimized.

The reasoning developed here is obviously not a proof, but it is sufficient for explaining the results obtained. As a matter of fact the m GPS procedure works better than the original GPS procedure only if applied to a very irregular graph.

4. Conclusions

The m GPS procedure has been tested against the GPS procedure on some real geodetic and photogrammetric problems. The values used for the parameter K, are :

K = 0	m GPS = GPS ; no weight to the component information
K = 25 K = 50 } K = 75	intermediate cases
K = 100	maximum weight to the component information

The component information (K = 75) proved to be useful in the case of a triangulation network of very irregular pattern, according to the unfavo-

rable terrain configuration.

On the other hand the classical GPS procedure gives better performance when applied to a more regular triangulation network, to some town levelling networks and to some photogrammetric blocks of regular shape.

The m GPS procedure is expected to be useful for every geodetic or photogrammetric problem of irregular pattern, such as vertical and horizontal (traverse and triangulation) networks, on the mountain or photogrammetric blocks obtained flying inside the valleys of a mountainous region.

REFERENCES

BENCIOLINI, B. and L. MUSSIO, 1981: Test on a reordering algorithm for geodetic networks and photogrammetric block adjustments. Proceedings of the International Symposium Management of Geodetic Data, Copenhagen, 24-26 Aug. 81.

GIBBS, N.E., W.G. POOLE, P.K. STOCKMEYER, 1976: An algorithm for reducing the bandwidth and profile of a sparse matrix. SIAM Journal Numerical Analysis, vol. 13, n. 2, April 1976.

STATUS AND PROBLEMS OF OFFICIAL HORIZONTAL DENSIFICATION NETWORKS AS SHOWN BY THE FEDERAL REPUBLIC OF GERMANY

Klaus BLACHNITZKY
München, Germany

ABSTRACT

A condensed account is rendered of accuracy status, working and personnel problems of official network densification routine in West-Germany, picking out the representative situations in the states of Baden-Württemberg (BW), Niedersachsen (Lower Saxony, LS) and in the free state of Bayern (Bavaria, BY) with particular focus on the determination of cadastral fix points in BY. The primary network as the base of densification work is outlined.

1. Introduction

In West-Germany, horizontal control is divided into major and minor control, the former consisting of the trigonometric points of 1st order (TP(1)-field for short). This primary control is broken down into 3 successive orders (TP(2), TP(3) and TP(4)-field). Outside BY and nowadays LS, Intermediate TP(1)s have been sandwiched between the regular TP(1)s and the TP(2)s. The name "trigonometric point" is an ever popular holdover since electronic range finders entered the picture three decades ago, and so is the classification into four orders since ADP has allowed us to avoid overlapping point connections by means of network adjustment and has made the lines of demarcation become blurred. As a rule, 1 TP is to be available for every 2 sq km generally. However, overdensified areas deprive us of that regular distribution. Whereas West-Germany on the whole has an average of 1.5 TPs per 2 sq km, this ratio is .6 in LS, 1.5 in BY and 3.6 in BW, although the TP-field has not been completed yet. At present, there are about 200,000 existing TPs. The annual increase is about 7,000, taking into account an annual loss of maybe 1,000.

The last densification stage is the cadastral fix points, called KFPs ("Katasterfestpunkte") in BY or APs ("Aufnahmepunkte") elsewhere. The latter term is untranslatable and, by the way, hardly a convenient label. Their number may exceed 150 mio; the annual increase is about 150,000, and the loss, mainly caused by construction work, is startling. Recommended KFP-intervals are 100 to 200 m in cities, 150 to 300 m in the outskirts and 200 to 500 m in rural areas. Point clusters are often used instead of single points.

Unfortunately there is no official German definition for the term "densification", and it is applied rather arbitrarily, e.g. in BY for TP(4) and KFP spacing only. TP-densification is primarily the domain of central state survey offices (LVA, Landesvermessungsamt). Since there is no federal authority liable for state surveys, a consultative board was established in 1948 by the state administrations. It created, among others, a working group AK Trig (Arbeitskreis Triangulierung), assigned to keep the TP-field in the various states of a similar type by working out draft regulations and standards. Considerably less uniformity is to be found in the KFP-field, where a lot of agencies and chartered surveyors (these not licensed in BY) are employed. Widespread attention is lavished on TP-determination. TPs serve for a wide scope of purposes, but most frequently for cadastral measurements. For these highly accurate TP-network is expedient, but the results of cadastral detailing depend to a much higher degree on the status of the KFP-field.

2. Cursory outline of the primary network

From the first observations in 1856 up to the last computations in 1956, it took exactly one century to install the German TP(1)-network. There was a 40-year-period of interruption between the 19th-century operations (which have rendered scale and positioning) in the north and the setting up of nets later, on in the south. Once a partial net of the latter had been observed, adjusted without constraint and fitted to the proceeding parts, the same process started in the next region. BY is covered by 7 partial nets, and altogether, we have a patchwork of 29 nets.

This scheme was not a bad one in itself, but it left behind a kind of geodetic chiaroscuro. According to claims of LS, major problems have arisen there (the oldest nets are located here), including LS to push for a complete redefinition of its TO(1)-field. But along that line of improvement the resistances are great, for such an intention is fancied neither by the neighboring states which do not want to alter point coordinates along their border nor by the Army which does not like the present network uniformity being jeopardized through introduction of double data for TPs. Furthermore, it influences both the sheet lines and the grid of large-scale cadastral maps when coordinates are altered more than the amount of drawing accuracy.

A particular test (diagnosis adjustment) is presently being carried out. Preliminary results attest to the bad-quality TP-field in LS and show a satis-

factory, almost tension-free situation in BW, slightly poorer accuracy in Northern BY and some severe defects in the Bavarian Alpine region. Since the diagnosis adjustment in BY is chiefly based on original observations and scarcely supported by modern ranging, the strains may actually be significantly greater there.

3. TP-densification

3.1 Situation in BW is characterized by four facts:

- a thoroughly satisfactory TP(1)-field relieved the manager from taking into account initial accuracy drawbacks;
- the existence of close-range, well monumented older Soldner-networks averted urgent detail-survey demands. Thus, the densification has been advanced systematically in time and place;
- a copious supply of scaffolding equipment and instruments was allocated to the task. This made it easy to overcome terrain obstacles and to reconnoiter and design highly regular-shaped nets, and to proceed on meticulous observation schedules;
- a numerous staff.

TP(2) and TP(3)-densification started simultaneously in 1948, using conventional methods (single point intercalation) at first and changing to multipoint intercalation and hybrid directions-distances adjustment in after years and to sole trilateration in 1973. Wooden observation towers from 1955 on and by wire-braced aluminum scaffold (Hi-way) from 1969 on. From the beginning of the sixties, electronic distance meters were fielded and crank-handled extension masts (Clark) used as reflector carriers. Attempts were made to introduce devices for turning the counterstations by remote control. The TP(2)-field was completed in 1978 (annual increase up to 40 points) and the TP(3)-field in 1980. Although it took 30-odd years to be completed, most of the work was done by modern means in the last decade. TP(4)-densification began parallel with the superior operations, fortunately also proceeding slowly in the first two decades. Up to 34 observers have been employed since 1974, the greater part of which recruited from the staff of regional survey offices, but always supervised by LVA-experts. Each survey team consists of 1 observer, 1 driver and 2 aids and has an electronic range finder at its disposal. The former beaconing parties have become monumenting parties and do preparatory

work: visiting, replacing or re-establishing monuments of old Soldner points which have to be included obligatorily in the new network, and marking newly reconnoitered points. As connecting link to the KFP-field, the TP(4)s succumb to constraints in locating. They are determined therefore by precise traverses primarily. To keep the scale factor stable, all distance meters are calibrated 3 to 4 times each survey period.

The new TP-network in BW is supposed to be completed in 1985. So far, all new TPs are accurate to better than 1.5 cm.

3.2 Situation in BY has been far from being so favorable as in BW:

- sundry defects sprang up and accompany the quality of the TP(1) field;
- the previous Soldner network was lacunary and of poor quality in divers places. Rural re-allotment operations, increasingly gathering momentum after the war, requested pressing TP(4)-densifications, and these demands weighed heavily upon the work. Hence the task force of the LVA was split in order to fulfill them and set up the new net as well;
- even the latter task was carried out partly in accordance with user's demands and therefore was scattered all over the country;
- circumstances enforced high-speed operations. So old observations had to be re-used regardless of age and origin, and the greater part of the TP(2)s and TP(3)s was determined before modern survey means and methods became operational in the sixties.

In spite of such disadvantages, the new network turned out to be of rather tolerable accuracy, except for a lot of not so little areas where regional survey offices complain of troublesome tension. The necessary resurvey there may or may not succeed at one go, because one tries to keep the renewal zone as small as possible and might have to repeat it later extending that zone. Furthermore, a good deal of the available staff is dedicated to taking missing or now obsolete observations. All field data, gathered within nearly one century, are stored in a data bank ("Meßelementendatei") for eventual use some day in the future. Time will show, whether this intention will be carried out then or of advanced space techniques or other developments, still only potential, might yield satisfactory results at a lower cost level.

Actually, densification started in 1950, and the number of observers hardly ever exceeded 20, though BY is West-Germany's largest state. The TP(2)-field was completed in 1970 (mostly by conventional methods) at an annual rate of

about 20 points, and some of them approach the error limit of 15 cm. Some conjuring tricks were inevitable to cope with the strains in the Alpine region. TP(3)-densification is still under way. Trilateration has not been applied, since additional direction-taking is necessary anyway to include the numerous edificational targets of the Soldner network. For TP(4)-densification, precise traverses are seldom used. TP(4)s are primarily determined in the same way as TP(3)s are and often in the same calculation run too. Some more decades will pass before the work comes to an end. Presently, there is still a backlog of unanswered requests.

The BY survey teams are composed of 1 engineer and 2 assistants. Most of the latter are seasonal workers and on the dole in winter. 3 to 4 teams form 1 survey group. One electronic distance-meter is allocated to each group. Pre-mission and post-mission calibrating is obligatory and is carried out both by means of a frequency meter and a special contrast base ("Deutscher Normalkilometer") near Munich. A number of steel towers (Clark-Köttgen) and Highway scaffolds have been bought from BW. Presently 21 steel towers up to 32 m and 8 Highway scaffolds up to 30 m are available. In 1980, steel towers were erected 19 times and crank-handled masts 77 times. A great many tree targets had to be hoisted, 55 of which were provided with reflectors rotated by ground controlled wind screen wiper motors. Successful tests concern the application of self-recording tacheometers (Zeiss Elta 2) to TP-densification, thus sparing the note-keeper and avoiding error-prone manual data transfer.

TP-networks are to be extended across national borders. So far, no problems have arisen concerning the cooperation with Austria, BW and Hesse, but only once in 1975, during an Intergerman survey campaign for border documentation, some TP connections were attainable beyond the German-German border. Afterwards the Iron Curtain fell again.

3.3 TP densification in LS is another particular business:

- accomplishment of the new superior network within the short period of one decade by exclusively modern means;
- subtle development of the subordinate network;
- postponed termination of the calculations, since the final choice of the fundamental TP(1)-coordinates is still slightly enveloped in the fog of circumstances.

From 1954 on, efforts were made to renew the TP-field systematically at the low annual rate of 1% of the state's size. It was only a prelude to the noteworthy operations that commenced in 1972 on new lines:

- independent establishment of a new TP(2)-field (including the former "Intermediate TP(1)s"); annual progress: 100 points, this meant an acceleration by the factor 8;
- TP(3)- and TP(4)-densification in selected areas of frequent and urgent demands within the framework of TP(2)-meshes and without expensive and time-wasting beaconing.

The new TP(2)-network has completely been trilaterated. It consists of ground points only and implies the old points wherever possible. As a spin-off, all 1st-order distances were measured and introduced into the diagnosis adjustment. The prevailingly flat country necessitated the employment of 30 wire-braced iron double-ladders up to 50 m in height to achieve the micro-wave distance measurements. It took five survey parties a fortnight to carry out the observations (distances only) on 35 stations. The meteorological data were thoroughly ascertained so as to have the opportunity to derive a new TP(1)-field some day from this very accurate TP(2)-field. The mean errors of the distances are calculated to about (1 - 2)E-6.

The TP(3)- and TP(4)-field are being determined at the same go, combined to relatively small network units of about 20 points, filling out a TP(2)-mesh). Since the average TP(2)-interval is 7 km (cf 14 km in BY), the average interval of the subordinate points actually is the recommended $\sqrt{2}$ = 1.4 km. The regional cadastral offices assist the LVA in reconnaissance and monumentation and thus are able to influence the point locating on behalf of the users. However, the observing work is a task of the LVA staff exclusively. The network is traced out as a net of traverses which are tied together in junction nodes. Each junction node is regarded as a TP(3), and the stations between the nodes as TP(4), and this makes the only difference. This filling-up network is said to be completed in about one decade. By 1981, 40% of the work was done already.

Unconstrained adjustments proved the high quality of the measured distances. To match these results, the TP(1)-field should be very precise, but here, as mentioned earlier, the shores of reality rise in front. Conditions of a re-newal are:

- same scale factor in all points;

- little translations to preserve the present sheet lines of cadastral maps;
- best possible fitting to the neighboring fields.

Respective feasibility studies have been elaborated.

4. KFP-densification in BY

was carried out from 1874 forth only for extensive work like renewal of urban cadastre or rural re-allotment, using traverses as sole determination method. Traverses show some advantages:

- clearly describable designing instructions: rectilinear shape, equilateral sides;
- favorable observation schemes;
- simple calculation and debugging methods suitable even for less trained personnel;
- since missing field data make the adjustment collapse, one cannot fail to take all the necessary observations and does not need sophisticated reliability indicators.

On the other hand, there are significant drawbacks:

- frequent difficulties in projecting a network of well-shaped traverses in view of terrain obstacles and irregularly distributed control stations;
- a lot of important, easily attainable neighborhood-connections (e.g. between parallel running traverses) cannot be utilized;
- the different weights of angles and distances cannot be taken into account, and this creates bigger effects, the greater the deviations from the straight-line traverse are. The simple adjustment does not agree with the field work efforts.

After World War II, a great many KFPs were determined by rural re-allotment agencies through cadastral photogrammetry. All in all, the then KFP-field left much to be desired in the article of accuracy due to uncertain distance measuring, comparatively small image scales (1 : 10,000 to 1 : 14,000) and modest evaluation means. Accordingly, the official error limits show a somewhat uneasy generosity. To cite an example: If traverses which stay just inside the confines of the maximum permitted errors are subdued to a rigid adjustment, MSR coordinate errors exceed 20 cm.

When ADP and EDM began to take the geodetic world by storm, the regional survey offices proceeded to applying coordinates, and that very idea has risen which is called "coordinate cadastre" or, more significantly, "computational cadastre". Its principles are repeated here:

- the coordinates are presumptive;
- the KFPs as the base of re-establishing or setting out property corners must be monumented and maintained in a painstaking way;
- the original field data shall never be re-used. Geometrical relations are defined or analyzed through coordinates exclusively.

Obviously, this thesis exacts KFP-field where accuracy is carried to an extreme, and all the older, less precise determinations have to be done away with. Strict standards and error limits are inevitable, and it is a question of time pressure, allocation of funds and the amount of already stored points whether or not the coordinate cadastre can be implemented. A critical attitude is now prevailing. As for BY, about 15 mio conventionally determined points have been stored, so that in case of a coordinate cadastre it would be necessary:

- either to install separate stores for first-class and for ordinary points
- or to affix an accuracy-status symbol to the numbers of the commonly stored points.

Neither way is applicable till date. BY, preparing land data bank and automized base maps, has been establishing

- a coordinate file, indiscriminately containing points of different quality
- an object file (Objektdatei) for the geometrical definitions of lots and buildings, and the basic field data may have to be reconsidered during follow-up surveys. This is called "numerical cadastre", and it is not that bad because the surveyor, contrary to the situation in an error-prone pure coordinate cadastre, is eased by the instantaneous availability of controlling data.

In 1961, the former national Soldner coordinate system (SS) was replaced by the Gauß-Krüger system (GKS), and about 15 mio SS-points fell due to be taken over into GKS by means of transformation, recalculation or new determination. This job has been proving a harder one than it was previously thought, and it might well be that it will last until the end of the current millennium. Thus,

cadastral surveys can face different initial situations:

- EDM-determined GKS-points only: a problem-free case;
- less accurate GKS-points: the most unpleasant case. As a rule, even modern surveys are to be fitted to the existing field, often causing recalculations of old points and trouble-prone alteration of areas of lots in the wake of it;
- less accurate SS-points: These do not hinder the establishing of a new GKS-field and are coordinated in GKS afterwards. Unfortunately, this idea sometimes conflicts with the setting-up of a ground data bank which enforces a uniform system all over the country.

Nowadays, the by far prevailing method of measuring distances in BY is EDM, but the KFP-determination depends on the purpose. Regional survey offices which mainly deal with smaller operations prefer traverses because they are suitable for desktop calculators and because a favorable view of traverses is still widely held. About 2/3 of all KFPs in BY are being determined this way. The rest is calculated by rigid adjustments and observed by means of self-recording tacheometers (RegElta, Elta 2) only. Such KFP-networks can show very complicated structures (Fig. 1), and optimizing the observation scheme is necessary. Since up to now there are no mathematically based procedures available, and hardly will be soon, network diagram maneuvers are being tried in order to find a favorable way. Influencing factors are: number of staff, equipment, cars, traffic conditions and roads, traffic safety conditions, temporary obstacles (which usually appear unexpected), weather etc. The only condition for the KFP-reconnaissance is: connection to neighboring points, and the diagrams are carefully checked because the ADP-programs indeed comprise stochastical reliability indicators and data snooping, but there is no known procedure which tells one that such important connections are missing.

About 1,000 KFPs are determined every year by photogrammetric means: multiple targets, comparator evaluation, bundle adjustment, and the achieved coordinate accuracy is, as in terrestric determination in the last two decades, better than 2 cm.

The regular types of KFP monuments are:

- at places where digging is possible: concrete stones 15x15x45 cm with a central boring through its longitudinal axis, put over an underground clay or iron pipe with the aid of a box-bubble equipped steel rod which can

glide in the boring (Fig. 2);

- in asphalt, tar and pavement: iron pipes protected by steel hoods;
- in street intersections: cemented, glued or shot-in bench marks in vertical walls, are combined with nearly ground points to point clusters;
- sharply definable house corners can take over the function of KFPs.

Although every endeavor has been made about carrying on KFP-densification and improving the older parts of the KFP-field, it remains a never-ending task for generations of surveyors.

REFERENCES

AUGATH, 1975: Die Mitwirkung der Katasterämter bei der Erneuerung des TP-Feldes, Nachrichten der Niedersächs. Verm. u. Katasterverwaltung 2/1975

AUGATH/NIEMEIER, 1978: On the point identity problem in control networks, VIth International Symposium on Geodetic Computations, Munich 1978

BLACHNITZKY, 1971: Eine neue Vermarkungsart für Lagefestpunkte, Mitteilungsblatt des Landesvereins Bayern des DVW 2/1971

BLACHNITZKY, 1977: The Project Moosach, Workshop on Standard and Specifications for Integrated Surveying and Mapping Systems, Munich 1977 (Proceedings 1978)

BLACHUT, 1979: Urban Surveying and Mapping, Springer-Verlag Berlin/New York 1979

HAUPT, 1979: Das Vermessungspunktfeld im Koordinatenkataster der Zukunft, AVN Allgemeine Vermessungsnachrichten 3/1979

MEIER, 1981: Erneuerung des trigonometrischen Netzes 4. Ordnung in Baden-Württemberg, ZfV Zeitschrift für Vermessungswesen 12/1981

PÖTZSCHNER, 1974: Hundert Jahre trigonometrisches Festpunktfeld, Nachrichten der Niedersächsischen Verm. u. Katasterverwaltung 4/1974

PÖTZSCHNER, 1977: Die Anpassung des Lagefestpunktfeldes an die heutigen Genauigkeitsansprüche, ZfV 4/1977

SCHMIDT, 1978: Die Diagnoseausgleichung des deutschen Hauptdreiecksnetzes, VIth International Symposium on Geodetic Computations, Munich 1978

TEGELER, 1979: Untersuchungen zur Neueinrichtung des Aufnahmepunktfeldes, Nachrichten der Niedersächs. Verm. u. Katasterverwaltung 2/1979

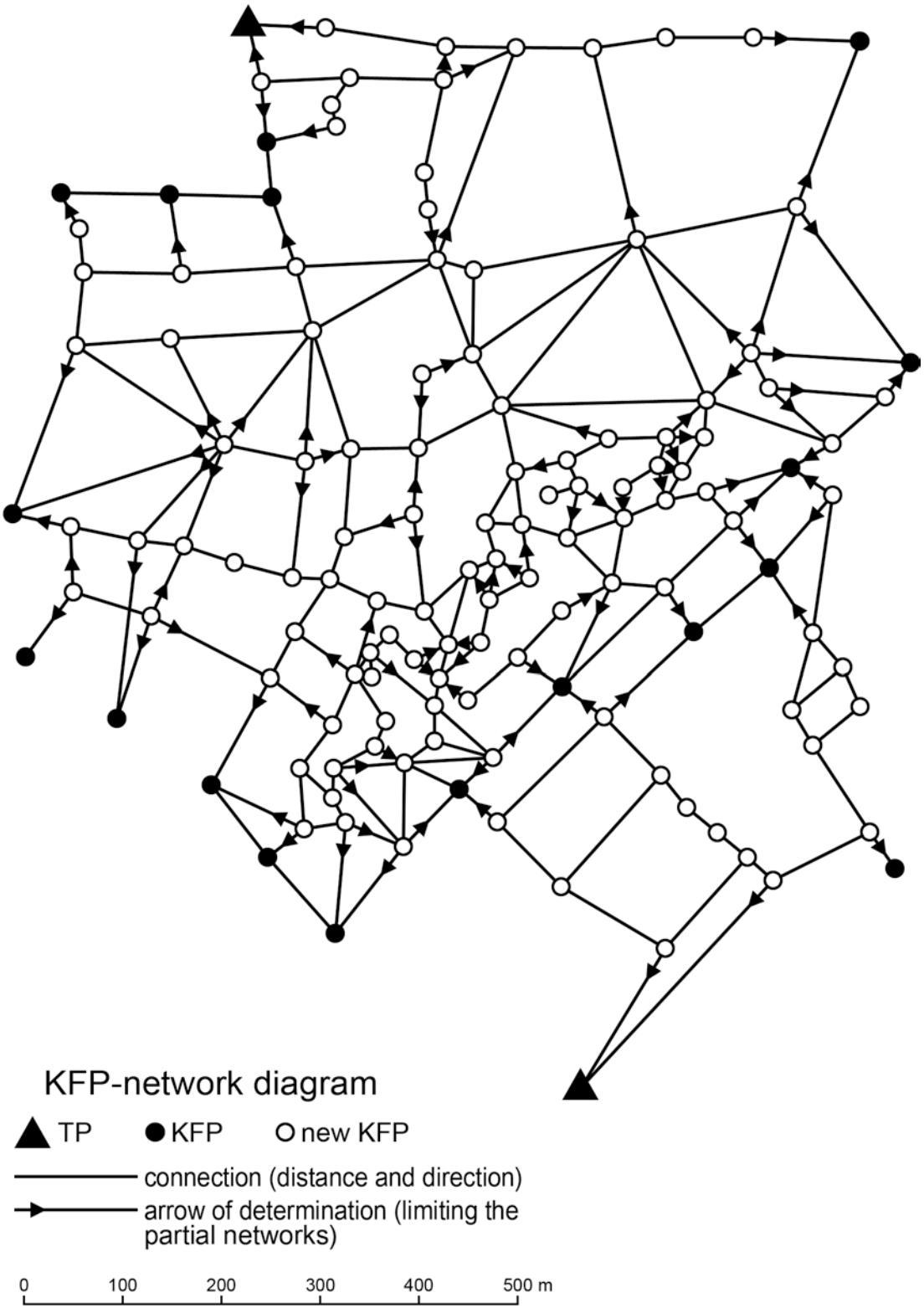

KFP-network diagram
TP
KFP
new KFP
connection (distance and direction)
arrow of determination (limiting the partial networks)
0
100
200
300
400
500 m

Figure 1

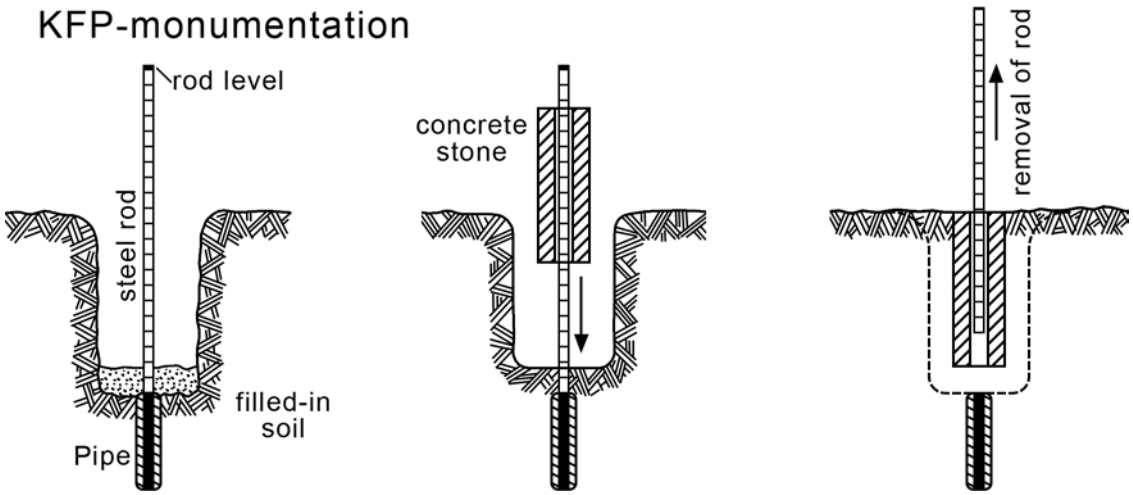

KFP-monumentation
rod level
steel rod
filled-in
soil
Pipe
concrete
stone
removal of rod

Figure 2

JUNCTION OF CONTROL SURVEYS BY ADJUSTMENT COMPARED WITH COORDINATE TRANSFORMATION

Hendrik J. BUITEN and Peter RICHARDUS
Wageningen, The Netherlands

ABSTRACT

The overdetermined similarity or "Helmert" transformation is a well known and commonly applied method to connect stations coordinated in a - subsidiary - system into another - principal - system, the coordinates of which are kept fixed.

The result is a regression type "best fit" figure with remnant discrepancies at the tie points used to calculate the transformation parameters. In this process the covariance matrices of both sets of coordinates are ignored. Those discrepancies may not be of importance for certain small and medium scale mapping. However, angles and distances calculated from a combination of the principle coordinates of the tie points and those of the other transformed stations may be affected in their relative position by local jumps to such an extent that they are unfit to be used for consistent densification in respect of large scale mapping, and checking and setting out in the field for precise engineering surveys.

A method of junction by adjustment is proposed, where the remnant discrepancies are eliminated. The covariance matrices are taken into account after having been submitted to a covariance transformation in order to relate them to a common covariance reference base. This makes also possible to apply a testing procedure to the misclosures at the tie points.

The structure is shown of a substitute covariance matrix to be put into practice where the real covariance matrices are unknown or not available. This matrix - originally proposed by Alberda and Baarda - is applied to the example of a junction of two control surveys nets at the conclusion of the paper.

1. General Introduction.

The least squares adjustment of an independent control survey according to the method of correlates leads to a uniquely closed figure. In addition the covariance matrix $Q_{PP}\sigma^2$ of the adjusted observations P becomes available (σ^2 is the variance factor), so that the precision and the internal reliability can be investigated.

The use of coordinates of the stations involves two decisions which are subjective in a way and based on many not further specified considerations some of which are obviously of a practical nature.

At first a coordinate system is defined by the selection of 4 coordinates of two stations, or two coordinates of one station, one orientation and one distance, or four linear functions of these quantities. This constitutes the four degrees of freedom to select the coordinate reference base of the control survey, from where all other coordinates are calculated by trigonometry.
The change from one coordinate system to another is assumed to be accomplished by a linear orthogonal or similarity transformation.

Secondly fixed (co)variances of 4 elements such as mentioned are selected as references for the standard curves of all other stations. This is the covariance reference base. It is usually tacitly assumed that the quantities selected according to the first decision have zero (co)variances. The coordinates of every station are expressed as functions of the coordinates to which these zero variances are assigned, and of the adjusted observations. The covariance matrix M of the coordinates is ultimately obtained by the application of the law of propagation of variance to these functions.
Changing from one covariance system to another may be carried out in two ways:

1) Indirectly by repeating the above procedure with newly selected reference (co)variances. In many cases it is not of interest to transform the coordinate system itself, leaving the nominal coordinate values unchanged;

2) directly by a covariance transformation, having one covariance matrix available (see section 3).

The adjustment of a control survey may be carried out with the method of parameters, these being the coordinates of the stations. In this case the decisions mentioned above have been taken implicitly, and may sometimes be cunningly hidden.

It should be emphasized, that the selection of a particular coordinate system does not necessarily lead to a corresponding selection of a set of covariances. The reverse is also true. The coordinate base and the covariance base do more often than not involve different respective sets of stations.

2. Introduction to the junction of surveys.

Suppose there are two of these control survey networks with respective independently defined coordinate systems and that one system (called the subsidiary system, indicated by X',Y') is to be fitted into the other - the principal - system (indicated by X,Y). Both networks may or may not be of an equal order of precision. Consistency must be achieved with respect to scale, orientation and position. Assume that m (<2) stations known by the coordinates in both systems are available for this purpose, two of which (X_1,Y_1; X_1',Y_1' and X_2,Y_2; X_2',Y_2' respectively) are selected as a common coordinate reference base for a provisional or approximate junction of the two nets. The similarity transformation shows:

$$\begin{pmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \\ \hline . \\ . \\ X_m \\ Y_m \end{pmatrix} = \left(\begin{array}{cccc} X_1' & Y_1' & 1 & 0 \\ Y_1' & -X_1' & 0 & 1 \\ X_2' & Y_2' & 1 & 0 \\ X_2' & -X_2' & 0 & 1 \\ \hline . & . & . & . \\ . & . & . & . \\ X_m' & Y_m' & 1 & 0 \\ Y_m' & -X_m' & 0 & 1 \end{array}\right) \begin{pmatrix} p \\ q \\ a \\ b \end{pmatrix} = \begin{pmatrix} E_1 \\ \hline E_2 \end{pmatrix} \begin{pmatrix} p \\ q \\ a \\ b \end{pmatrix} \qquad (1)$$

where $p = \lambda \cos\Theta$ and $q = \lambda \sin\Theta$, λ being a multiplication factor; Θ the angle of rotation; and a and b translation parameters. These are obtained from the solution of the first four equations of (1). The parameters p and q are linear functions of coordinates, these being variates (stochastical quantities), and therefore in general are variates themselves. It can then be submitted that the expectation Exp{p} and Exp{q} are similar functions of Exp{X,Y} and Exp{X',Y'}.

Let us assume that the m stations of the control survey in the subsid-

iary system X',Y' have (co)variances referring to two stations A_1 and A_2, and that in the principal system the covariances refer to the base B_1B_2. These base points do not necessarily belong to the set of m stations (fig. 1). The standard curves of the m stations transformed into the principal system are calculated by applying the law of propagation of variances to (1), taking the calculated (sample) values of p and q. Since they are still referring to the base stations A_1 and A_2, there are now 2 sets of covariances in the same X,Y coordinate system. In order to make the two sets comparable they should be transformed so as to refer to the same reference base. This transformation is derived in the next section.

3. Covariance transformation.

We now want to transfer the set of weight coefficients of the m transformed stations on base A_1A_2 (denoted by $M_{A_1A_2}$) to the base R_1R_2 leading to $\bar{M}_{R_1R_2}$ (R_1,R_2 appertinent to both the subsidiary and principal system. *We do not change the coordinate reference base*. Use is made of the general similarity transformation formulae as in (1) (see fig. 1), only to derive the differential equations.

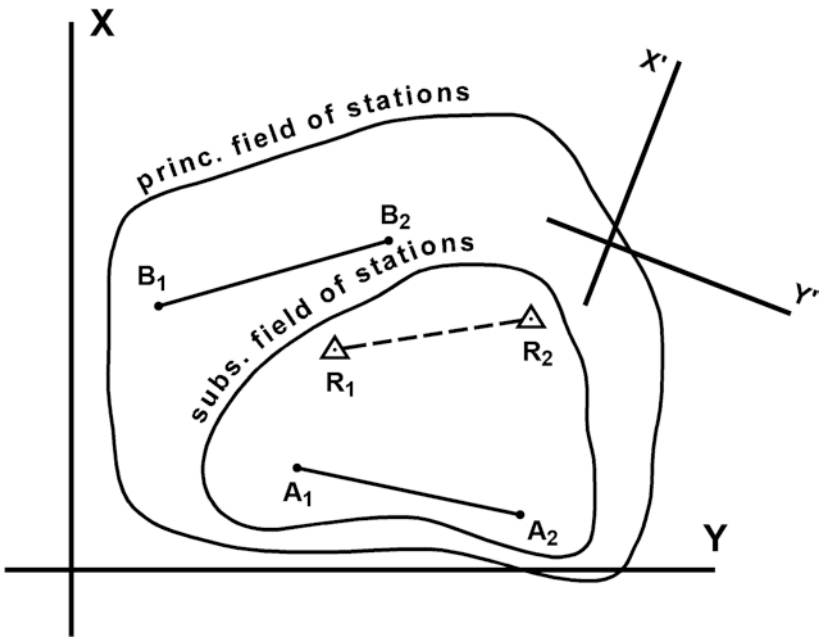


fig. 1

In respect of the covariance reference base A_1A_2, the weight coefficients $Q_{X_{A_1}X_{A_2}} = Q_{X_{A_1}Y_{A_1}} = \ldots = Q_{Y_{A_2}Y_{A_2}} = 0$; (2a)

In respect of the covariance base R_1R_2:

$\overline{Q}_{X_{R_1}X_{R_2}} = \overline{Q}_{X_{R_1}Y_{R_1}} = \ldots = \overline{Q}_{Y_{R_2}Y_{R_2}} = 0$; (2b), Considering that to each sample of X Y coordinates belongs a corresponding sample value of p and q, it follows that with $\lambda = 1$; $\cos\Theta = 1$ and $\sin\Theta = 0$, the Exp{p} = 1 and Exp{q} = 0. Further Exp{a] = Exp{b} = 0.

Taking coordinate *differences* leads to

$$\begin{pmatrix} X_{R_2} - X_{R_1} \\ Y_{R_2} - Y_{R_1} \end{pmatrix}_{(R_1R_2)} = \begin{pmatrix} X_{R_2} - X_{R_1} & Y_{R_2} - Y_{R_1} \\ Y_{R_2} - Y_{R_1} & -(X_{R_2} - X_{R_1}) \end{pmatrix}_{(A_1A_2)} \begin{pmatrix} p \\ q \end{pmatrix} \qquad (3)$$

Differentiation of (3) gives (in "Q" notation) in combination with (2)

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} Q_{X_{R_2}} - Q_{X_{R_1}} & Q_{Y_{R_2}} - Q_{Y_{R_1}} \\ Q_{Y_{R_2}} - Q_{Y_{R_1}} & -(Q_{X_{R_2}} - Q_{X_{R_1}}) \end{pmatrix}_{(A_1A_1)} \begin{pmatrix} p \\ q \end{pmatrix} + \begin{pmatrix} X_{R_2} - X_{R_1} & Y_{R_2} - Y_{R_1} \\ Y_{R_2} - Y_{R_1} & -(X_{R_2} - X_{R_1}) \end{pmatrix} \begin{pmatrix} Q_p \\ Q_q \end{pmatrix}_{(A_1A_1)} \qquad (4)$$

or, with Exp{p} = 1 and Exp{q} = 0, and putting

$x_{R_{12}} = X_{R_2} - X_{R_1}$; $y_{R_{12}} = Y_{R_2} - Y_{R_1}$

$$-\begin{pmatrix} Q_{x_{R_{12}}} \\ Q_{y_{R_{12}}} \end{pmatrix}_{(A_1A_2)} = \begin{pmatrix} x_{R_{12}} & y_{R_{12}} \\ y_{R_{12}} & -x_{R_{12}} \end{pmatrix} \begin{pmatrix} Q_p \\ Q_q \end{pmatrix}_{(A_1A_2)} \qquad (5)$$

The unknowns Q_p and Q_q are solved from (5) by

$$\begin{pmatrix} Q_p \\ Q_q \end{pmatrix}_{(A_1A_2)} = -\begin{pmatrix} x_{R_{12}} & y_{R_{12}} \\ y_{R_{12}} & -x_{R_{12}} \end{pmatrix}^{-1} \begin{pmatrix} Q_{x_{R_{12}}} \\ Q_{y_{R_{12}}} \end{pmatrix}_{(A_1A_2)} = -\frac{1}{d_{R_{12}}^2} \begin{pmatrix} x_{R_{12}} & y_{R_{12}} \\ y_{R_{12}} & -x_{R_{12}} \end{pmatrix} \begin{pmatrix} Q_{x_{R_{12}}} \\ Q_{y_{R_{12}}} \end{pmatrix}_{(A_1A_2)} \qquad (6)$$

where $d_{R_{12}}$ is the distance between the reference stations R_1 and R_2.

For each station i (i = 3,....,m) it can be derived that

$$\begin{pmatrix} X_i - X_{R_2} \\ Y_i - Y_{R_2} \end{pmatrix}_{(R_1R_2)} = \begin{pmatrix} X_i - X_{R_2} & Y_i - Y_{R_2} \\ Y_i - Y_{R_2} & -(X_i - X_{R_2}) \end{pmatrix}_{(A_1A_2)} \begin{pmatrix} p \\ q \end{pmatrix}$$

and subsequently, referring to (2a and b)

$$\begin{pmatrix} \overline{Q}_{x_i} \\ \overline{Q}_{y_i} \end{pmatrix}_{(R_1R_2)} = \begin{pmatrix} x_{R_2i} & y_{R_2i} \\ y_{R_2i} & -x_{R_2i} \end{pmatrix} \begin{pmatrix} Q_p \\ Q_q \end{pmatrix}_{(A_1A_2)} + \begin{pmatrix} Q_{x_{R_2i}} \\ Q_{y_{R_2i}} \end{pmatrix}_{(A_1A_2)} \qquad (7)$$

The substitution of (6) ultimately transposes this into

$$\begin{pmatrix} \overline{Q}_{x_i} \\ \overline{Q}_{x_i} \end{pmatrix}_{(R_1R_2)} = -\frac{1}{d_{R_{12}}^2} \begin{pmatrix} x_{R_2i}x_{R_{12}} + y_{R_2i}y_{R_{12}} & x_{R_2i}y_{R_{12}} - y_{R_2i}x_{R_{12}} \\ y_{R_2i}x_{R_{12}} - x_{R_2i}y_{R_{12}} & y_{R_2i}y_{R_{12}} + x_{R_2i}x_{R_{12}} \end{pmatrix}_{(A_1A_2)} \text{ times}$$

$$\begin{pmatrix} Q_{x_{R_{12}}} \\ Q_{y_{R_{12}}} \end{pmatrix}_{(A_1A_2)} + \begin{pmatrix} Q_{x_{R_2i}} \\ Q_{y_{R_2i}} \end{pmatrix}_{(A_1A_2)} \qquad (8)$$

Denoting the matrix of coefficients of $\left(Q_{x_{R_{12}}}, Q_{y_{R_{12}}}\right)^T$ by + F_i (inclusive the scalar $-1/d_{R_{12}}^2$) we have after some manipulations

$$\begin{pmatrix} \overline{Q}_{X_i} \\ \overline{Q}_{y_i} \end{pmatrix}_{(R_1R_2)} = \left(I \vdots (-F_i) \vdots (F_i - I)\right) \left(Q_{X_i}\ Q_{Y_i}\ Q_{X_{R_1}}\ Q_{Y_{R_1}}\ Q_{X_{R_2}}\ Q_{Y_{R_2}}\right)^T_{(A_1A_2)} \qquad (9)$$

where $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Application of the law of propagation of variances to (9) gives the transformation of the matrix of weight coefficients in respect of a selected covariance base R_1R_2, of the weight coefficients are given in respect of some other covariance base A_1A_2.

The corresponding *covariance transformation* then becomes

$$\begin{pmatrix} \overline{\sigma}_{X_i} \\ \overline{\sigma}_{Y_i} \end{pmatrix} \begin{pmatrix} \overline{\sigma}_{X_k} \\ \overline{\sigma}_{Y_k} \end{pmatrix}^T_{(R_1R_2)} = \left(I \vdots (-F_i) \vdots (F_i - I)\right) \begin{pmatrix} \sigma_{X_i} \\ \sigma_{Y_i} \\ \sigma_{X_{R_1}} \\ \sigma_{Y_{R_1}} \\ \sigma_{X_{R_2}} \\ \sigma_{Y_{R_2}} \end{pmatrix} \begin{pmatrix} \sigma_{X_k} \\ \sigma_{Y_k} \\ \sigma_{X_{R_1}} \\ \sigma_{Y_{R_1}} \\ \sigma_{X_{R_2}} \\ \sigma_{Y_{R_2}} \end{pmatrix}^T_{(A_1A_2)} \begin{pmatrix} I \\ \dots \\ (-F_k) \\ \dots \\ (F_k - I) \end{pmatrix} \qquad (10)$$

These equations have been derived already by Baarda in [4] in the notation of complex numbers.

4. Conditions for external consistency with redundant data; adjustment and testing.

4.1 Junction by adjustment.

In section 2 it has been assumed that m stations are given in both the subsidiary and principal systems, both systems were connected by a similarity transformation at two points viz. the coordinate reference base.

Within the scope of this paper, the two point covariance reference base will be adhered to. For densification purposes the coordinates in the principal system will be held fixed.

Then the following data are available:

1) The principle system with a total of m stations indicated individually by i or j ($m > 2$; $i, j = 1...m$); and the covariance matrix M_1;

2) The subsidiary system having a total of n stations. Of these, m stations are known by the coordinates in both systems ($n > m$). Indices k and l are used running from (m+1) to n.

Further we know the 2n x 2n covariance matrix M_2 of the coordinates in this subsidiary system.

4.1.1. If we deal with a free independent network (no coordinates) the first above mentioned decision leads directly to the coordinate system being selected to be that of the principal system.

The network is oriented and brought to scale (necessitating two stations), and coordinates of all n stations calculated. Then at (m-2) stations these will show discrepancies with the given coordinates in the principal system. This leads to the well known coordinate condition equations, two for each redundant station. They can be formulated as misclosures for the X- and for the Y-coordinates. *Only the original observations in the network receive corrections* since the given coordinates are considered as having been introduced into the adjustment with zero (co)variances.

However, the <u>real</u> covariance matrix of the principal coordinates is used to calculate the covariance matrix of the new coordinates in relation to the covariance reference base of the principal system.

4.1.2. Now suppose the subsidiary system is defined by the two decisions, the covariance reference base being A_1A_2. The corresponding rows and columns in the covariance matrix M_2, where the elements with respect to $Q_{X_{A_1}}$, $Q_{Y_{A_1}}$, $Q_{X_{A_2}}$, $Q_{Y_{A_2}}$ occur, consist of zeros. We write

$$M_{2\ 2n2n} = \begin{pmatrix} M_{ij} & | & M_{il} \\ - - & + & - - \\ M_{kj} & | & M_{kl} \end{pmatrix}_{2\ (A_1A_2)} = M_{2\ (A_1A_2)}^{T} = \begin{pmatrix} M_I & | & M_{II} \\ - - & + & - - \\ M_{IV} & | & M_{III} \end{pmatrix}_{2\ (A_1A_2)} \qquad (11)$$

where the submatrix M_{ij} denotes the covariance matrix of the m stations common to both systems. Out of these m points a new reference base - say R_1R_2 - is selected. A geometrical rotation and a change of scale of the axes of the standard curves, and a covariance transformation lead to a matrix $(\bar{M}_2)_{2n2n(R_1R_2)}$. This is the covariance matrix of the coordinates which are to be considered as "observations" in the principal system.
A covariance matrix $(\bar{M}_1)_{(R_1R_2)}$ of the m stations should be calculated in the principal system, transformed if necessary to this very base from B_1B_2 (fig. 1), similar to (10).

The matrix $\bar{M}_1$ and the submatrix $\bar{M}_I$ should be identical if the two respective sets of observations would be so with regard to type and accuracy. A different level of accuracy would only lead to similar standard curves at a different scale. If, however, various types of measurement have been applied (e.g. angle observations only with regard to $\bar{M}_{1_{i.j}}$ and angle and/or distance measurement in respect of $\bar{M}_I$), the standard curves would have different shapes.
The matrix of the discrepancies t mentioned earlier then

is $$\bar{M}_{(t)_{i.j}} = (\bar{M}_1 + \bar{M}_I)_{(R_1 R_2)} \tag{12}$$

in the latter case, or

$$\bar{M}_{(t)_{i.j}} = (\bar{M}_1 + \rho^2 \bar{M}_1)_{(R_1 R_2)} \tag{13}$$

in the former case, where ρ is a scale factor. Both sets of coordinates are considered independent and free of correlation.

Written in full the covariance matrix for all n stations is: (see page 124).

$$(\bar{M}_2)_{(R_1R_2)} = \begin{pmatrix} \bar{M}_{1.1} & \cdots & \bar{M}_{1.m} & | & \bar{M}_{1.m+1} & \cdots & \bar{M}_{1.n} \\ \vdots & & \vdots & | & \vdots & & \vdots \\ \bar{M}_{m.1} & \cdots & \bar{M}_{m.m} & | & \bar{M}_{m.m+1} & \cdots & \bar{M}_{m.n} \\ \hline \bar{M}_{m+1.1} & \cdots & \bar{M}_{m+1.m} & | & \bar{M}_{m+1.m+1} & \cdots & \bar{M}_{m+1.n} \\ \vdots & & \vdots & | & \vdots & & \vdots \\ \bar{M}_{n.1} & \cdots & \bar{M}_{n.m} & | & \bar{M}_{n.m+1} & \cdots & \bar{M}_{n.n} \end{pmatrix}_{2(R_1R_2)} =$$

$$\begin{pmatrix} \bar{M}_I & | & \bar{M}_{II} \\ \hline \bar{M}_{IV} & | & \bar{M}_{III} \end{pmatrix}_{(R_1R_2)}$$

The zero rows/columns in both the $\bar{M}_1$ and $\bar{M}_2$ matrices may be omitted, but are maintained for the sake of clarity.

The junction can now be carried out according to the following rules

1) Select two points 1 and 2 out of the m stations (as in (1)) which will not receive corrections as a matter of course (the misclosures $t_{X_1} = t_{Y_1} = t_{X_2} = t_{Y_2} = 0$).

Theoretically the adjustment may be formulated, as if both the principal (1) and the transformed subsidiary coordinates (2) are variates. Then,

$$\underset{2m.4m}{\begin{pmatrix} 1 \cdots 0 & | & -1 \cdots 0 \\ \vdots & | & \vdots \\ 0 \cdots 1 & | & 0 \cdots -1 \end{pmatrix}} \begin{pmatrix} X_1^{(1)} + \varepsilon_{X_1}^{(1)} \\ \vdots \\ Y_m^{(1)} + \varepsilon_{Y_m}^{(1)} \\ \hline X_1^{(2)} + \varepsilon_{X_1}^{(2)} \\ \vdots \\ Y_m^{(2)} + \varepsilon_{Y_m}^{(2)} \end{pmatrix}_{4m.1} = \underset{2m.1}{(0)} \quad \text{with } g = \begin{pmatrix} \underset{2m.2m}{\bar{M}_1} & | & 0 \\ \hline 0 & | & \underset{2m.2m}{\bar{M}_I} \end{pmatrix}_{(R_1R_2)} \qquad (15)$$

This adjustment can be solved in two phases; in the first phase the principal coordinates are considered free of error.

We define

$$X_i^{(2)} + \varepsilon_{X_i}^{(2)} - \varepsilon_{X_i}^{(1)} = X_i^{(1)} = X_i^{(2)} + v_{X_i}$$

$$Y_i^{(2)} + \varepsilon_{Y_i}^{(2)} - \varepsilon_{Y_i}^{(1)} = Y_i^{(1)} = Y_i^{(2)} + v_{Y_i}$$

and the corresponding covariance matrix $\bar{M} = \bar{M}_1 + \bar{M}_I$ of (12).
The corrections v are then found with

$$\begin{pmatrix} v_{X_1} \\ \vdots \\ v_{Y_m} \end{pmatrix} = (I) \begin{pmatrix} X_1^{(1)} - X_1^{(2)} \\ \vdots \\ Y_m^{(1)} - Y_m^{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ t_{X_3} \\ \vdots \\ t_{Y_m} \end{pmatrix} \tag{16}$$

Since the principal coordinates in the prevailing junction of the two networks cannot be corrected from a practical point of view, the second phase of adjustment is to be omitted.

The transformed subsidiary coordinates of the m-2 stations other than 1 and 2 receive corrections equal to the misclosures t.

2) The principal coordinates of the remaining n-m stations in the subsidiary system are calculated as follows.

We write

$$\left.\begin{aligned} X_1^{(1)} &= X_1^{(2)} \\ Y_1^{(1)} &= Y_1^{(2)} \\ &\vdots \quad \vdots \\ X_3^{(1)} &= X_3^{(2)} + t_{X_3} \\ Y_3^{(1)} &= Y_3^{(2)} + t_{Y_3} \\ &\vdots \quad \vdots \\ X_m^{(1)} &= X_m^{(2)} + t_{X_m} \\ Y_m^{(1)} &= Y_m^{(2)} + t_{Y_m} \end{aligned}\right\} \text{"constrained"}$$

$$\text{----------}$$

$$\text{----------} \tag{17}$$

$$\left.\begin{aligned} X_{m+1}^{(1)} &= X_{m+1}^{(2)} + v_{X_{m+1}} \\ Y_{m+1}^{(1)} &= Y_{m+1}^{(2)} + v_{Y_{m+1}} \\ &\vdots \quad \vdots \\ X_n^{(1)} &= X_n^{(2)} + v_{X_n} \\ Y_n^{(1)} &= Y_n^{(2)} + v_{Y_n} \end{aligned}\right\} \text{"free"}$$

These conditions are not disturbed if

$$\left(\begin{array}{c|c} 1_{1.1}\dots 0_{1.2m} & 1_{1.2m+1}\dots 0_{1.2n} \\ \vdots \quad\quad \vdots & \vdots \quad\quad\quad \vdots \\ 0_{2m.1}\dots 1_{2m.2m} & 0_{2m.2m+1}\dots 1_{2m.2n} \end{array}\right) \underset{2n.1}{\left(\begin{array}{c} v_{X_1} \\ \vdots \\ v_{Y_1} \\ \hline v_{X_{m+1}} \\ \vdots \\ v_{Y_n} \end{array}\right)} = \underset{2m.1}{\left(\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ t_{X_3} \\ t_{Y_3} \\ \vdots \\ t_{X_m} \\ t_{Y_m} \end{array}\right)} \tag{18}$$

The solution is given by the well known formula

$$v = gU^T(UgU^T)^{-1}\,t.$$

With $U=(I\vdots 0)$ of (18) the corrections v are calculated from

$$v = \left(\begin{array}{c|c} \bar{M} & \bar{M}_{II} \\ \hline \bar{M}_{IV} & \bar{M}_{III} \end{array}\right)\left(\begin{array}{c} I \\ \hline 0 \end{array}\right)\left\{(I|0)\left(\begin{array}{c|c} \bar{M} & \bar{M}_{II} \\ \hline \bar{M}_{IV} & \bar{M}_{III} \end{array}\right)\left(\begin{array}{c} I \\ \hline 0 \end{array}\right)\right\}^{-1} t$$

which transposes into

$$v = \left(\begin{array}{c} I \\ \hline \bar{M}_{IV}\bar{M}^{-1} \end{array}\right) t \tag{19}$$

where $\bar{M}$ equals $\bar{M}_{(t)_{ij}}$ of (12).

It should be noted that the conditions of (17) are independent of the selection of the "zero" points 1 and 2; and the new principal coordinates of the stations m+1 to n are invariant to this selection.

4.2. Testing. From the point of view of reliability it is not indifferent which points of the common stations constitute the "zero" base. The selection can be supported by a testing procedure [3].

The estimated variance factor s^2 is equal to

$$s^2 = \frac{t(\bar{M}_t)^{-1}t^T}{2m-4}\ , \tag{20}$$

2m-4 being the number of redundancies. This value should be compared with $\sigma^2 = 1$.

If $s^2 > F_{(1-a),(2m-4),\infty}$ the testing should be continued with a data snooping.

Let the coordinate reference basepoints 1 and 2 be put to the test first. The influence of a deviation of these points can be formulated as

$$\begin{pmatrix} \nabla_{(X_i^{(2)}-X_i^{(1)})} \\ \hline \nabla_{(Y_i^{(2)}-Y_i^{(1)})} \end{pmatrix} = \left(\begin{array}{c|c} & \\ F_i & (I-F_i) \\ & \end{array}\right) \begin{pmatrix} \nabla_{(X_i^{(2)}-X_i^{(1)})} \\ \nabla_{(Y_i^{(2)}-Y_i^{(1)})} \\ \hline \nabla_{(X_2^{(2)}-X_2^{(1)})} \\ \nabla_{(Y_2^{(2)}-Y_2^{(1)})} \end{pmatrix} \qquad (21)$$

where i runs from 1...m.

This formula is obtained by the differentiation of the formulae of the similarity transformation, omitting the second and higher order terms.

Further

$$(F_{11})_i = (F_{22})_i = -\frac{1}{d_{12}^2}(x_{2i}\ x_{12} + y_{2i}\ y_{12}) \qquad (22)$$

$$(F_{12})_i = -(F_{21})_i = +\frac{1}{d_{12}^2}(y_{2i}\ x_{12} - x_{2i}\ y_{12}) \qquad (22)$$

The four alternative hypotheses $(H_a)_r$ for the two zero points are respectively (r = 1,2,3,4), simplifying the notation by putting $\nabla X_i = \nabla_{(X_i^{(2)}-X_i^{(1)})}$ etc.

$$\begin{pmatrix} \nabla X_1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \nabla Y_1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ \nabla X_2 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ \nabla Y_2 \end{pmatrix}; \qquad (23)$$

whence

$$\begin{pmatrix}\nabla X_i\\ \\ \nabla Y_i\end{pmatrix}=\begin{pmatrix}(F_{11})_i\\ \\ (F_{21})_i\end{pmatrix}\nabla X_i \ , \quad \begin{pmatrix}\nabla X_i\\ \\ \nabla Y_i\end{pmatrix}=\begin{pmatrix}(F_{12})_i\\ \\ (F_{22})_i\end{pmatrix}\nabla Y_i \quad \text{etc.}$$

The first hypothesis $(H_a)_1$ indicates for the other (m-2) stations that, referring to (21)

$$\begin{pmatrix}\nabla X_1\\ \\ \nabla Y_1\\ \hline \vdots\\ \hline \nabla X_m\\ \\ \nabla Y_m\end{pmatrix} = \begin{pmatrix}(F_{11})_1 & (F_{12})_1 & 1-(F_{11})_1 & (F_{12})_1\\ (F_{21})_1 & (F_{22})_1 & (F_{21})_1 & 1-(F_{22})_1\\ \hline \vdots & \vdots & \vdots & \vdots\\ \hline (F_{11})_m & (F_{12})_m & 1-(F_{11})_m & (F_{12})_m\\ \\ (F_{21})_m & (F_{22})_m & (F_{21})_m & 1-(F_{22})_m\end{pmatrix}\begin{pmatrix}\nabla X_1\\ \\ 0\\ \hline \vdots\\ \hline 0\\ \\ 0\end{pmatrix}$$

$$= \begin{pmatrix}(F_{11})_1\\ \\ (F_{21})_1\\ \hline \vdots\\ \hline (F_{11})_m\\ \\ (F_{21})_m\end{pmatrix} \nabla X_1 = c_1 \nabla X_1 \qquad (24)$$

In the same way we get under the 2nd hypothesis $(H_a)_2$

$$\begin{pmatrix}\nabla X_1\\ \\ \nabla Y_1\\ \hline \vdots\\ \hline \nabla X_m\\ \\ \nabla Y_m\end{pmatrix} = \begin{pmatrix}(F_{11})_1\\ \\ (F_{21})_1\\ \hline \vdots\\ \hline (F_{11})_m\\ \\ (F_{21})_m\end{pmatrix} \nabla Y_1 = c_2 \nabla Y_1 \quad \text{etc.}$$

The statistic T_1 is given by the well known formula

$$T_1 = -\frac{c^T w v}{\sigma\sqrt{c^T w Q_{vv} w c}}$$

with 1) v as given in (19) but only for the junction points

2) $w = g^{-1} = (\overline{M}_t)^{-1}$ (omitting the zero rows and columns with regard to the base stations)

3) $wQ_{vv}w = g^{-1} = w = (\overline{M}_t)^{-1}$

4) $\sigma = 1$

as U = I according to (16).

Hence

$$(T_1)_r = -\frac{c_r^T(\overline{M}_t)^{-1}t}{\sqrt{c_r^T(\overline{M}_t)^{-1}c_r}} \qquad (25a)$$

where c_1, c_2, c_3 and c_4 are the corresponding vector columns of (24). The T_1 value of the other (m-2) stations are obtained by the substitution of $c^T = (0 \ldots c_{X_i}, c_{Y_i} \ldots 0)$, talking respectively

$c_{X_i} = 1$ and $c_{Y_i} = 0$; and $c_{X_i} = 0$ and $c_{Y_i} = 1$ per station.

Then

$$(T_1)_{X_i} = -\frac{c_{X_i}^T \ (\overline{M}_t)^{-1} \ t}{\sqrt{c_{X_i}^T \ (\overline{M}_t)^{-1} \ c_{X_i}}} \qquad (25b)$$

$$(T_1)_{Y_i} = -\frac{c_{Y_i}^T \ (\overline{M}_t)^{-1} \ t}{\sqrt{c_{Y_i}^T \ (\overline{M}_t)^{-1} \ c_{Y_i}}}$$

(i = 3 ..... m)

Testing follows the prescriptive sequence, leading to the acception of the H_0 hypothesis (H_a rejected) if this hypothesis is also accepted in the variance ratio test.

This makes possible the data snooping on the misclosures at the m points in the principal system, and the selection of the better set of "zero" points in respect of the reliability.

In cases where $\bar{M}_t = (1+\rho^2)\ \bar{M}_1$ is applicable (se (13)), (25) becomes

$$(T_1)_r = -\frac{c_r^T\ (\bar{M}_1)^{-1}\ t}{\sqrt{(1+\rho^2)\ c_r^T\ (\bar{M}_t)^{-1}\ c_r}} \qquad (26)$$

It often happens in practice that neither the covariance in the principal system not those in the subsidiary system are known, or that they are only partially given. Estimated covariances (see section 6) may then be introduced, which should be realistic and compatible in the one reference system. If not, they may unsettle the testing results, or even make impossible to interpret the quality of the observations or parameters for a subsequent action. It is obvious that the subjective judgment and decisions of the geodetic engineer cannot be excluded in this process.

5. Comparison with the overdetermined similarity transformation.

A frequently applied method of combining networks into the one system is the overdetermined similarity or Helmert transformation. As in the previous section, the coordinates of m stations are given in both systems. The formulae are well known and need not be repeated here.

The phases of the computations are:

1) Calculation of the transformation parameters p, q, a and b in (1) by a least squares procedure. In almost all applications identity covariance matrices are used, but one may of course refine the method by introducing the real matrix or a suitably estimated one.

The position of the covariance reference bases in both networks is left out of consideration altogether.

The condition of adjustment holds the minimization of the sum of the squares of the misclosures.

$$t_x^T t_x + t_y^T t_y = t^T t \text{ is a Minimum.} \qquad (27)$$

2) Having solved the p and q (a = b = 0), the remainder of the (n-m) stations can be coordinated into the principal system.

This method commonly put into practice in this way is a statistical regression process of "best fit". As opposed to the previous method it leaves remnant differences in the given principal coordinates since these cannot be corrected; and as such it represents a different purpose. The local jumps in the relative position and precision may often cause intolerable discrepancies, when angles or distances computed from a combination of the principal coordinates of the m stations and those of the "new" (formerly subsidiary) n-m stations are checked out in the field. For certain medium and small scale mapping these discrepancies may not be of importance. For precise engineering surveys and large scale mapping the application of this method is prohibitive.

Testing becomes of little significance in the light of the nature of the Helmert transformation.

6. An estimated substitute covariance matrix.

6.1. In practice especially in existing networks, it frequently occurs that the covariance matrix of the coordinates is unknown. In engineering surveying it is not always possible to design new independent networks, not the least for economical reasons. In addition it is not recommended to use several coordinate systems in a project at the same time. It may lead not only to administrative difficulties but also to most annoying complications in the processing of field observations and the subsequent engineering action.

It is therefore important an estimation can be made of the covariance matrix of coordinates (at least regional) where

1) the position of the covariance reference base is of little importance or sufficiently remote so as to have an approximately evenly distributed influence on the regional control stations;

2) a reasonably reliable guidance is ascertained for the design of densification control surveys in order to maintain a homogeneous precision;

3) the matrix should be of a simple structure, and easy to handle.

According to a theory originally developed by Alberda (1963) (see also Baarda [4]) and more recently described by him in [1] and Buiten [5], the size of the standard ellipse can be estimated according to the

following rules.

1) The standard curve of a point is estimated by a circle with the semi major axis of the appertinent ellipse as a radius (fig. 2), ignoring the correlation of the coordinates at that point:

$$r^2 = \sigma_{X_i}^2 = \sigma_{Y_i}^2 \text{ and } \sigma_{X_iY_i} = 0 \tag{28}$$

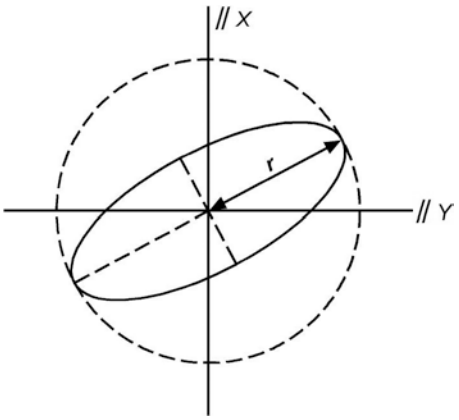


Fig. 2

2) Within the region of the points, homogeneity and isotropy is assumed, making the relative precision of any two points i and k dependent only on the distance d_{ik} and independent of the orientation of that distance. The estimated standard deviation of either point at the terminal is linked with the distance d_{ik} taking

$$r = c\sqrt{d_{ik}+c_R} \tag{29}$$

whereas the relative standard circle of two points i and k is estimated by the radius $r\sqrt{2}$.

The constant c is dependent on the precision; one may take c = 3, or c = 6, according to the precision required; c_R is a constant, securing a minimum value of r not equal to zero, if d_{ik} approaches to zero. In the literature one finds c_R = 0.05 [2]. If d_{ik} is expressed in units of km, the r is obtained in cm.

3) Coordinates of two points i and k are assumed free of correlation if they are separated by a distance d_R, for which it is valid that this length is greater than any length d_{ik} in the region

$$d_R > \text{maximum } d_{ik} \tag{30}$$

In agreement with (29) all points for which the mutual distance is equal

to d_R have standard circles

$$r_R = c\sqrt{d_R + c_R} \tag{31}$$

In view of the first requirement of the substitute matrix, any standard curve of points in the region will by estimated by (31), the correlation terms giving the stochastical dependence. (32)

Referring to the formulae of the relative standard ellipse, re-collecting $x_{ik} = X_k - X_i$ and $y_{ik} = Y_k - Y_i$, namely

$$\begin{aligned} \sigma^2_{x_{ik}} &= \sigma^2_{X_k} - 2\sigma_{X_i X_k} + \sigma^2_{X_i} \\ \sigma^2_{y_{ik}} &= \sigma^2_{Y_k} - 2\sigma_{Y_i Y_k} + \sigma^2_{Y_i} \\ \sigma_{x_{ik} y_{ik}} &= \sigma_{X_k Y_k} - \sigma_{X_k Y_i} - \sigma_{X_i Y_k} + \sigma_{X_i Y_i} \end{aligned} \tag{32}$$

and using (29), this estimation leads to the formulae

$$\begin{aligned} \sigma^2_{x_{ik}} &= 2r^2 = 2r_R^2 - 2\sigma_{X_i X_k} \\ \sigma^2_{y_{ik}} &= 2r^2 = 2r_R^2 - 2\sigma_{Y_i Y_k} \\ \sigma_{x_{ik} y_{ik}} &= 0 = 0 - \sigma_{X_k Y_i} - \sigma_{X_i Y_k} + 0 \end{aligned} \tag{33}$$

The <u>covariance function</u> follows directly from (33)

$$\sigma_{X_i X_k} = \sigma_{Y_i Y_k} = r_R^2 - r^2 = c^2(d_R - d_{ik}) \tag{34}$$

It should be noted that the constant c_R is eliminated.

The third equation of (33) also shows that

$$\sigma_{X_i X_k} = -\sigma_{X_k Y_i} \tag{35}$$

We now formulate rule

4) The correlation is considered positive for any pair of points with

$$d_{ik} < d_R \tag{36}$$

Under the assumptions of homogeneity and isotropy there should be no preference of one index to the other. No orientation being assigned to a line ($d_{ik} = d_{ki}$) whence $\sigma_{X_iY_k}$ should be equal to $\sigma_{X_kY_i}$.

In view of (35) this can be satisfied only if

$$\sigma_{X_iY_k} = -\sigma_{X_kY_i} = 0 \tag{37}$$

As a result the variances/covariances of the coordinates of any pair of stations in a region where the distance between these stations is smaller than a maximum distance d_R according to (36), may be estimated by (31) and (34).

$$M = \begin{array}{c|cc:cc} & \sigma_{X_i} & \sigma_{Y_i} & \sigma_{X_k} & \sigma_{Y_k} \\ \hline \sigma_{X_i} & c^2 d_R & 0 & c^2(d_R - d_{ik}) & 0 \\ \sigma_{Y_i} & & c^2 d_R & 0 & c^2(d_R - d_{ik}) \\ \hdashline \sigma_{X_k} & & & c^2 d_R & 0 \\ \sigma_{Y_k} & & & & c^2 d_R \end{array} = M^T \tag{38}$$

This matrix is positive definite. The correlation coefficient is equal to

$$\frac{\sigma_{X_iX_k}}{\sigma_{X_iX_k}} = \frac{\sigma_{Y_iY_k}}{\sigma_{Y_iY_k}} = \frac{d_R - d_{ik}}{d_R} = 1 - \frac{d_{ik}}{d_R} \tag{39}$$

which is always positive.

The covariance of the function of d_{ik}. The question arises what kind of covariance function should be selected. In the given derivation the relationship (34) is a linear one, with respect to distance. Another function $f(d_{ik})$ may be selected in which case, however, the d_R should be redefined so as to keep the correlation positive.

The simple relationship as derived gives satisfactory results within a region of some 30 x 30 km^2 or smaller.

In the reconnaissance use is made of (29). In proportion to larger distances, the constant c_R becomes negligible.

6.2. We may submit the substitute matrix to the covariance transformation towards the base points R_1R_2 used in (14). Using k and ℓ as general indices we get for any pair of points (see (10))

$$\begin{pmatrix} \overline{\sigma}_{X_kX_l} & \overline{\sigma}_{X_kY_l} \\ \overline{\sigma}_{Y_kX_l} & \overline{\sigma}_{Y_kY_l} \end{pmatrix} = \tag{40}$$

$$c^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -(F_{11})_k & -(F_{21})_k \\ -(F_{12})_k & -(F_{22})_k \\ (F_{11})_{k^{-1}} & (F_{21})_k \\ (F_{12})_k & (F_{22})_{k^{-1}} \end{pmatrix}^T \left(\begin{array}{cc|cc|cc} d_R-d_{kl} & 0 & d_R-d_{kR_1} & 0 & d_R-d_{kR_2} & 0 \\ 0 & d_R-d_{kl} & 0 & d_R-d_{kR_1} & 0 & d_R-d_{kR_2} \\ \hline d_R-d_{lR_1} & 0 & d_R & 0 & d_R-d_{R_1R_2} & 0 \\ 0 & d_R-d_{lR_1} & 0 & d_R & 0 & d_R-d_{R_1R_2} \\ \hline d_R-d_{lR_2} & 0 & d_R-d_{R_1R_2} & 0 & d_R & 0 \\ 0 & d_R-d_{lR_2} & 0 & d_R-d_{R_1R_2} & 0 & d_R \end{array}\right) \text{ times}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -(F_{11})_\ell & -(F_{12})_\ell \\ -(F_{21})_\ell & -(F_{22})_k \\ (F_{11})_{\ell^{-1}} & (F_{12})_\ell \\ (F_{21})_\ell & (F_{22})_{\ell^{-1}} \end{pmatrix}^{\cdot\cdot}_{\cdot\cdot}$$

where $(F_{11})_k = -(F_{22})_k = \frac{1}{d^2_{R_1R_2}}\left(x_{kR_2}x_{R_1R_2}+y_{kR_2}y_{R_1R_2}\right)$

and $(F_{12})_k = -(F_{21})_k = \frac{1}{d^2_{R_1R_2}}\left(x_{kR_2}y_{R_1R_2}-y_{kR_2}x_{R_1R_2}\right)$.

Substitution of the suffix ℓ gives corresponding expressions. The dimensions of c^2 are cm^2/km; those of $\sigma_{X_kX_\ell}$ and $\sigma_{X_kY_\ell}$ are cm, and those of $d_{k\ell}$ etc. are km.

Working this out, it is seen that

1) $\bar{\sigma}_{X_k X_\ell} = \bar{\sigma}_{Y_k Y_\ell}$ and $\bar{\sigma}_{X_k Y_\ell} = -\bar{\sigma}_{Y_k X_\ell}$;

2) if k = ℓ

$\bar{\sigma}^2_{X_k} = \bar{\sigma}^2_{Y_k}$ and $\bar{\sigma}_{X_k Y_k} = \bar{\sigma}_{Y_k X_k} = 0$,

so that the standard curves of the points remain circular;

3) Also the relative standard curves remain circular since

$\bar{\sigma}^2_{X_{k\ell}} = \bar{\sigma}^2_{y_{k\ell}} = \bar{\sigma}^2_{x_k} - 2\bar{\sigma}_{X_k X_\ell} + \bar{\sigma}^2_{X_\ell} = \bar{\sigma}^2_{Y_k} - 2\bar{\sigma}_{Y_k Y_\ell} + \bar{\sigma}^2_{Y_\ell}$

and $\bar{\sigma}_{X_{k\ell} Y_{k\ell}} = 0$;

4) The transformed variances/covariances are formations of the distances $d_{k\ell}$, d_{kR_1}, d_{kR_2}, $d_{\ell R}$, $d_{\ell R_2}$ and $d_{R_1 R_2}$ only. The reference distance d_R is eliminated (!);

5) The zero covariances of (40) do not necessarily remain equal to zero after the covariance transformation.

6) The radii of the standard circles in the reference points R_1 and R_2 are equal to zero, the further a point moves away from the basepoints the larger the radius of the standard circles becomes. This is according to the trend.

Important remark: The junction of networks as treated in section 4.1 is independent of the factor c^2 since it is eliminated in the coefficient of the misclosure t in (19). The proper estimation of c^2 is of great importance in the testing procedure of section 4.2.

7. Applications.

There are many possibilities to apply the described method of junction, some of which are mentioned below:

a. The junction of existing fields of points in some local coordinate system into the national coordinate system.

b. The junction of an existing field of points coordinated in the national system into some stations of which the coordinates have been redetermined.
   This is generally the case when control stations are observed in order to check the coordinates in relation to be permanent marking after a series of years. If it appears that the new coordinates have changed, the coordinates of the surrounding survey points must in consequence be submitted to a corresponding change in order to maintain the proper coherence.
   An application of this type has been made in 1976 for the network of regional main control points (the so-called 5^{th} order points) in an area of about 12 x 16 km^2. It is the task of the regional office of the Cadastral Survey in the Netherlands to establish and to maintain this network as an accurate densification of the first, second and third order national triangulation network, serving also surveys of general purposes. The results are published by the National Triangulation Board.
   A revision in 1976 of previous computations of the national triangulation combined with precise field measurements, undertaken by this Board, gave subsequently rise to a recomputed set of coordinates.
   It appeared that in the area mentioned above 8 of the 14 national triangulation points to which the densification previously was attached, obtained slightly but still significantly different coordinates.
   In order to avoid a complete recomputation of the 5^{th} order densification and for reasons of efficiency and of avoiding mistakes, the method of junction by adjustment was applied, using the substitute covariance matrix according to section 6.
   In this way the vectors of misclosure in the 14 national triangulation points (6 of which being zero) were distributed over the regional main control points in a manner, based on a model of statistical behaviour of densification measurements as described by the substitute covariance matrix. In addition it does not leave discrepancies in the total set of the registered coordinates of the 1^{st} to 5^{th} order (figure 3)

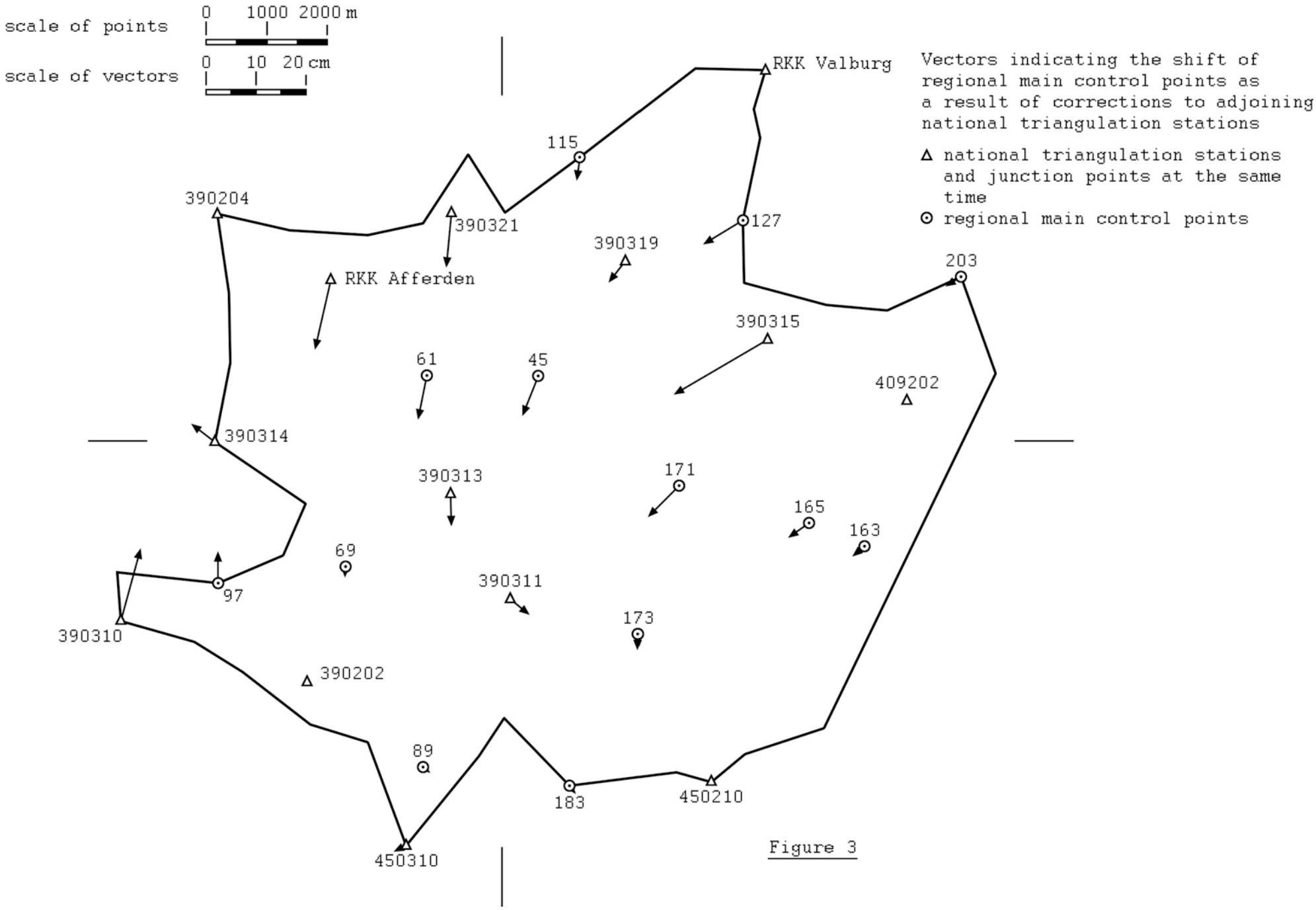

scale of points
0 1000 2000 m
scale of vectors
0 10 20 cm
Vectors indicating the shift of regional main control points as a result of corrections to adjoining national triangulation stations
national triangulation stations and junction points at the same time
regional main control points
RKK Valburg
115
390204
390321
127
390319
203
RKK Afferden
390315
61
45
409202
390314
390313
171
165
163
69
390311
97
173
390310
390202
89
183
450210
450310

Figure 3

control points.
In this way a homogeneous set of coordinates in the national coordinate system becomes available for all surveying institutes.

c. The junction of a photogrammetric field of points (the measured photogrammetric model including both minor control points and all the digitized details) to the coordinate system of the photogrammetric aerotriangulation.

d. The junction of remote sensing images (which generally contain system deformations and other distortions) to the coordinate system of existing topographical maps. The geometrical deformations can be determined at a series of reference points, after which the junction accomplishes the rectifying of all the features of the remote sensing image.

e. The junction of existing maps to a new mapping framework. This may for instance be the case where gridded maps of a good survey quality but mapped in a local coordinate system, have to be fitted into a national coordinate system. A similar junction problem arises when a set of existing maps without any grid have to be joined and provided with a grid system.
This question deals with the updating of a series of maps with irregular map boundaries to a series of maps with a fixed rectangular framework (Hektor [8]).
An application of this type has recently been made to the construction of large scale base maps ("GBKN", scale 1:1,000), where terrestrial cadastral surveying data of 1972 - available for part of an area in a digital form - had to be fitted into a photogrammetric (digital) mapping of the topographical terrain features to be displayed at the standardised large scale base maps.
At first one is able to correct the usually less accurate photogrammetric mapping of buildings (insufficient stereoview, roofs hanging over etc.); secondly a map overlay concerning cadastral boundaries may be manufactured, serving the users of the topographical oriented large scale base map within and outside the cadastral service.
The photogrammetric mapping required a set of base points, leading to the establishment of a new control survey network over the total area of the planned base maps in 1979.
The digital cadastral coordinated data of detail points (in the subsidiary system), were determined in 1972 by means of a densification network according to the national standards.

For the comparison of terrestrial and photogrammetric data and the construction of a joint data bank for the large scale base maps, a junction of coordinate systems is needed.

For this purpose 45 control points surrounding the area were selected for connection to the new control survey network (principal system), resulting in a set of 45 junction points having coordinates known in both systems. To these points a set of 2,381 detail points (with respect to boundaries and buildings) in the subsidiary system has been attached.

The junction was accomplished by Buiten by the method of junction by adjustment using the substitute covariance matrix according to section 6.

This misclosure vectors (after a provisional junction of the two nets at a selected coordinate reference base) ranged between 0 and 33 cm. These misclosures were assumed to be of a stochastical nature, so that the proposed junction was justified.

The testing of the junction points by data snooping showed that by taking the constant c = 6 for the subsidiary system and c = 3 for the primary system, none of the points needed to be rejected.

Moreover the estimation of the variance factor in the variance ratio test was accepted under these assumptions.

The resulting coordinates of the detail points after junction were stored on magnetic tape and delivered to the Office of the Cadastral Surveys in the region of the planned base maps.

There the updating of buildings and the construction of cadastral overlay data was prepared with the help of a M. and S. interactive graphics computer system.

Though the activities have not yet come to a conclusion it can be said that the results of junction are of great value for the final mapping of buildings and for the data snooping of mistakes in the photogrammetric mapping of terrain features.

In addition a map overlay of cadastral features is obtained in the same system as the base map. This is of great importance for the comparison of topographic and cadastral details.

Without doubt there will be more possibilities for practical application than mentioned here.

REFERENCES

1. ALBERDA, J.E.: "A Review of Analysis Techniques for Engineering Survey Control Schemes." Paper presented at the Industrial and Engineering Survey Conference, London, Sept. 1980.

2. BAARDA, W.; de GROOT, D.; HARKINK, F.: "Manual of Cadastral Surveys". (Dutch), 1956.

3. BAARDA, W.: "A Testing procedure for use in geodetic networks." Neth. Geod. Comm. New Series 2.5, Delft, 1968.

4. BAARDA, W.: "S-transformations and Criterion matrices." Neth. Geod. Comm. New Series 5.1, Delft, 1973.

5. BUITEN, H.J.: "Junction of networks by adjustment." (Dutch) Neth. Geod. Per. 5.10, Dec. 1975, p. 181.

6. BUITEN, H.J.; SCHOK, H.A.: "Junction of existing control survey networks by adjustment." (Dutch) Geodesia, Vol. 18, Sept. 1976, p. 219.

7. BUITEN, H.J.: "Junction of networks by collocation." Manuscripta Geodaetica, Vol. 3, 1978, p. 253-297.

8. HEKTOR, E.H.: "Datenerfassung für automatische Verarbeitung und Zeichnung einschließlich der Probleme der Digitalisierung." Ber. XV, FIG Kongr. (1977), Nr. 514.1.

9. RICHARDUS, P.: "Project Surveying." Manuscr., 2nd revised edition, 1981.

A CRITERION MATRIX FOR THE SECOND ORDER DESIGN OF CONTROL NETWORKS

Fabio CROSILLA
Trieste, Italy

ABSTRACT

This paper proposes an original method of constructing a criterion matrix for the optimal design of control networks by means of the contraction of the eigenvalues and the rotation of the eigenvectors of a covariance matrix. A Second Order Design problem is then resolved, that is the optimization of the precision of the observations of a local free distance network to be constructed for the study of recent crustal movements in the seismogenetic area of Friuli (Italy).

1. Introduction.

Sprinsky (1978) suggested a method, subsequently developed by Wimmer (1981), of constructing a criterion matrix by means of the reduction of the trace of the covariance matrix Q_{xx} of the adjusted net coordinates vector x. It involves proceeding to a singular value decomposition of a (mxm) covariance matrix $Q_{xx} = V \Lambda V'$ with $r(Q_{xx}) \leq m$; where Λ is a diagonal matrix whose terms correspond in descending order to the eigenvalues of the matrix Q_{xx}, and V corresponds to the orthonormalized eigenvectors. The matrix Q_{xx} can be interpreted geometrically as an m-dimensional error ellipsoid in which the length and direction of the semi-axes correspond to the square root of the eigenvalues and the eigen-vectors respectively.

Reducing the dimensions of this ellipsoid by a contraction of the greater eigenvalues, it is possible to obtain a higher global precision. Wimmer (1981) thus proposed to consider as a criterion matrix, the matrix $\tilde{Q}_{SVD} = V \tilde{\Lambda} V'$ obtained from the covariance matrix Q_{xx} whose greater eigenvalues (λ_i) are reduced by a parameter of contraction t

$$\tilde{\lambda}_i = \lambda_i - t\,(\lambda_i - \lambda_r) \quad , \; 0 \leq t \leq 1 \quad \text{and where} \quad r = r\,(Q_{xx}) \qquad (1)$$

Sprinsky mentioned the possibility of a redistribution of the allocated variances through a rotation of the m-dimensional error ellipsoid by

means of a change of the orthonormal basis of the vector· space. The rotation procedures imply the definition of a class of covariance matrices similar to the matrix of the contracted eigenvalues $\tilde{\Lambda}$.
First, this paper presents two methods of rotation of the error ellipsoid for the construction of a criterion matrix for control networks. This criterion matrix is then used for the solution of a S.O.D. problem in a network to be constructed in the seismogenetic area of Friuli. Finally the limits to the application of this criterion matrix are defined.

2. Definition of a type of criterion matrix for control networks

Let $\underline{d}$ be defined as the deformation vector of the net coordinates characterized by a covariance matrix Q_{dd} and given by $\underline{d} = \underline{x}_1 - \underline{x}_2$, where $\underline{x}_1$ and $\underline{x}_2$ are the vectors of the coordinates relative to two different periods of measurement. On the hypothesis that the observations and their precision are the same for both periods, according to the law of variance propagation the covariance matrix of the deformation vector $\underline{d}$ will be given by:

$$Q_{dd} = 2Q_{xx} = 2(V \Lambda V') = 2\left(\sum_{i=1}^{r} \lambda_i v_i v'_i\right) = \sum_{i=1}^{r} 2\lambda_i v_i v'_i \qquad (2)$$

where v_i are the eigenvectors related to the λ_i eigenvalues.
The matrices Q_{xx} and Q_{dd} have the same eigenvectors and the eigenvalues are different by a factor of 2.
Regarding the possibility of a rotation of the error ellipsoid mention has been made above of a class of covariance matrices similar to that of the contracted eigenvalues $\tilde{\lambda}$. In this class could be considered as a criterion matrix the matrix for which the components of the "essential eigenvectors" relative to the pairs of variables x_i, y_i (i = l...n) (Pelzer, 1976, 1980) (Dupraz and Niemeier, 1979) are disposed in a direction as orthogonal as possible with respect to the predicted deformation. This criterion should be satisfied in particular by the components of the essential eigenvector, that is the eigenvector relative to the greatest eigenvalue. It therefore represents the direction of the greatest semi-axis of the error ellipsoid of the coordinate vector $\underline{x}$ and, from

what has been said above, of the deformation vector $\underline{d}$. In the direction defined by the essential eigenvector possible deformations $\underline{d}$ thus cannot be established with any great precision.

Furthermore, in the case of non circular error ellipses, there is generally an isodirectionality between the greatest semi-axis of the ellipse and the components x_i, y_i (i = l...n) of the essential eigenvector. The greater the difference between the semi-axes of the ellipse, the more evident this isodirectionality becomes. From this it follows that considering as a criterion matrix the covariance matrix characterised by essential eigenvector components which are as orthogonal as possible to the direction of the predicted deformation, error ellipses are obtained in such a way that their greatest semi-axes are also disposed in a direction as orthogonal as possible to that of the predicted deformation. The probability P (χ^2) that the real values of the coordinates are contained in the area defined by the dimension of the ellipse, may be applied to all error ellipses. From this derives the necessity that the greatest semi-axis of the ellipse be orthogonal to the predicted direction of deformation.

3. Construction of the criterion matrix

The singular value decomposition of the covariance matrix Q_{xx} (mxm), $r(Q_{xx}) \leq m$, relating to a free net in its eigenvalues and eigenvectors may be considered thus:

$$Q_{xx} = \left[\begin{array}{c|c} & \\ V & U \\ & \end{array}\right] \left[\begin{array}{c|c} \Lambda & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} V' \\ \hline U' \end{array}\right] \qquad (3)$$

The criterion matrix proposed by Wimmer (1981) follows from:

$$\tilde{Q}_{SVD} = \left[\begin{array}{c|c} & \\ V & U \\ & \end{array}\right] \left[\begin{array}{c|c} \tilde{\Lambda} & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} V' \\ \hline U' \end{array}\right] \qquad (4)$$

where: $\tilde{\Lambda}$ is a matrix of eigenvalues $\tilde{\lambda}_i$ (i = l...r);
$\tilde{\lambda}_i$ are calculated in formula (1);
V (mxr) is a matrix of eigenvectors relative to Λ and $\tilde{\Lambda}$;
U (mxm-r) is a matrix of eigenvectors relative to m-r eigenvalues = 0.

The orthogonality of the essential eigenvector components with respect

to the predicted direction of deformation may be achieved by rotating the eigenvector matrix V in an orthonormal matrix V^o in which the elements of the first column correspond to the components required by the essential eigenvector. The construction of the matrix V^o can be carried out in two different ways:

- by means of independent rotations of all the r eigenvector component pairs v_{xij}, v_{yij} (j = 1...r) relative to the n pairs of variables x_i, y_i (i = 1...n);
- by means of an orthogonal procrustean transformation of the matrix V.

In the first case any pair of components of the eigenvectors v_{xij}, v_{yij} (j 1...r) relative to the variables x_i, y_i is rotated through an angle $\varphi_{x_i y_i}$ by means of

$$\begin{bmatrix} v^o_{xij} \\ v^o_{yij} \end{bmatrix} = \begin{bmatrix} \cos\left(\varphi_{x_i y_i}\right) & \sin\left(\varphi_{x_i y_i}\right) \\ -\sin\left(\varphi_{x_i y_i}\right) & \cos\left(\varphi_{x_i y_i}\right) \end{bmatrix} \begin{bmatrix} v_{xij} \\ v_{yij} \end{bmatrix} \qquad (5)$$

where the angle $\varphi_{x_i y_i}$ is understood as > 0 if the direction is clockwise.

In this way the introduction of new components v^o_{xij}, v^o_{yij} in the original matrix of the eigenvectors V does not invalidate the properties of normality and orthogonality of the matrix itself. The criterion matrix derived from this method of rotation of the eigenvector components relative to the variables x_i, y_i is given by:

$$\tilde{Q}_{IR} = V^o_{IR}\, \tilde{\Lambda}\, V^{o'}_{IR} \qquad (6)$$

The construction of the matrix V^o can also be achieved through an orthogonal procrustean transformation of the matrix V. In the factor analysis "procrustean transformation" is understood to mean any linear transformation which under certain specified conditions allows the transformation of a given matrix into a matrix as near as possible equal to a preconstructed one.

For example, let V and V^o be two matrices (mxr) $m \geq r$

where: V is an original orthonormal eigenvector matrix;

V^o is a preconstructed matrix, not necessarily orthonormal, containing the estimates of k ($1 \leq k \leq r$) rotated eigenvectors. For k = 1 the rotated eigenvector estimate corresponds to the required essential eigenvector one.

The problem is to find an orthogonal transformation matrix T (rxr) such

that $$VT \cong V^o \quad (7)$$

where the approximation $VT \cong V^o$ is a least square approximation and the matrix VT is an orthonormal matrix. To calculate the matrix T which transforms the matrix V into a least squares approximation of V^o, the sum of the squares of the elements of the matrix $E = (V^o - VT)$ that is

$$tr\,(E'\,E) = tr\,((V^o - VT)'\,(V^o - VT)) \quad (8)$$

must be minimal under the condition of orthogonality of T so that

$$(TT' - I) = 0 \quad (9)$$

Now in accordance with Lagrange's method the minimum condition of tr(E'E) under the condition that (TT' - I) = 0 is given by

$$\frac{\partial\, tr(E'E)}{\partial T} + \frac{\partial\left(tr\,\Theta(TT'-I)\right)}{\partial T} = 0 \quad (10)$$

where Θ is a matrix of Lagrangian multipliers.

Developing the matrix derivatives of the matrix traces (Schonemann, 1965) the following equation is obtained

$$2\,V'VT - 2\,V'V^o + 2\,\Theta\,T = 0 \quad (11)$$

Dividing both parts by 2 and multiplying both parts by T' the result is $$V'V + \Theta = V'V^oT' \quad (12)$$

Now given that V'V and Θ are both symmetrical $V'V^oT'$ must also be symmetrical. It thus follows that $$V'V^oT' = TV^o'V \quad (13)$$

that is $$V'V^o = TV^o'VT \quad (14)$$

Let $R = V'V^o$ (15), therefore $R = TR'T$ (16)

Let two orthonormal matrices be defined P and Q (rxr) calculated from the singular value decomposition of RR' and R'R

$$RR' = PDP' \quad (17a) \qquad R'R = QMQ' \quad (17b)$$

where D and M are the diagonal matrices of the eigenvalues of RR' and R'R respectively.

Now RR' and R'R have the same eigenvalues (Johnson, 1963), therefore D = M. Putting the terms of (16) in (17a) and using (17b) it results that

$$RR' = TQDQ'T' = PDP' \quad (18)$$

clearly TQ = P, that is to say $$T = PQ' \quad (19)$$

This method is also valid for the case in which the matrices V and V^o are not of full rank. Moreover, in the particular case where the matrix D contains roots equal to each other and different from zero the matrices

P and Q will not be unique and consequently T will not be unique either.

To satisfy the condition that the trace tr (E'E) = minimum, the matrices P and Q will also have to satisfy the following condition (Schonemann, 1966):

$$P'RQ = D^{1/2} \qquad (20)$$

where $R = V'V^o$ and $D^{1/2}$ is the matrix of the square root of the eigenvalues of D. Now let

$$H = P^{o\prime}RQ^o \qquad (21)$$

where P^o and Q^o are two orthonormal matrices which satisfy only (17). The matrix H (Schonemann, Bock, Tucker, 1965) is diagonal with diagonal elements of $D^{1/2}$ exept for the blocks H_j of elements of order $n_j x n_j$ along the diagonal, corresponding to multiple roots λ_j of multiplicity n_j in D. Since the matrices P^o and Q^o satisfy (17a) and (l7b) the matrix H_j has the property that $H'_j H_j = D_j$ and in turn H'H = D.

Each of these square blocks H_j can be considered proportional by a scalar λ_j to an orthonormal matrix W_j ($n_j x n_j$). H_j can therefore be decomposed into

$$H_j = \lambda_j^{1/2} W_j = D_j^{1/2} W_j \qquad (22)$$

Let $P = P^o$ in (20). To find a matrix Q which satisfies (20), on the basis of (22), matrix H is multiplied by a diagonal matrix K where

$$W^o = KH \qquad k_{ii} = \begin{cases} \lambda_i^{-1/2} & \lambda_i > 0 \\ 0 & \lambda_i = 0 \end{cases} \quad (i = 1 \ldots r) \qquad (23)$$

from which is obtained a matrix W^o which can in its turn be transformed into an orthonormalised matrix W inserting 1s in the positions of the principal diagonal of W^o which contain 0s.

Finally the matrix Q can be obtained from

$$Q = Q^o W' \qquad (24)$$

This general solution to the problem of procrustean transformation, even in the case where the matrices V and V^o are of rank $r < m$, gives a notable flexibility to the definition of the preconstructed matrix V^o.

In this regard it is sufficient to define the estimates of k eigenvectors ($k \geq 1$) and to insert zeros in the residual r - k columns of V^o. For k = 1 the estimates of the essential eigenvector components correspond.

The matrix VT which results from the orthogonal procrustean transformation of V to approximate to V^o, will contain estimated orthonormalised components of the essential eigenvector in its first column and orthonormal r - 1 vectors in the residual ones.

The criterion matrix resulting from this method of procrustean transfor-

mation of the eigenvector matrix V is finally given by

$$\tilde{Q}_{PT} = V^{0}_{PT}\, \tilde{\Lambda}\, V^{0'}_{PT} \qquad (25)$$

4. Solution of a Second Order Design problem for a control network

The S.O.D. problem consists in designing a matrix of the observation weights P (sxs) in such a way that the covariance matrix of the coordinates Q_{xx} (mxm) resulting from $(A'PA)^{-}$, where A (sxm) is the design matrix and $()^{-}$ is a generalised inverse, is equivalent to a criterion matrix Q_{xx} (Baarda, 1973; Grafarend, 1972; Grafarend, Schaffrin, 1979). The generalised inverse corresponds to a Cayley inverse $()^{-1}$ for constrained nets and to a Moore Penrose inverse $()^{+}$ for a minimum norm solution of a free net. For the solution of the S.O.D. problem the Kronecker product (Bossler et al., 1973) was proposed for correlated observations and the Katri Rao product (Rao, Mitra, 1971) for uncorrelated observations.

Starring from the matrix equation $(A'PA)^{+} = Q_{xx}$ the extraction of the observation weights matrix P requires first of all a decomposition $K = (A'PA)^{+} A'P$.

This, applying the Katri Rao product gives $(K \odot K)$ vecd P^{+} = vec Q_{xx} where $(K \odot K)$ is of m^2xs dimension, vecd P^{+} is of sxl dimension and vec Q_{xx} is of m^2xl dimension. Since the matrix Q_{xx} is symmetrical it is sufficient to consider only its lower or upper triangle.

The reduced matrix equation is given by $(K \underline{\odot} K)$ vecd P^{+} = vech Q_{xx} with $(K \underline{\odot} K)$ of gxs dimension, vech Q_{xx} of gxl dimension and where $g = {}^{m(m+1)}/_{2}$. Let 0 be a gxg diagonal weight matrix in which $o_{ff} = 0.5 (f=1...g)$ if the f^{th} component of vech Q_{xx} corresponds to a diagonal element of Q_{xx}; $o_{ff} = 1$ otherwise (Wimmer, 1978). The solution for vecd P^{+}, whenever g > s, is given by

$$\text{vecd } P^{+} = ((K \underline{\odot} K)'0\ (K \underline{\odot} K))^{+}\ (K \underline{\odot} K)'0 \text{ vech } Q_{xx} \qquad (26)$$

from which vecd P = (vecd $P^{+})^{+}$.

This calculation method is iterative (Wimmer, 1981) since the matrix of the weights is contained in the decomposition K. The calculation must therefore be repeated until the matrix of the weights no longer varies.

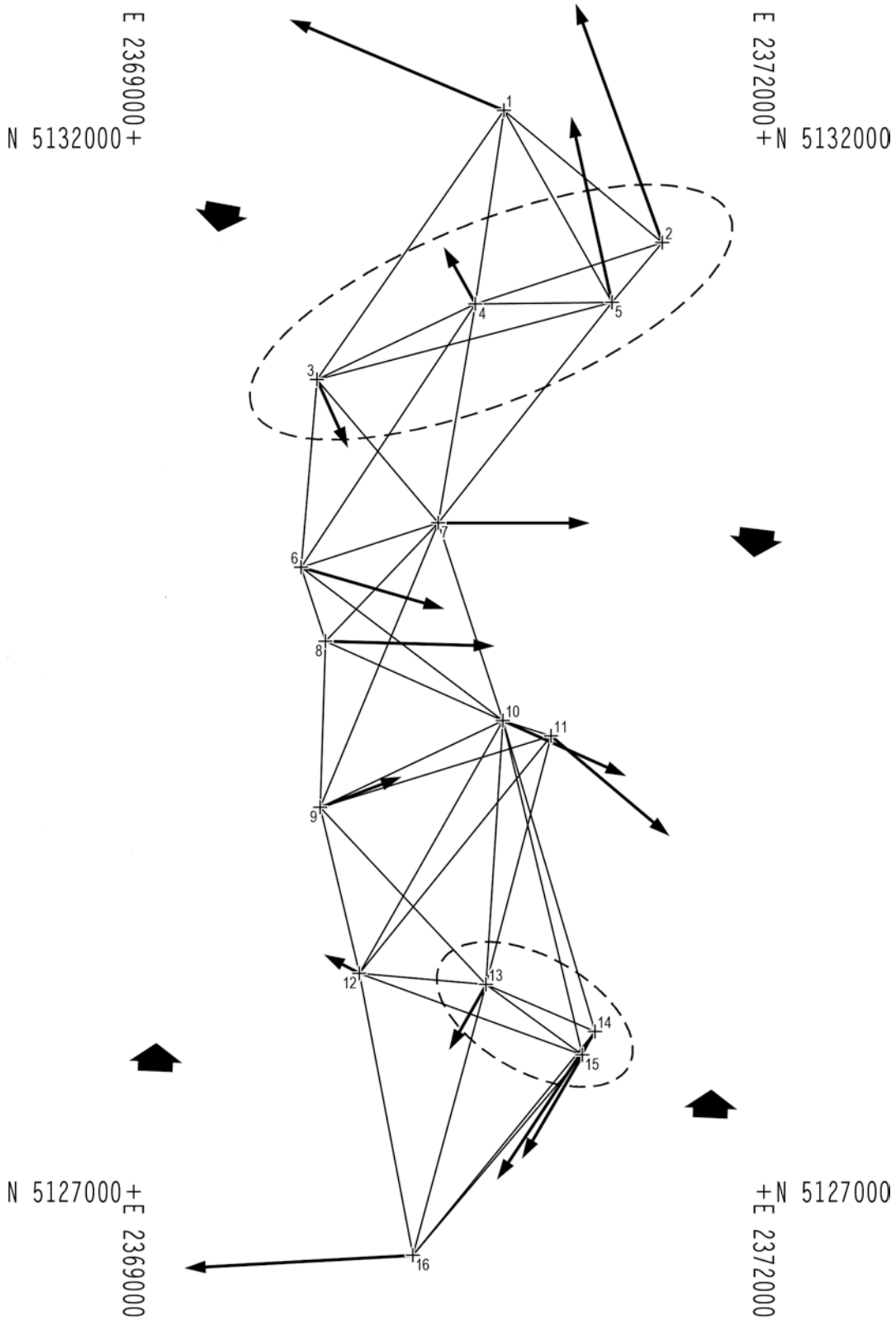


Fig. 1 Control network design to be constructed in the seismogenetic area of Friuli (Lake Cavazzo valley). The original essential eigenvector components and the predicted direction of ground deformation are also shown.

The net for which a S.O.D. problem is solved is a pure trilateration net made up of 16 vertices and 42 distances, to be constructed in the Lake Cavazzo Valley for the study of recent crustal movements in the seismogenetic area of Friuli (Crosilla, Marchesini, 1982).

Fig. 1 shows the design of the net and the essential eigenvector components relative to the coordinates of each point obtained from the singular value decomposition of the covariance matrix of the coordinates Q_{XX} calculated with the unit weight matrix of the observations . The predicted direction of ground deformation is also shown.

Previous considerations have shown that in the control nets the components of the essential eigenvector must be orthogonal with respect to the direction of predicted movement. It can thus be observed from fig. 1 that in the case of this net it is necessary to rotate the essential eigenvector components of two distinct groups of points.

The first group comprises points 2, 3, 4, 5 and the second group points 13, 14, 15. The essential eigenvector components relative to the points of the first group were rotated each time by -10 gon, -20 gon, -30 gon, -40 gon, and those of the second group of points by +10 gon, +20 gon, +30 gon, +40 gon. Two criterion matrices were constructed for each of the 5 rotations. One was by means of independent rotation $\tilde{Q}_{IR}$ and the other by means of procrustean transformation $\tilde{Q}_{PT}$, after contracting the greatest eigenvalue of the covariance matrix Q_{XX} through the parameter of contraction t = 0.5 (Wimmer, 1981). A S.O.D. problem was then resolved in each case with the iterative method reported above.

The precision of the observations obtained after one rotation of -10 gon of the essential eigenvector components of the first group of points and +10 gon of the components of the second group of points are shown in figs. 2a and 2b for $\tilde{Q}_{IR}$ and $\tilde{Q}_{PT}$ respectively.

The error ellipses $(P(\chi^2) > 0.99)$ for the 16 points of the net are shown in fig. 3 and 4. They were obtained from a covariance matrix calculated with a unit weight matrix of the observations (dotted line), from the criterion matrix $\tilde{Q}_{IR}$ in fig. 3 and $\tilde{Q}_{PT}$ in fig. 4 (thin line) relative to a rotation of ± 10 gon, and from the covariance matrix calculated

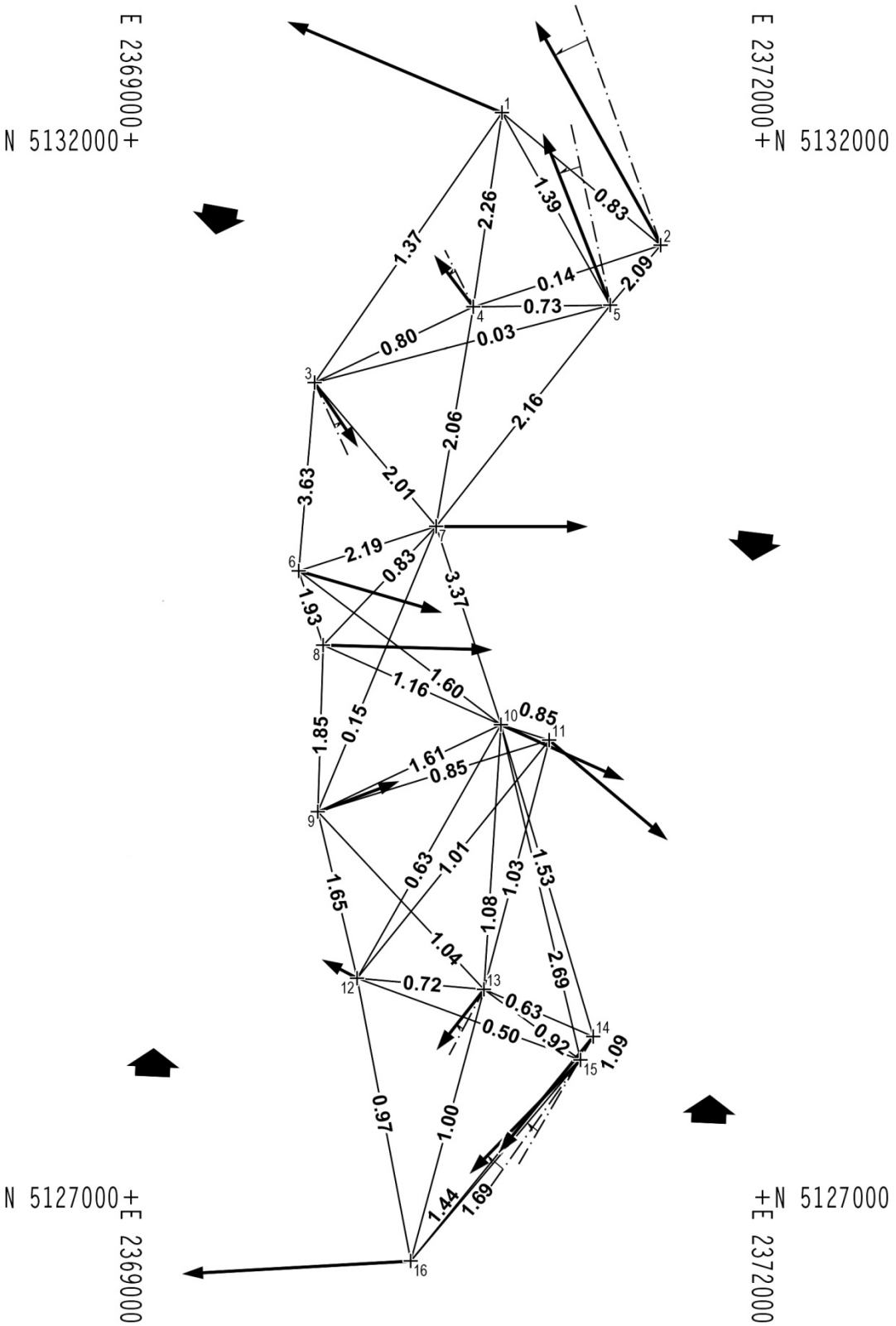


Fig. 2a Weight distribution of the observations calculated by the Second Order Design considering the matrix $\tilde{Q}_{IR}$ as a criterion matrix.

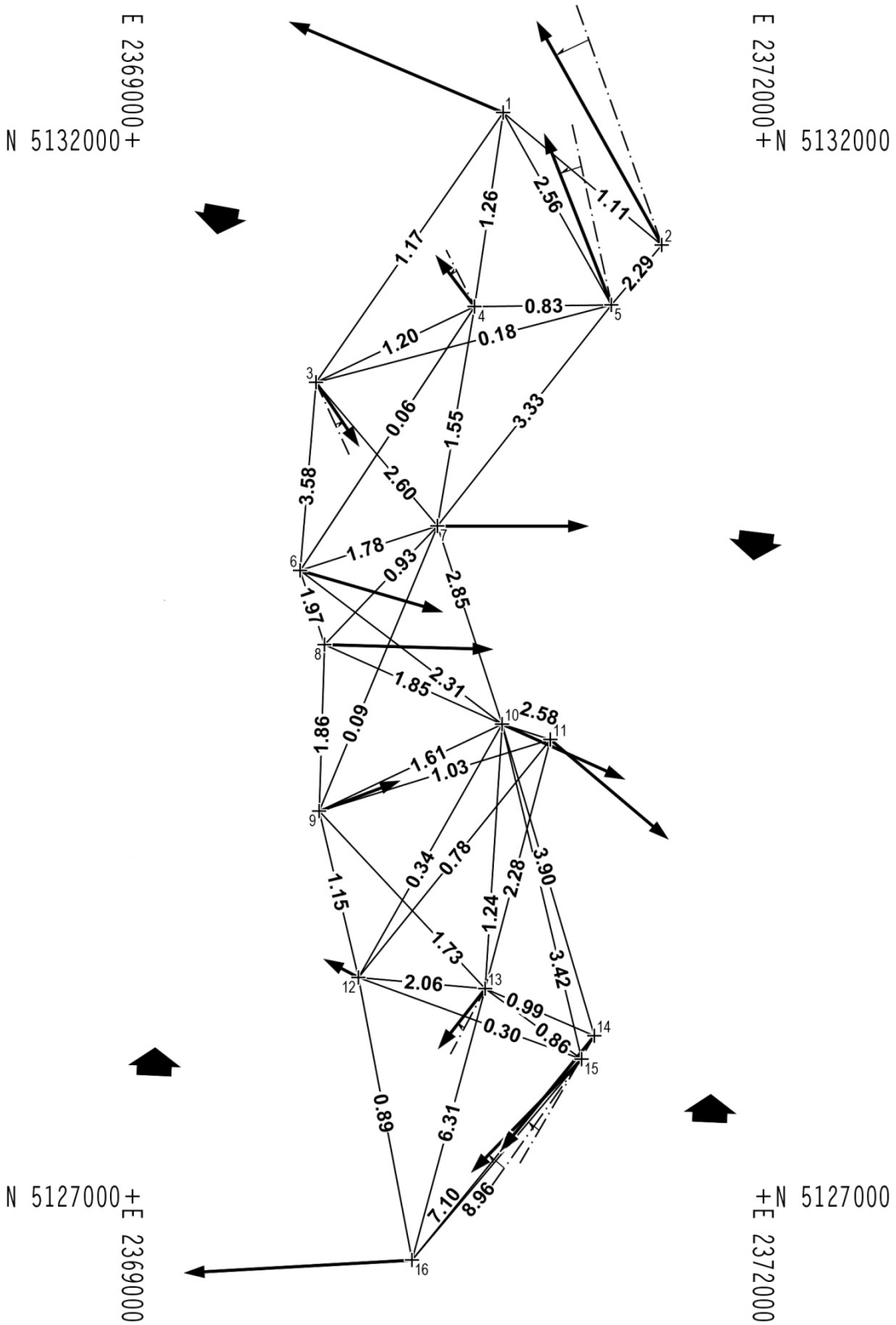


Fig. 2b Weight distribution of the observations calculated by the Second Order Design considering the matrix $\tilde{Q}_{PT}$ as a criterion matrix.

with the weights resulting from the S.O.D. (thick line). As can be seen in figs. 3 and 4 there is a high level of correspondence between the error ellipses postulated by the criterion matrix and the ones obtained. It can also be seen that the ellipses of the points 2, 3, 4, 5 and 13, 14, 15, actually undergo the rotation imposed on the eigenvectors in the construction of the criterion matrix.

Finally it can be seen that the results obtained with the two methods of rotation are substantially identical.

The solution to the S.O.D. problem for criterion matrices obtained with rotations of ± 20 gon, ± 30 gon, ± 40 gon of the essential eigenvector components does not give satisfactory results. In fact the weights of many observations are often prone to be negative and weights which are very different from each other often result in the other observations. In the case of the rotation of ± 40 gon, for example, 7 observations with negative weights are obtained from the criterion matrix $\tilde{Q}_{IR}$ with a ratio pmax/pmin equal to 21.16/0.10 = 211.6.

5. Conclusion

The results obtained from the solution of this S.O.D. problem confirm the validity of the method, suggested by Sprinsky and used by Wimmer, for the construction of the criterion matrix $\tilde{Q}_{SVD}$, provided that the contraction parameter of the eigenvalues of the covariance matrix is not taken to be too high.

The criterion of rotation of the essential eigenvector components introduced in this paper also make it possible to improve the definition of a criterion matrix for control nets. The results have made it clear, however, that rotations of this type must be limited. Rotations of great amplitude give physically unreal criterion matrices.

Finally, the two methods of rotation here proposed, that is independent rotation and procrustean transformation, give substantially identical results for limited rotations of the eigenvectors.

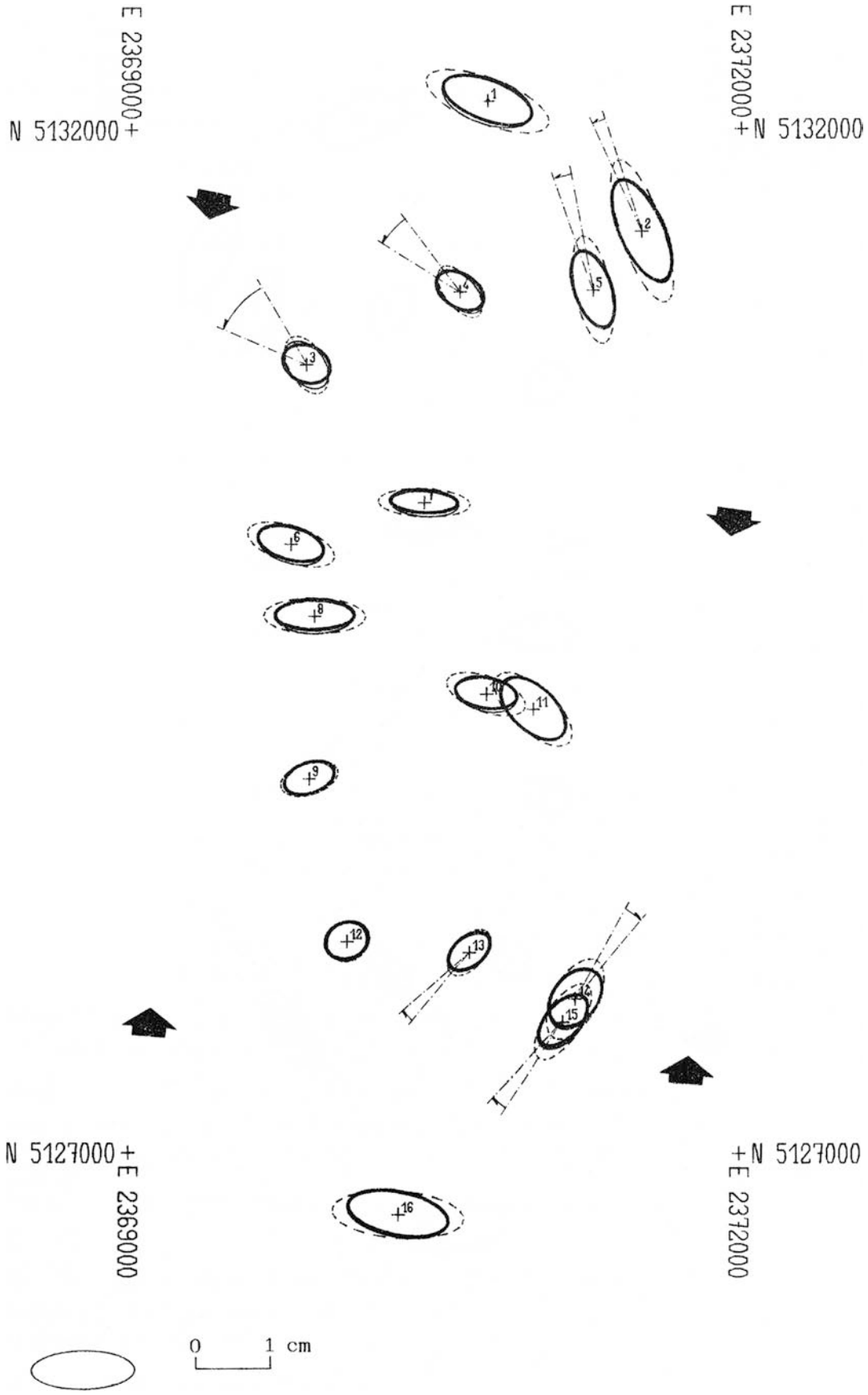


Fig. 3 Error Ellipses obtained from a unit weight matrix of the observations (dotted line), criterion matrix $\tilde{Q}_{IR}$ (thin line) and the S.O.D. covariance matrix (thick line).

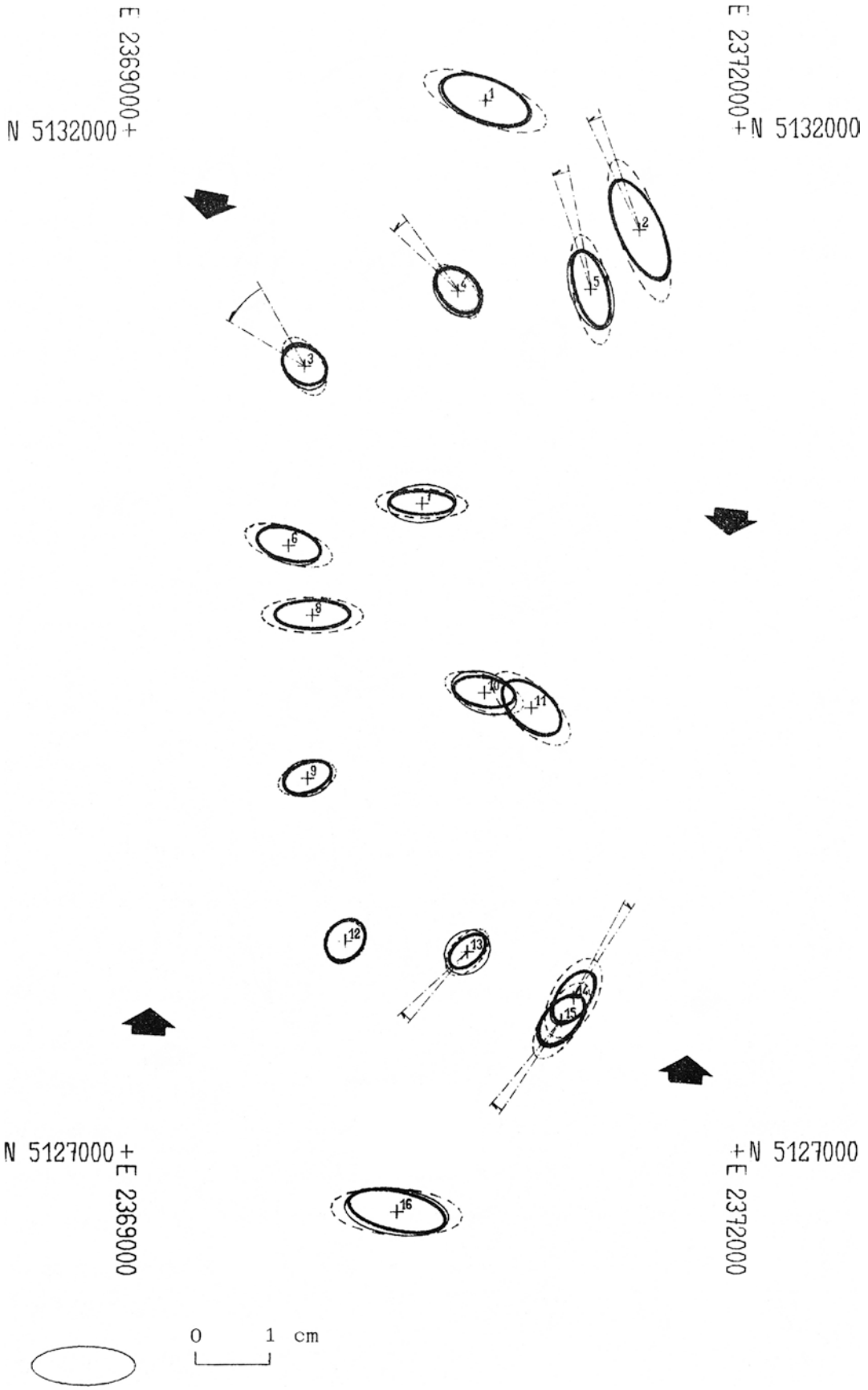


Fig. 4 Error Ellipses obtained from a unit weight matrix of the observations (dotted line), criterion matrix $\tilde{Q}_{PT}$ (thin line) and the S.O.D. covariance matrix (thick line).

REFERENCES

BAARDA, W., 1973: S-transformations and criterion matrices, Neth. Geod. Comm., Publ. on Geod., New Series 5, n° 1, Delft (1973).

BOSSLER, J., GRAFAREND, E., KELM, R., 1973: Optimal design of geodetic nets II, Journal Geophysical Research, 78 (1973), pp. 5887-5897.

CROSILLA, F., MARCHESINI, C., 1982: Geodetic network optimization for the detection of crustal movements using a mekometer, Bollettino di Geodesia e Sc. Affini, to appear.

DUPRAZ, H., NIEMEIER, H., 1979: Un critère pour l'analyse des résaux géodésiques de controle, Mensuration, Photogrammetrie, Génie rural, 4 (1979), pp. 70-76.

GRAFAREND, E., 1972: Genauigkeitsmaße geodätischer Netze, Deutsche Geodätische Kommission, Reihe A, n° 73 (1972).

GRAFAREND, E., SCHAFFRIN, B., 1979: Kriterion Matrizen I, Zeitschrift für Vermessungswesen, 104 (1979), pp. 133-149.

JOHNSON, R.M., 1963: On a theorem stated by Eckart and Young, Psychometrika, 28 (1963), pp. 259-263.

PELZER, H., 1976: Genauigkeit und Zuverlässigkeit geodätischer Netze. Tagung, Mathematische Probleme der Geodäsie, Oberwolfach (1976).

PELZER, H., 1980: Some criteria for the accuracy and the reliability of networks, Deutsche Geodätische Kommission, Reihe B, n° 252 (1980).

RAO, C.R., MITRA, S.K., 1971: Generalized inverse of matrices and its applications. New York/London/Sydney/Toronto, (1971).

SCHONEMANN, P.H., 1965: On the formal differentiation of traces and determinants. Research Memorandum, n° 27, Chapel Hill, N.C.: The Psychometric Laboratory, (1965).

SCHONEMANN, P.H., 1966: The generalized solution of the orthogonal procrustes problem. Psychometrika, n° 31, (1966), pp. 1-16.

SCHONEMANN, P.H., BOCK, R.D., TUCKER, L.R., 1965: Some notes on a theorem by Eckart and Young. Research Memorandum, n° 25, Chapel Hill, N.C.: The Psychometric Laboratory, (1965).

SPRINSKY, W.H., 1978: Improvement of parameter accuracy by choice and quality of observation, Bollettin géodésique, 52 (1978), pp. 269-279.

WIMMER, H., 1978: Eine Bemerkung zur günstigsten Gewichtsverteilung bei vorgegebener Dispersionsmatrix der Unbekannten. Allgemeine Vermessungs-Nachrichten, 85 (1978), pp. 375-377.

WIMMER, H., 1981: Second Order Design of geodetic networks by an iterative approximation of a given criterion matrix. Paper presented at the Symposium on Geodetic Networks and Computations, Munich, August 31 - September 5, (1981).

PROCEDURES FOR THE FIRST AND SECOND ORDER DESIGN OF VERTICAL CONTROL NETWORKS

Paul CROSS and Ali FAGIR
London, England

ABSTRACT

Optimal network design problems can be posed in many ways but for the special case of level networks the following is most common.

GIVEN (a) the approximate position of the required bench marks,

(b) a list of possible level routes,

(c) the precision expected of the field parties and

(d) the intended use of the network.

FIND The set of level routes whose observation will yield a network suitable for the intended purpose with the minimum of cost (i.e. minimum total length of levelling).

The above problem is of course, an example, of the now standard "first order design". To date no direct solution to this problem has been found but various iterative strategies using the second order design procedure (i.e. solution for observation weights and subsequent discard of observations with small weights) have been employed.

This paper critically reviews second order design procedures that have been applied to the design of vertical control networks and gives details of two sets of experiments with a new method which has a number of advantages including not needing the explicit formation of a criterion matrix.

1. Introduction

Grafarend (1974) has defined the first and second order design problems as

First Order Design: The selection of the configuration of observations.

Second Order Design: The choice of observation weights.

Mathematically we can say that the two design problems are equivalent to solving for A and W respectively in the well known equation

$$\left(A^T W A\right)^{-1} = C_{\hat{x}} \qquad (1)$$

where A is the design matrix, W the weight matrix (i.e. inverse of the covariance matrix of the observations) and $C_{\hat{x}}$ the required covariance matrix of the least squares estimates of the station coordinates (often called a criterion matrix).

In the case of levelling networks the first order design problem is far more important than the second order design problem. This is because, in practice, we rarely have a choice as to the precision of the measuring process and hence cannot tailor our field methods and instrumentation to fit a required weight matrix W. Note the contrast with two-dimensional networks where, for example, there is a large range of distance measuring instruments all with different precisions. Hence almost all level network design problems are first order and can be posed as follows:

GIVEN

(a) a set of stations whose heights are to be found,

(b) a set of all possible level routes,

(c) the required precision of the least squares estimates of the station heights, and

(d) the required reliability of the measurements.

FIND

The subset of level routes that will satisfy the precision and reliability criteria with the minimum of cost, i.e. find the optimum level network.

Note that, in the case of levelling, costs are simple to model as they are proportional to the length of level route and the design problem can be restated as

FIND

The shortest subset of level routes that will satisfy the precision and reliability criteria.

Unfortunately there is no known analytical solution to the first order design problem so an "indirect" method needs to be employed. There are two possibilities: computer simulation and second order design.

Computer simulation

(i) "Guess" a possible network configuration

(ii) compute precision and reliability criteria

(iii) if unsatisfactory go to (i)

(iv) compute total length of levelling

(v) go to (i)

(vi) The process is stopped when the network with the shortest total length of levelling appears to be found.

The obvious disadvantages of the method are that the optimum network may never be found and also a very large amount of work may be involved. Cross and Whiting (1981) have, however, found a way of automating the process that seems to counter both these disadvantages. The difficulty with their technique is that it relies, for its claim of optimality for any given problem, on an extrapolation made from exhaustive, time consuming tests, on a series of small (up to twelve stations) networks. There is no theoretical proof that the technique will always give the optimum solution. Nevertheless it is probably the most practical of the currently available techniques and a number of researchers are working in this area, e.g. Niemeier and Rohde (1981).

Second order design

(i) Simulate a network with every possible observation

(ii) solve the second order design problem

(iii) discard all observations whose required weights are small (variance large)

(iv) go to (ii)

(v) stop when reliability criteria are reached or when required precision can only just be met, i.e. variances better than those required could not be achieved with the available instrumentation.

Note that stage (iii) above is justified by the fact that observations with very low variances do not significantly contribute to the precision or reliability of the network.

This paper is concerned with techniques for the second order design of levelling networks that can be conveniently employed in the above iterative process. A critical review of the methods that have been used will be given followed by the results of tests with two new methods. Note that the emphasis will be on precision rather than reliability so first some remarks will be made on precision criteria suitable for levelling networks.

2. Precision criteria for levelling networks

It is extremely rare for absolute height values, derived from levelling networks, to be of interest. What are of more practical use are height differences derived from the least squares estimates of the station heights. Hence the variances of these derived quantities are the most useful precision criteria when designing levelling networks. Usually we would require that the variance of a height difference be less than some specified function of the distance between the points, i.e.

$$\sigma^2_{\Delta h} = f(d_{ij}) \qquad (2)$$

where $\sigma^2_{\Delta h}$ is the variance of the height difference of two stations a distance d_{ij} apart. Occasionally, as pointed out by Cross and Whiting (1980), $\sigma^2_{\Delta h}$ may be desired to be a constant but such a requirement does not alter the basic philosophy of the approach, it simply means that a different function is used in (2).

Some design methods require the specification of a complete criterion matrix. This can be built up as follows

(i) Select a station to act as origin; this will usually be arbitrary and the origin will be assigned zero variance but for some problems we may begin with a station of known height with variance σ^2.

(ii) Compute the required variance of every station from

$$\sigma^2_i = \sigma^2 + f(d_i) \qquad (3)$$

where σ^2_i is the variance of station i and d_i the distance of the ith station from the origin $f(d_i)$ is as defined by (2) and σ^2 will be zero for an arbitrary origin. Note that there is no covariance between the origin and any other station.

(iii) The covariance of every pair of stations, σ_{ij}, is then computed from the well known relationship

$$\sigma_{ij} = \frac{1}{2}(\sigma_i^2 + \sigma_j^2 - \sigma_h^2) \qquad (4)$$

where $\sigma_{\Delta h}^2$ is given by (2).

The use of such a process to built up a covariance matrix suffers from two disadvantages. Firstly many levelling networks are for specific purposes - such as connecting tide gauges to study the mean sea level surface and consequently the design requirements apply only to a limited number of stations. Hence we cannot always completely define a covariance matrix in this manner as we really may not care at all which variances and covariances result at and between stations whose purpose is only to act as intermediate connections between the stations of special interest. Clearly if a complete criterion matrix is needed some (arbitrary) function in (2) must be decided upon.

A second disadvantage of this approach is that the function specified in (2), although it could genuinely reflect a particular user's precision requirements, may lead to a covariance matrix that is extremely atypical and hence cause problems in applying certain design strategies. Very little work has been done on covariance functions for levelling although for two-dimensional networks a lot of research has taken place, both for theoretical covariance functions, e.g. Grafarend and Schaffrin (1979) and Schmitt (1977), and for idealised network shapes, e.g. Wimmer (1981).

If now seems clear that all precision requirements, whether in the form of complete covariance (criterion) matrices, or otherwise, must (a) satisfy the desires of the network users and (b) be of a form that can lead to the design of networks with covariance matrices having similar structures to those established by practical networks.

3. Review of second order design procedures

A number of methods have been used for the second order design of levelling networks. What follows is a brief description of them followed by some critical remarks as to their suitability.

3.1 Method 1

Equation (1) is inverted to give

$$A^T W A = C_{\hat{x}}^{-1} = Q_{\hat{x}} \tag{5}$$

and, assuming W to be diagonal i.e. uncorrelated observations (5), rearranged, using the Khatri-Rao product, to yield

$$\left(A^T \odot A^T\right) w = \text{vech } Q_{\hat{x}} \tag{6}$$

where w is a vector containing the diagonal elements of W. Notice that, in a design problem where m new stations are connected by n observed height differences, (6) will be a set of $m(m+1)/2$ equations in n unknowns. The basic set up was proposed by E. Grafarend, e.g. Grafarend (1974), and further developed by Schaffrin (1977). There are two ways to solve (6): by generalised matrix algebra or linear programming.

3.1.1 Generalised matrix algebra

Bossler et al. (1973) suggested that (6) may be solved by using the Moore-Penrose inverse

$$w = \left(A^T \odot A^T\right)^+ \text{vech } Q_{\hat{x}} \tag{7}$$

Thus will produce a set of observation weights which, if later achieved in practice, will yield a network whose covariance matrix best fits the criterion matrix in a least squares sense, also it will be optimum in the sense that $w^T w$ will be a minimum. This will tend to minimise the total length of levelling because if w is made small more observations will be discarded, but, of course, it is not possible to claim that the solution can yield an optimum network. Note that if required coefficients can be introduced, i.e. $w^T P w$ can be minimised, by use of the minimum v-norm generalised inverse.

A numerical problem with this solution arises because, even for quits small networks, $\left(A^T \odot A^T\right)$ is a large matrix, with size $m(m+1)/2$ by n, and the computation of its inverse is time consuming. The fact that it is spare is not very helpful in practice and, of course, $\left(A^T \odot A^T\right)^+$ will always be a full matrix. Furthermore the inversion involves poorly conditioned matrices leading to numerical difficulties. These can, to

some extent, be lessened by rewriting (6) in its canonical form (see Schmitt (1977) and (1978))

$$w = (Z^T \odot Z^T)\ \mathrm{vech}\, D \tag{8}$$

where $Q_{\hat{x}}$ has been decomposed by the similarly transformation

$$Q_{\hat{x}} = EDE^T \tag{9}$$

and

$$Z = AE \tag{10}$$

Of course equations (7) and (8) give identical solutions.

Cross and Whiting (1981) have carried out a number of tests applying this technique to levelling networks with the disappointing result that it regularly produced negative observation weights. These clearly have no physical meaning and are therefore difficult to interpret. One approach is simply to discard observations with negative weights but this leads to disconnected networks, i.e. networks split into several independent sections. Alternatively only the observation with the least negative weight can be discarded and the process repeated but tests have shown that the observation with the least negative weight is rarely the least valuable.

The foregoing tests were carried out using a criterion matrix computed as in section 2 of the paper and using as a covariance function

$$\sigma^2_{\Delta h} = \text{constant x } d_{ij} \tag{11}$$

It was thought that perhaps this was unrealistic as Schmitt (1977) has reported that, for the design of two-dimensional networks, using criterion matrices with the Taylor Karman structure, negative weights did not arise. To test this a "real" network was analysed and the design procedure carried out using its a posteriori covariance matrix as the criterion matrix. Negative weights were again found and it must be concluded that the Moore-Penrose inverse is fundamentally unsuitable for solving a design problem set up as in (6).

3.1.2 Linear Programming

One way of avoiding negative weights is to use linear programming. Following the suggestion of Boedecker (1977), who solved (6) for the

case of gravity networks by linear programming, Cross and Thapa (1979) attempted to find a solution for w in (6) such that the resulting network would have a covariance matrix that would, in some sense, be better than the criterion matrix. They reasoned that, since the variance of a computed height difference is given by

$$\sigma^2_{\Delta h} = \sigma^2_i + \sigma^2_j - 2\sigma_{ij} \tag{12}$$

a network is bound to satisfy its design criteria if the variances in the covariance matrix are forced to be smaller than those in the criterion matrix and, conversely, the covariances larger. Since (6) involves an inversion of the criterion matrix, and since inversion is the matrix equivalent of a reciprocal, these inequalities were reversed and the linear programming constraint equations written as:

$$\left(A^T \odot A^T\right) w \geq \operatorname{vech} Q_{\hat{x}} \quad \text{(diagonal elements)} \tag{13}$$

$$\left(A^T \odot A^T\right) w \leq \operatorname{vech} Q_{\hat{x}} \quad \text{(off-diagonal elements)} \tag{14}$$

$$w_i \geq 0 \qquad \text{for all } w \tag{15}$$

The objective function was given coefficients of unity, i.e. the sum of the weights was minimised which, as explained in 3.1.1 tends to reduce the total length of levelling required to be observed. Unfortunately the method sometimes yield networks which do not satisfy the design criteria. The reason is that the simple reversal of inequality signs due to the inversion of the criterion matrix is not valid. It seems impossible to predict, for a given choice of method, the correct inequality signs for (13) and (14), and it must be concluded that linear programming can only be applied if the inversion of the criterion matrix is avoided.

3.2 Method 2

Cross and Whiting (1980) have suggested that the inversion of the criterion matrix can be avoided by expanding the left hand side of (1) using an unspecified generalised inverse

$$A^- W^- \left(A^T\right)^- = C_{\hat{x}} \tag{16}$$

which, after application of the Khatri-Rao product, becomes

$$\left(A^{-} \odot A^{-}\right) v = \operatorname{vech} C_{\hat{x}} \tag{17}$$

where v is a vector containing the required variances of the observations (the diagonal elements of W^{-}). Equation (17) can now be restated, using the correct part of the logic in 3.2.2, as a linear programming problem with the following constraint equations

$$\left(A^{-} \odot A^{-}\right) v \geq \operatorname{vech} C_{\hat{x}} \quad \text{(diagonal elements)} \tag{18}$$

$$\left(A^{-} \odot A^{-}\right) v \leq \operatorname{vech} C_{\hat{x}} \quad \text{(off-diagonal elements)} \tag{19}$$

$$v_i \geq 0 \qquad \text{for all } i \tag{20}$$

As in 3.1.2 unity objective coefficients can be used but for this set up the objective function must be maximised in order to reduce the total length of levelling (observations with large required variances will be discarded).

A special advantage of the method is that the complete criterion matrix is not required as equations (18) and (19) need not be formed for all elements of C_x. Moreover the need for a criterion matrix can be completely eliminated by making suitable linear combinations of the equations (similar to section 4.2 of this paper). Unfortunately the method proved impractical as a suitable generalised inverse could not be found. The Moore-Penrose inverse was tried by Cross and Whiting (1980) (even though they showed that theoretically it was not valid) but it resulted, in general, in designed networks being much more precise than required and hence too expensive.

3.3 Method 3

The negative weight problem may also be overcome by use of the linear complimentarily algorithm as suggested by Schaffrin (1980). This involves determining a best-fit solution (in the least squares sense) to (6) subject to a number of linear constraints which, as well as describing the required precision and cost of the network, also ensure that w is non-negative. The mathematical set up is essentially equivalent to a quadratic programming problem and can be written as

minimise $$\left(\left(A^T \odot A^T\right) w - \operatorname{vech} Q_{\hat{x}}\right)^T \left(\left(A^T \odot A^T\right) w - \operatorname{vech} Q_{\hat{x}}\right) \quad (21)$$

subject to $$\left(A^T \odot A^T\right) w \ (\geq, =, \leq) \ \operatorname{vech} Q_{\hat{x}} \quad (22)$$

$$p^T w \leq c \quad (23)$$

$$w_i \geq 0 \quad \text{for all } i \quad (24)$$

where p is a vector of coefficients (which are difficult to determine) relating observation weight to cost and c the total allowable cost. Liew and Shim (1978) contains details of a computer program suitable for the solution of this problem. Note that the difficulty regarding the inequality signs for use in (22) arises again but Schaffrin (1980) states that it may be avoided by reforming (21) and (22) using the canonical formulation (se 3.1.1) and restricting (22) to the rows which correspond to the eigenvalues of Q_x within $\operatorname{vech} D$, then (22) becomes

$$\left(Z^T \odot Z^T\right) w \geq \operatorname{vech} D \quad (25)$$

Schaffrin et al. (1980) have successfully applied the method to two-dimensional networks with Taylor-Karman criterion matrices but, according to Whiting (1980), it does not work with levelling networks using criteria matrices as described in 3.1.1. The result is a disconnected network, i.e. too many of the elements of w becomes zero indicating that they need not be measured. Also it should be noted that in no part of the process is w being minimised so there is no way that the method is reducing the total amount of levelling (although perhaps this could be done by judicious choice of p and c in (23)).

4. A new method

Following the ideas of Sprinsky (1978) and Wimmer (1981) we reform the basic mathematical statement of the second order design problem as follows. First we write

$$W = W W^{-1} W \quad (26)$$

Then postmultiplying both sides of (26) by $A\left(A^T W A\right)^{-1}$ and premultiplying by $\left(A^T W A\right)^{-1} A^T$ yields

$$\left(A^T W A\right)^{-1} A^T W A \left(A^T W A\right)^{-1} = \left(A^T W A\right)^{-1} A^T W W^{-1} W A \left(A^T W A\right)^{-1} \quad (27)$$

which simplifies to

$$\left(A^T W A\right)^{-1} = \left\{\left(A^T W A\right)^{-1} A^T W\right\} W^{-1} \left\{W A \left(A^T W A\right)^{-1}\right\} \tag{28}$$

Then if we put

$$B = W A \left(A^T W A\right)^{-1} \tag{29}$$

and substitute (1) and (29) into (28) we get

$$C_{\hat{x}} = B^T W^{-1} B \tag{30}$$

Applying the Khatri-Rao product to (30) and rearranging them yields

$$\left(B^T \odot B^T\right) v = \text{vech}\, C_{\hat{x}} \tag{31}$$

Then putting $H = B^T \odot B^T$ (32)

and substituting in (31) we get

$$H v = \text{vech}\, C_{\hat{x}} \tag{33}$$

where v contains the diagonal elements of W^{-1}, i.e. v is a vector of the required observation variances. This formulation is of a similar structure to (6) but has the considerable advantage of being in terms of the criterion matrix itself rather than its inverse. All solutions to (33) must, of course, be iterative because, according to (29), the matrix B is itself in terms of W. Hence we must first guess a set of values for W, solve equation (31) for v (and hence W^{-1}) and use this value to recompute B. The process is repeated until W ceases to change. Two techniques are suggested as being suitable for solving (31) (or (33)). They are, in principle, very similar to those described in 3.1 for solving (6): ordinary least squares and linear programming.

Both methods will now be described along with tests of their application to the simple first order design problem described below (note the problem is stated in terms equivalent to those used in section 1 of this paper).

GIVEN

(a) Four stations (station 1 assumed fixed with zero variance) located as in Fig. 1.

(b) Six possible level routes with lengths as indicated in Fig. 1.

(c) A precision requirement that all variances of computed height differences between station d metres apart should be less than $0.001\,d$ mm.

(d) No reliability criteria were considered for these tests.

FIND

The shortest subset of level routes whose observation will yield the required precision.

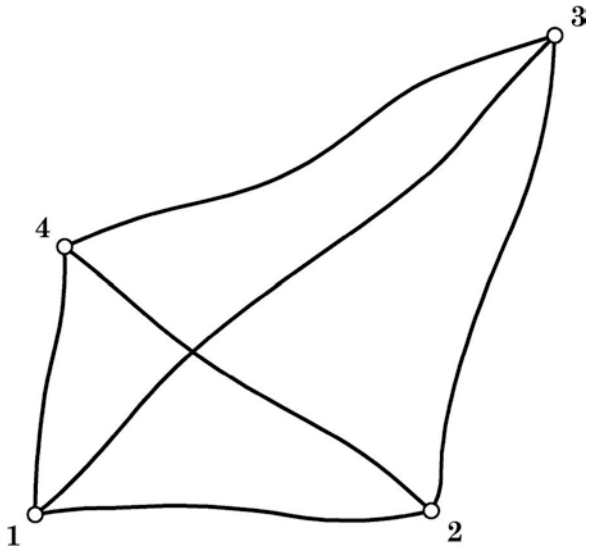


Fig. 1

Furthermore it was assumed that the field levelling process could not be carried out with a standard error of better than $0.001\,d$ mm, where d is the length of the level route in metres.

4.1 Ordinary least squares

The application of ordinary least squares to (33) gives

$$v = \left(H^{T}H\right) H^{T} \operatorname{vech} C_{\hat{x}} \tag{34}$$

The network design then proceeds as follows.

(i) Specify $C_{\hat{x}}$. For the numerical examples this was done by the procedure outlined in section 2 of this paper using

$$f(d) = 0.001\,d$$

(ii) Make an initial estimate of W, in the examples in this paper variances were put proportional to the length of the level routes, viz

$$w_i = 1/(0.001\, d_i)$$

(iii) Compute B and H from (29) and (32) respectively, assuming all possible level lines are to be observed.

(iv) Derive the necessary observation variances from (34) and hence compute W. If W is significantly different from its latest estimate go to step (iii).

(v) Discard all observations with large variances (note that in practice this decision could be affected by reliability considerations).

(vi) Check that the designed network fulfills the original design criteria.

This process has been applied to the foregoing example with the results after the 10th iteration of stages (iii) and (iv) as given in Table 1.

LINE	LENGTH (m)	BEST POSSIBLE STANDARD ERROR (mm)	DESIGN STANDARD ERROR (mm)	DESIGN WEIGHT (mm^{-2})	
1-2	1000	1.000	1.106	0.818	
1-3	1813	1.346	∞	0.000	*
1-4	722	0.850	0.901	1.232	
2-3	1252	1.119	1.260	0.630	
2-4	1184	1.088	4.049	0.061	*
3-4	1385	1.177	1.334	0.562	

* Discarded

TABLE (1)

Note that the third column, headed "best possible standard error", contains the highest achievable precision for each line assuming that the levelling process could not be carried out with a precision better than $0.001d$ mm. It should be noted that all designed standard errors, column four of Table 1, are greater than this, hence the observations are all feasible. Clearly proposed observations along the lines 1-3 and 2-4 can be discarded as their required standard errors are so large that they cannot significantly contribute to the precision of the network. After the removal of these two lines the network, with a total length of levelling of 4,359 m, in Fig. 2 results. Its covariance

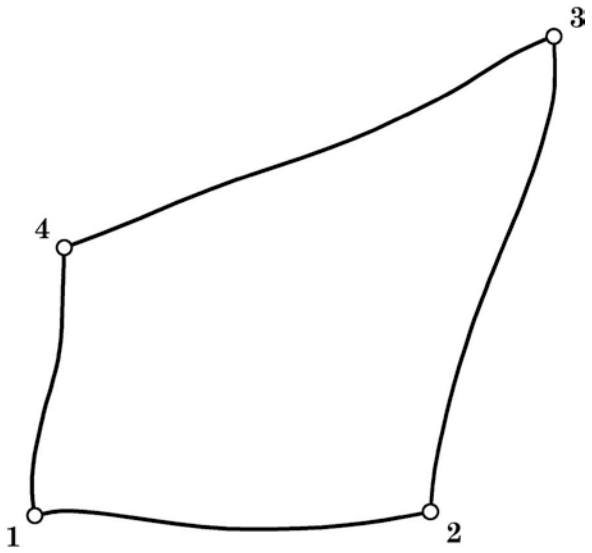


Fig. 2

matrix computed using the design standard errors is

$$\begin{bmatrix} 0.9458 & 0.5865 & 0.1837 \\ 0.5865 & 1.3480 & 0.4223 \\ 0.1837 & 0.4223 & 0.6897 \end{bmatrix}$$

which can be compared with the criterion matrix of

$$\begin{bmatrix} 1.0000 & 0.7804 & 0.2692 \\ 0.7804 & 1.8137 & 0.5752 \\ 0.2692 & 0.5752 & 0.7225 \end{bmatrix}$$

Note that when the relative standard errors are computed from the covariance matrix they are all less than or equal to the criteria of $0.001d$ mm. Also note that the covariance matrices are of size 3 x 3 because station 1 is fixed.

Encouraged by these results the same design process was applied to a twelve station network whose configuration and possible observations are given in Fig. 3. The results of the design are in Table 2 and, after discarding all the observations with zero weight, the configuration is as in Fig. 4. A major problem is now immediately apparent: four of the lines (marked with a +) have to be measured with a precision greater than that which is possible with the available instrumentation. Also

LINE	LENGTH (m)	BEST POSSIBLE STANDARD ERROR (mm)	DESIGN STANDARD ERROR (mm)	DESIGN WEIGHT (mm^{-2})	
1-2	1431	1.196	1.790	0.312	
1-3	1202	1.096	∞	0.000	*
1-4	1106	1.052	∞	0.000	*
2-3	865	0.930	0.968	1.066	
2-5	1610	1.269	∞	0.000	*
3-4	665	0.815	0.427^{+}	5.492	
3-5	1316	1.147	∞	0.000	*
3-6	1504	1.226	∞	0.000	*
4-6	1247	1.117	1.023^{+}	0.956	
4-7	1578	1.256	∞	0.000	*
5-6	1434	1.197	0.624^{+}	2.564	
5-8	1460	1.208	∞	0.000	*
5-9	1077	1.038	∞	0.000	*
6-7	605	0.778	0.889	1.266	
6-8	456	0.675	0.913	1.199	
6-10	1044	1.022	∞	0.000	*
7-10	690	0.831	1.344	0.554	
8-9	824	0.908	0.670^{+}	2.224	
8-10	794	0.891	1.671	0.358	
8-11	800	0.894	∞	0.000	*
9-11	655	0.809	0.986	1.028	
10-11	1001	1.000	∞	0.000	*
10-12	850	0.922	1.256	0.634	
11-12	632	0.795	1.006	0.989	

* Discarded

\+ Standard errors smaller than best possible

TABLE (2)

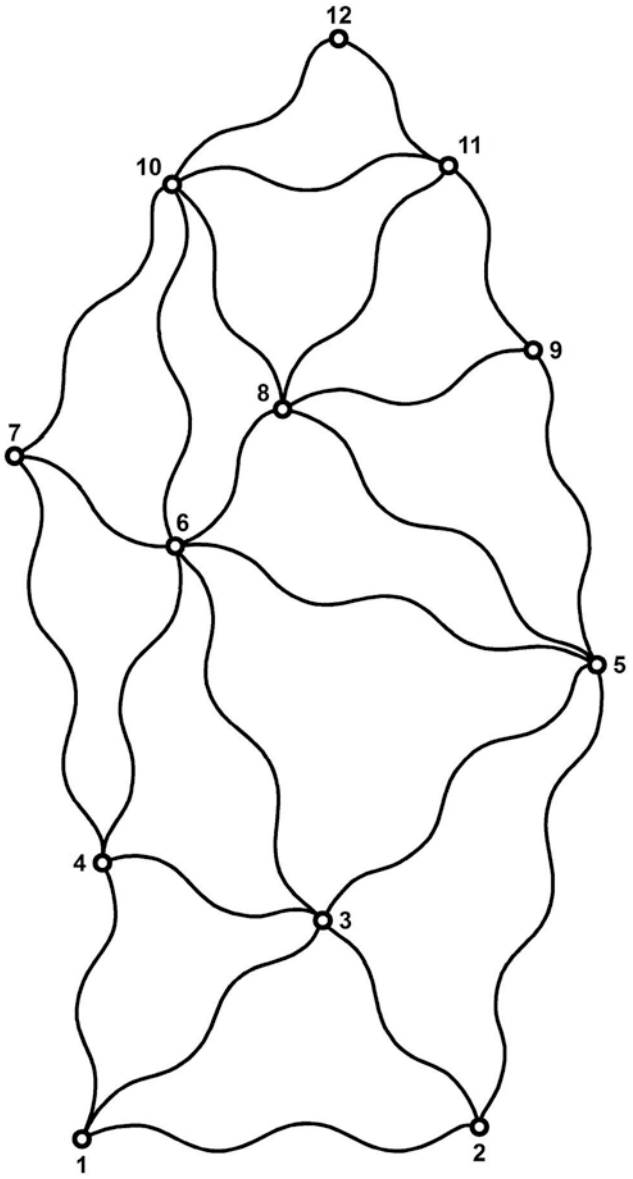


FIG 3. THE POSTULATED NETWORK.

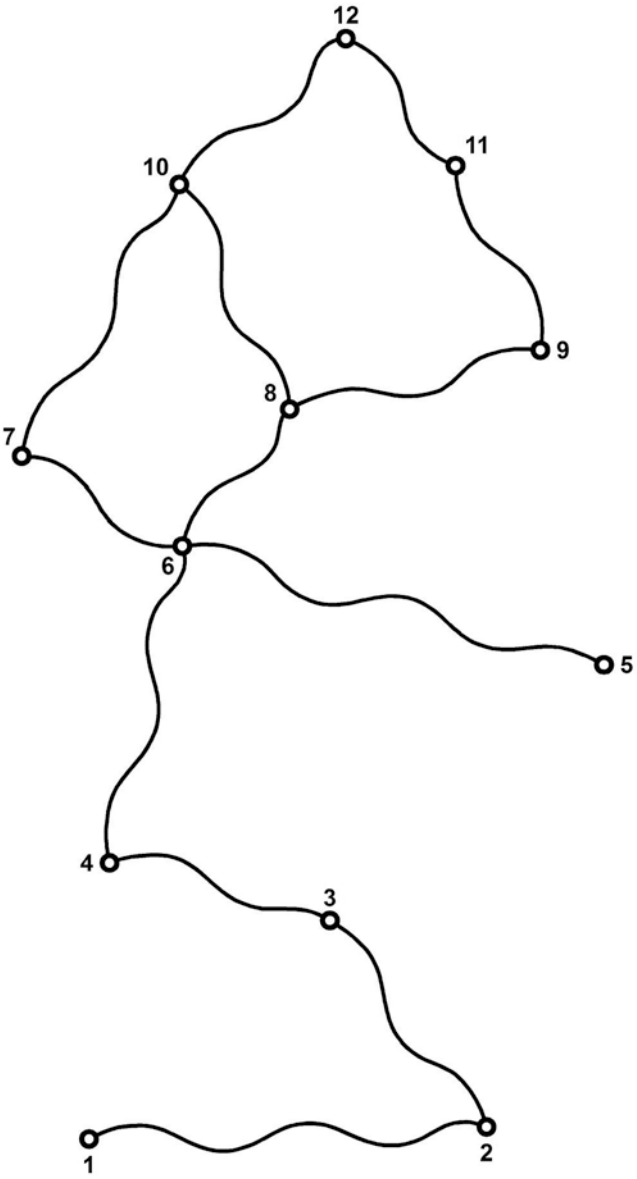


FIG 4. THE DESIGNED NETWORK.

examination of Fig. 4 shows that an unreliable network has resulted. Clearly the procedure, as it stands, is not then practically useful but it is anticipated that the two difficulties could be overcome by

(i) including additional equations of the form

$$v_i = 0.001\, d_i \qquad \text{(for this example)}$$

with a "high weight", for all variances which, after iteration, became too small and

(ii) computing the reliability before discarding an observation.

Two important positive conclusions can be drawn from the tests:

(i) The method converges, producing a set of weights which satisfy, in a least squares sense, the design criteria.

(ii) Negative weights do not arise in the final solution (although it has been noticed that they appear during the iteration process).

4.2 Linear Programming

It would be possible to solve (33) using linear programming following a similar procedure to that described in 3.1.2 but with a change of inequality because we are now dealing with $C_{\hat{x}}$ rather than $Q_{\hat{x}}$. The constraint equations are then

$$H\, v \leq \text{vech}\, C_{\hat{x}} \qquad \text{(diagonal elements)} \tag{35}$$

$$H\, v \geq \text{vech}\, C_{\hat{x}} \qquad \text{(off-diagonal elements)} \tag{36}$$

$$v_i \geq 0 \qquad \text{for all } v \tag{37}$$

and the objective function $\sum_{i=1}^{n} v_i$ is maximised.

This set up does not suffer from the difficulties explained in 3.1.2 because the criterion matrix does not need to be inverted. However, because the basic design precision criteria can be expressed as linear transformation of the a posteriori covariance matrix it is actually possible to avoid completely the need for the criterion matrix. We proceed as follows.

Let there be k pairs of stations, i and j, between which the variance is required to be less than some specified amount, e.g. less than

some function of the distance d_{ij}. Then for each pair we write

$$\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij} \leq f(d_{ij}) \tag{38}$$

which can be expressed in the form

$$p_{ij}\, \mathrm{vech}\, C_{\hat{x}} \leq f(d_{ij}) \tag{39}$$

where p_{ij} is a null vector except for the elements which correspond to σ_i^2, σ_j^2 and σ_{ij}^2 in $\mathrm{vech}\, C_{\hat{x}}$ which are +1, +1 and -2 respectively. For all k pairs of lines the precision criteria are

$$P\, \mathrm{vech}\, C_{\hat{x}} \leq F(d) \tag{40}$$

where P is a $k \times m(m+1)/2$ matrix with each row of the form of p_{ij}.

Substituting (33) in (40) gives

$$P H v \leq F(d) \tag{41}$$

which is in a form ready for immediate application of linear programming.

The network design proceeds as follows

(i) Specify the precision criteria in the form of a linear transformation of the a posteriori covariance matrix (note this does not involve explicit formation of the criterion matrix), i.e. form the matrix P and vector $F(d)$ in (40).

(ii) Make an initial estimate of W, in the example in this paper the variances were put proportional to the length of the level routes, viz

$$w_i = 1/(0.001\, d_i)$$

(iii) Compute B and H from (29) and (32) respectively, assuming all possible level lines are to be observed.

(iv) Set up the linear programming problem as follows:

Constraints:

$$\{PH\} v \leq F(d) \tag{41}$$

$$v_i \geq g(d_i) \quad \text{(for all } i\text{)} \tag{42}$$

Objective:

$$\sum_{i=1}^{n} v_i = \text{maximum} \tag{43}$$

Note that $g(d_i)$ in (42) represents the best possible variance that could be achieved in practice, for the example

$$g(d_i) = 0.001\, d_i \tag{44}$$

(v) Solve the linear programming problem and compute W from v, using

$$w_i = 1/v_i$$

If W is significantly different from its latest estimate go back to stage (iii).

(vi) Discard all observations with small weights (large variances), note that, in practice, reliability would now be considered.

(vii) Check the designed matrix meets the precision requirements.

Application of the above process to the example in Fig. 1 yields, after three iterations, the results in Table 3. The same two proposed

LINE	LENGTH (m)	BEST POSSIBLE STANDARD ERROR (mm)	DESIGN STANDARD ERROR (mm)	DESIGN WEIGHT (mm^{-2})	
1-2	1000	1.000	1.000	1.000	
1-3	1814	1.346	∞	0.000	*
1-4	722	0.850	0.850	1.384	
2-3	1253	1.119	1.119	0.798	
2-4	1184	1.088	∞	0.000	*
3-4	1386	1.177	1.177	0.722	

* Discarded

TABLE (3)

observations are discarded as for the application of the least squares design process in 4.1 and the designed network is again as in Fig. 2. The only difference is in the numerical values of the designed weights but, so long as they are feasible, they are not important because in practice the same levelling procedure is likely to be used for all lines.

Of course the method suffers from the same problems as in 4.1 regarding network reliability but these could be solved at stage (vi). The method does, however, offer the following important advantages.

(i) It converges, yielding a set of weights which lead to a network which is bound to have the required precision (unlike the method of 4.1 where we only "best fit" the precision requirements).

(ii) It is bound to yield non-negative, feasible weights.

(iii) It does not need the explicit formation of a criterion matrix, moreover precision requirements can be specified at or between only a limited number of stations.

Tests on larger, more realistic, networks are in hand.

5. Conclusions

Apart from computer simulation all the currently known methods of designing level networks suffer from difficulties which make their application to real problems impractical.

A new iterative approach, involving either a least squares or linear programming solution, can overcome these difficulties and, if combined with a consideration of reliability, promises a complete solution to the second and first order level network design problems.

REFERENCES

BOEDECKER, G., 1977: The development of an observation scheme from the variance covariance matrix of the unknowns. Proc. Int. Symposium on Optimisation of Design and Computation of Control Networks, Sopron.

BOSSLER, J., E. GRAFAREND and R. KELM, 1973: Optimal design of geodetic Nets, 2. Journal of Geophysical Research, Vol. 78, No. 26.

CROSS, P.A. and K. THAPA, 1979: The optimal design of levelling networks. Survey Review, Vol. XXV, No. 192, pp. 68-79.

CROSS, P.A. and B.M. WHITING, 1980: The design of national vertical control networks. Proc. 2nd Int. Symposium on Problems Related to the Redefinition of North American Vertical Geodetic Networks, Ottawa, Canada.

CROSS, P.A. and B.M. WHITING, 1981: On the design of vertical geodetic control networks by iterative methods. VI. International Symposium on Geodetic Computations, Munich.

GRAFAREND, E.W., 1974: Optimisation of geodetic networks. Bolletino di Geodesia a Science Affini, Vol. 33, No. 4, pp. 351-406.

GRAFAREND, E.W. and B. SCHAFFRIN, 1979: Kriterion-matrizen I-zweidimensionale homogene und isotrope geodätische Netze. Zeitschrift für Vermessungswesen, No. 4, pp. 133-149.

LIEW, C.K. and J.K. SHIM, 1978: Computer program for inequality constrained least squares estimation. Econometrica, Vol. 46, p. 237.

W. NIEMEIER and G. ROHDE, 1981: On the optimization of levelling nets with respect to the determination of crustal movements. IAG Symposium on Geodetic Networks and Computations, Munich.

RAO, C.R. and S.K. MITRA, 1971: Generalized inverses of matrices and its applications. John Wiley & Sons.

SCHAFFRIN, B., F. KRUMM, and D. FRITSCH, 1980: Positive-diagonale Genauigkeitsoptimierung von Realnetzen über den Komplementaritäts-algorithmus. VIII. Internationaler Kurs für Ingenieurvermessung, Zürich.

SCHAFFRIN, B., 1980): Some proposals concerning the diagonal second order design of geodetic networks. Manuscripta Geodaetica, Vol. 6, No. 3, pp. 303-326.

SCHMITT, G., 1977: Experiences with the second order design of geodetic networks. Proc. International Symposium of Design and Computation of Control Networks, Sopron, Hungary.

SCHMITT, G., 1978: Numerical problems concerning the second order design of geodetic networks. Second International Symposium Related to the Redefinition of North American Geodetic Networks, Washington.

SPRINSKY, W.H., 1978: Improvement of parameter accuracy by choice and quality of observation. Bulletin Géodésique, Vol. 52, pp. 269-279.

WHITING, B.M., 1980: Unpublished paper.

WIMMER, H., 1981: Second order design of geodetic networks by an iterative approximation of a given criterion matrix. VI. International Symposium on Geodetic Computations, Munich.

STRENGTH ANALYSIS OF HORIZONTAL NETWORKS USING STRAIN

Peter DARE and Petr VANÍČEK*
Toronto, Canada

ABSTRACT

A method for the strength analysis of horizontal geodetic networks is developed. While the existing approaches are based on studying the behaviour of random errors, our study is based on the ability of the network to deform. How easy it is for the network to experience a scale change, twist or shear is determined. "Forces" to deform the network come from changing the values of observations by an amount equal to their standard deviation.

Displacements due to a change in an observed value are determined as an intermediate step; this is done through a matrix transforming changes in observed values to displacements. The displacements are then transformed to strain parameters of which the largest at each station are plotted to show the strength of the network. Both a simulated and a real network are used to test the new approach.

1. INTRODUCTION

Most of the existing techniques for strength analysis are based on the study of the behaviour of random errors throughout the network. One of the simplest methods uses absolute error ellipses. The main use of these ellipses is to study the "strength" of the network in orientation and scale. By looking at the shape of the absolute error ellipses, a general weakness of the network in either orientation or scale can be judged. However, small variations in "strength" are obscured by the growth of the ellipses in size with distance from the origin of the network. Examples of this approach are given in Allman and Hoar (1973), and Joó (1978).

Relative error ellipses computed between any two points of the network, alleviate the problem of the consistent growth pattern present in absolute error ellipses.

---

* Currently visiting professor at the University of Stuttgart, West Germany

Although these ellipses show the local "strength" of the network (e.g. Joó, 1978), it is not particularly easy to see what type of weakness exists. Error ellipses constructed from the relative standard deviation of distances and standard deviation of azimuths have been studied (e.g. Gaździcki, 1975).

To study weaknesses in scale and orientation separately, an approach has been devised by Ashkenazi and Cross (1972). They compute the standard errors of a selected number of adjusted distances and azimuths. However, the picture one gets of the "strength" of the network depends to some extent on the selected distribution of the distances and azimuths.

Other techniques have been developed by Meissl (1974), and Bartelme and Meissl (1974) that work for infinitely large networks made up of equally sized cells. These are useful for studying how random errors behave in hypothetical networks. Their application to actual networks, which are unlikely to conform to the idealised designs is, however, limited.

Considering the network from the civil engineering structural analysis point of view can give more insight into the strength of the network. An example of this type of analysis is in Borre (1972). A problem with this type of analysis is that the network has to be constructed of triangles; some alteration in the design of the network may thus have to take place. Also, observations extending over the length of the network (e.g. VLBI or satellite determined distances) cannot be accommodated. Factor analysis (Harman, 1960) is one topic that may warrant further investigation.

Baarda (1967) introduced the idea of invariants into the analysis of geodetic networks. These invariant quantities are freed from the effect of systematic errors in scale and azimuth. Grafarend (1972) develops geodetic observation equations in terms of dilatation, shear and rotation. Lamé constants are computed for certain network designs.

With so many existing methods claiming to analyse strength, justification for yet another approach is needed. A dictionary definition of strength is "The ability to withstand stress or deformation" (Webster's Third New International Dictionary, 1976). It is this definition that our approach is based upon. To see how strong a network is, a stress must be applied to it. If it can withstand this stress, the network is strong (with respect to that stress) and vice-versa.

The crucial question is what kind of stress should be applied. Clearly, there must be no preference for particular stresses there being no reason to presume that the network would be actually subjected to any such forces. The hypothetical stress must therefore be applied evenhandedly throughout the network.

A reasonable approach appears to assume the largest possible stress to act on each element (i.e. distance, angle, azimuth) of the network. This "worst stress" will if otherwise unobstructed deform each element inversely proportional to the element's strength which we will assume to be given by the weight of the element. In other words, we shall assume that the "worst stress" can deform each element by as much as a constant multiple of its standard deviation. To simulate the external stresses evenhandedly, we therefore can alter each observation by, typically, one standard deviation and study the resulting deformation.

The next question is then which observations in the network should be so varied. Clearly, changing all the observations simultaneously makes little sense. To see why, imagine a trilateration network of equal sides with distances observed to the same accuracy. Changing all these distances simultaneously would result in a homogeneous and isotropic deformation which would indicate the same strength in the inner as well as outer reaches of the network.

Again, a reasonable approach appears to look for the largest possible deformation in response to our simulated stress. The neighbourhood of a point of a network will thus be stronger the less it will succumb to the most effective of all the possible stresses. We thus investigate at each point the maximum possible deformation (in scale, twist and shear) in response to the change of whichever element is capable of producing this maximum deformation.

2. Description of the Method

The observation equations for a least-squares positions adjustment are given by

$$\underline{A}\,\underline{x} = \underline{\ell} \qquad (1)$$

where $\underline{A}$ is the design matrix, $\underline{x}$ is a vector of coordinate corrections, and $\underline{\ell}$ is a vector of observational misclosures. The least squares solution

$$\underline{x} = \left(\underline{A}^T \underline{C}_\ell^{-1} \underline{A}\right)^{-1} \underline{A}^T \underline{C}_\ell^{-1} \underline{\ell} = \underline{T}\,\underline{\ell} \qquad (2)$$

can be regarded as showing the linearized effect of observations (or their misclosures $\underline{\ell}$) on positions. Clearly, if an observation misclosure ℓ_i is changed by $\Delta\ell_i$, the effect $\Delta\underline{x}$ of this change, i.e. the "displacement response", can be easily obtained from eqn. (2). Denoting $(\ell_1, \ell_2, \ldots, \ell_i + \Delta\ell_i, \ldots, \ell_n)$ by $\underline{\ell} + \Delta\underline{\ell}$, where $\Delta\underline{\ell}$ is a vector with only one non-zero element, we get

$$\Delta\underline{x} = \underline{T}\ \Delta\underline{\ell} \ . \qquad (3)$$

The "displacement response" $\Delta\underline{x}$ can be readily converted into a "strain response". For each point in the network the strain matrix (or partial derivatives of the displacement)

$$\underline{e} = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix} \qquad (4)$$

is evaluated from the displacements of the adjacent points following the least squares procedure described in Vaníček et al. (1981). Then (for each point P_i) the four elements of $\underline{e}_i$ are written as a vector $\underline{s}_i = (e_{11}, e_{12}, e_{21}, e_{22})_i^T$. It is possible to express the hypervector $\underline{S} = (\underline{s}_1, \underline{s}_2, \ldots, \underline{s}_n)^T$, containing 4n elements (in a network of n points) as a linear combination of all the displacements $\Delta\underline{x}$:

$$\underline{S} = \underline{Q}\ \Delta\underline{x} \ . \qquad (5)$$

Substitution of eqn. (3) into (5) yields

$$\underline{S} = \underline{Q}\ \underline{T}\ \Delta\underline{\ell} = \underline{R}\ \Delta\underline{\ell} \ , \qquad (6)$$

where $\underline{R}$ is the strain response matrix.

Clearly, each four-tuple or rows of $\underline{R}$ shows the strain effect at point P_i due to a unit change in $\underline{\ell}$. Conversely, j-th column of $\underline{R}$ shows all the strain effects (at all the points) of a unit change in the corresponding observation ℓ_j. Thus the 4nxm matrix $\underline{R}$ contains all the information needed to do the strength analysis. The four-tuples of elements in each row are the strain elements (at point P_i) caused by a unit change in observation ℓ_j. Multiplication of these by the standard deviation σ_j of ℓ_j gives the strain caused by the hypothetical stress causing the most deformation in element ℓ_j (the "worst stress").

All that remains to be done is to convert these strain elements into some more readily interpretable quantities. These are: total strain ρ, twist ω, and total shear γ. These are related to the strain parameters $e_{11}, e_{12}, e_{21}, e_{22}$ through the following formulae (see e.g. Vaníček & Krakiwsky (1982))

$$\rho = \left(\alpha^2 + \beta^2\right)^{1/2} \qquad (7)$$

where α and β are the semi lengths of the major and minor axes of the strain quadric,

$$\omega = \tfrac{1}{2}\left(e_{12} - e_{21}\right) \ , \qquad (8)$$

and

$$\gamma = \left[\tfrac{1}{2}(e_{11}-e_{22})^2+\tfrac{1}{2}(e_{12}-e_{21})^2\right]^{\frac{1}{2}} \quad . \qquad (9)$$

Total strain ρ describes the amount of scale change experienced at P_i, ω shows how much rotation has occurred at P_i and γ depicts the change of the shape of the neighbourhood of P_i.

The above formulae transform the four-tuples of row elements of $\underline{R}$ into triplets $(\rho,\omega,\gamma)^T$, and $\underline{R}$ becomes $\underline{R}^*$, a table in with ρ_{max}, ω_{max}, γ_{max} can be found for each point. These quantities are then plotted and contoured to display the 3 strength-characterising maps of the analysed network.

3. Strength analysis of a simulated network

To test the method, the strength of a simulated network (see fig. 1) was studied. The network consists of 26 equilateral triangles of side length 20 kms. The basic network configuration consists of all directions (although angles are actually shown in fig. 1), two azimuths (from stations 2 to 3 and 21 to 22) and two distances (from 1 to 19 and 2 to 22). The origin chosen for the adjustments was station 1; exactly the same results would have been obtained if any other station had been held fixed. All directions were given a standard deviation of 1.0", azimuths 0,8" and distances 0.02 m + 2 ppm. The basic design was then systematically changed and a strength analysis of each configuration carried out.

The total strain plots showing strength in scale are shown in fig. 3. In the first plot, the strongest areas are found around the two distances themselves, and around station 12 where their effects combine. Plot B shows very clearly the weakness in scale produced by leaving out distance 19-22. In plot C, a strengthening in scale can be seen (coming from the added azimuth 11-14) in the formerly weak areas (see plot A). However, the strengthening is only very slight, as one would expect. The omission of 3 directions (11-9, 11-8, 11-7) weakens the network in scale but not very appreciably (see plot D). The most significant change occurs in the centre of the network.

The addition of weighted stations can greatly improve the strength of a network. This is evident from plot E (in which stations 1 and 21 were weighted) by the greater strength in scale than shown in plot A. The degree of strengthening depends, of course, upon the weight of the added stations.

Strength in twist is displayed in fig. 4. Plot A was produced with the basic network design. As expected, the least twisting has occurred around the two azimuths as these restrict the rotation of members. In plot B slightly larger twisting is present in the network, although the change is not very drastic.

The strength of the network in twist has been greatly improved in plot C with the strongest area now being around the additional azimuth. In plot D, the strength in twist is now reduced, predominantly in the centre of the network where the directions were omitted. As can be seen from plot E, strength in twist is greatly improved by adding weighted stations. Not only do the twist values become much smaller, but also the range of values is reduced showing much more homogeneous strength in twist.

Fig. 5 shows the plots for the total shear for the same network designs used in the scale and twist analyses. Plot A tells us that the strongest areas, in terms of shear, are in the centre of the network, while the weakest regions are on the extremities. This occurs because there are more geometrical ties in the centre of the network than at the edges to prevent it from changing its shape. Plot B is identical to plot A showing that the omission of a distance from the network does not alter the strength of the network in shear. The reason for an apparent slight weakening of the network is shear in response to an added azimuth (plot C) is not clear. A dramatic weakening in shear has occurred around station 11 in plot D. It can be seen that directions play a large role in the strength of a network in shear.

Plot E shows only a small improvement over plot A. Thus strength in the shape of the network cannot be improved much by adding two weighted stations.

4. Strength analysis of an actual network

A strength analysis of an actual first-order geodetic network in Nova Scotia, Canada (Network 293) was carried out. A total of 135 directions with an estimated standard deviation of 0.7", 68 tellurometer distances (ρ = 3.5 ppm) and two azimuths with a standard deviation of 1.09" were used for the analysis (see fig. 2). Both contour plots and 3D diagrams were produced for this network (figs. 6-8).

The total strain plots show an area of strength in scale near 691012. The northern region has greater strength in scale then the southern region because of the two long distances (11100-11101B and 11101B-19101) observed there.

The twist plots show no twist greater than 1.7" between stations 11100 and 691003. Thus it can be concluded that the network is quite strong in twist in this region. This is due to the azimuths observed at stations 691009 and 691002. Further to the east and south from this strong region, values of twist increase showing a weakening of the network in twist.

The total shear plots have three distinctly high areas. Two of them occur in the vicinity of stations 19112 and 6419181. This is because station 6419181 is connected to just two stations of the main framework, and station 19112, the weaker of the two, is only intersected. The third weak area in total shear is around station 691013 where the geometry, we conclude, is not favourable. The strongest station with respect to shear is, as expected, 691010 which is well inside the network, and has many observations to and from it.

5. Conclusions

In this paper a novel approach to strength analysis of horizontal networks has been proposed. The approach investigates mainly the geometry of a network and uses the standard statistical aspects only to access the relative strength of individual elements observations to resist changes. Weak elements are taken as being prone to geometrically distort the network more, subject to their lower statistical weights. Constantly the "most pessimistic attitude" has been adopted in so far that the potentially worst possible effect is taken as a measure of strength.

In our opinion it does not make sense to measure strength by means of one scalar quantity. Rather one has to distinguish strength in scale, in twist and in shape. Thus 3 different plots of strength are always produced. The results of extensive testing on a simulated as well as a real network, some of which have been shown here, seem to support our views. Generally, the strength plots of the individual networks are intuitively understandable and pleasing. We are thus convinced that the presented technique can and should be used, possibly side by side with a standard variance-covariance technique, to analyse horizontal networks. Software for the analysis can be obtained from any of the authors.

Acknowledgements

This paper is based on the results of research carried out in the Department of Survey Science of the University of Toronto. The final preparation of this paper took place in the Geodetic Institute of the University of Stuttgart. The cooperation of the Geodetic Institute is gratefully acknowledged.

The work was carried out under the auspices of the Surveys and Mapping Branch of the Canadian Federal Department of Energy, Mines and Resources (continuation of research agreement number 3-5-80).

REFERENCES

ALLMAN, J.S., and G.J. HOAR, 1973: Optimisation of Geodetic Networks. Survey Review, No. 167, Vol. 22.

ASHKENAZI, V., and P.A. CROSS, 1972: Strength Analysis of Block VI of the European Triangulation. Bulletin Géodésique, No. 103.

BAARDA, W., 1967: A Generalization of the Concept Strength of Figure. In: Statistical Concepts in Geodesy. Netherlands Geodetic Commission, Publications on Geodesy, Vol. 2, No. 4.

BARTELME, N., and P. MEISSL, 1974: Strength Analysis of Distance Networks. Report No. 15 of the Geodetic Institutes, Technical University Graz, Austria.

GAŹDZICKI, J., 1975: Strength Analysis of Control Networks. Paper presented at the First Venezuelan Conference on Geodesy, Maracaibo, Venezuela.

GRAFAREND, E., 1977: Stress-Strain Relations in Geodetic Networks. Communication from the Geodetic Institute of Uppsala University, No. 16.

HARMAN, H., 1960: Modern Factor Analysis. The University of Chicago Press, Chicago, U.S.A.

JOÓ, I., 1978: The New Horizontal Control Network of Hungary. Paper presented at the Second International Symposium on Problems Related to the Redefinition of North American Geodetic Networks, Arlington, Virginia, U.S.A.

LINKWITZ, K., 1979: Least-squares methods as a tool to calculate and optimise the design of prestressed cable nets and hanging nets. Paper presented at the International Symposium on Optimisation of Design and Computation of Control Networks, Sopron, Hungary.

MEISSL, P., 1974: The Strength of Continental Terrestrial Networks. Canadian Surveyor, Vol. 28, No. 5.

VANÍČEK, P., and E.J. KRAKIWSKY, 1982: Geodesy: The Concepts. North Holland.

VANÍČEK, P., K. THAPA, and D. SCHNEIDER, 1981: The Use of Strain to Identify Incompatible Observations and Constraints in Horizontal Geodetic Networks. Manuscripta Geodaetica, Vol. 6, pp. 257-281.

WEBSTER'S Third New International Dictionary, 1976: G. and C. Merriam Company, Massachusetts, U.S.A.

Captions for Figures

Figure 1: Simulated Network

Figure 2: Network 293 in Nova Scotia, Canada.

Legend: Ⓣ denotes a distance measured with a Tellurometer

Ⓖ denotes a distance measured with a Geodimeter

→ denotes a measured azimuth

Ⓞ denotes a Doppler point

Figure 3: Strength in scale of the simulated network (microstrains)

Plot A: Basic network design

Plot B: Leaving out a distance

Plot C: Adding an azimuth

Plot D: Omitting directions

Plot E: Adding weighted stations

Legend:

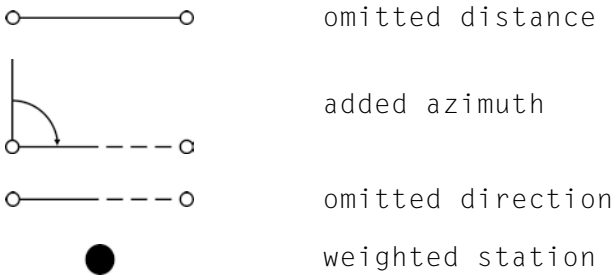


Figure 4: Strength in twist of the simulated network (seconds of arc) (for explanation see fig. 3)

Figure 5: Strength in shear of the simulated network (microstrains) (for explanation see fig. 3)

Figure 6: Strength in scale of Network 293 (microstrains)

Figure 7: Strength in twist of Network 293 (seconds)

Figure 8: Strength in shear of Network 293 (microstrains)

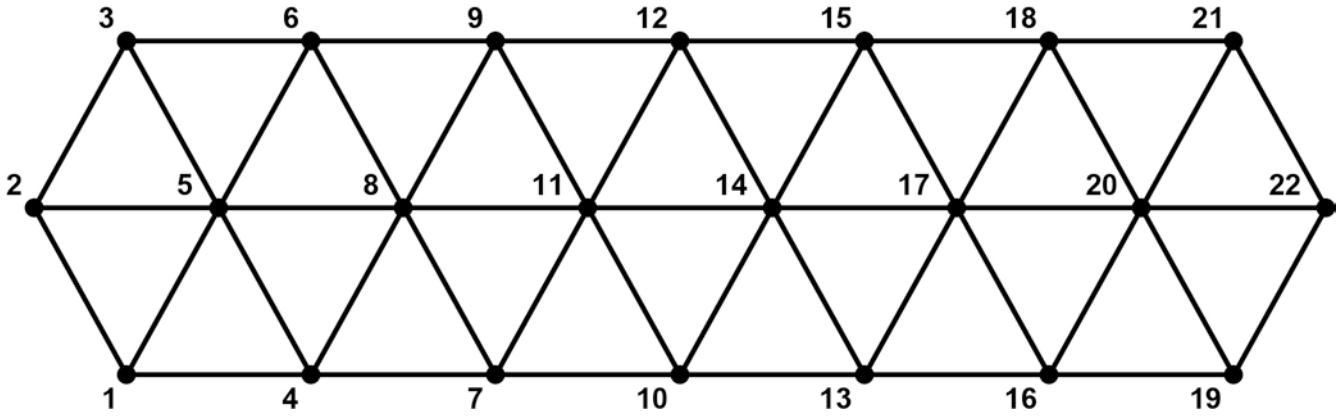


Figure 1

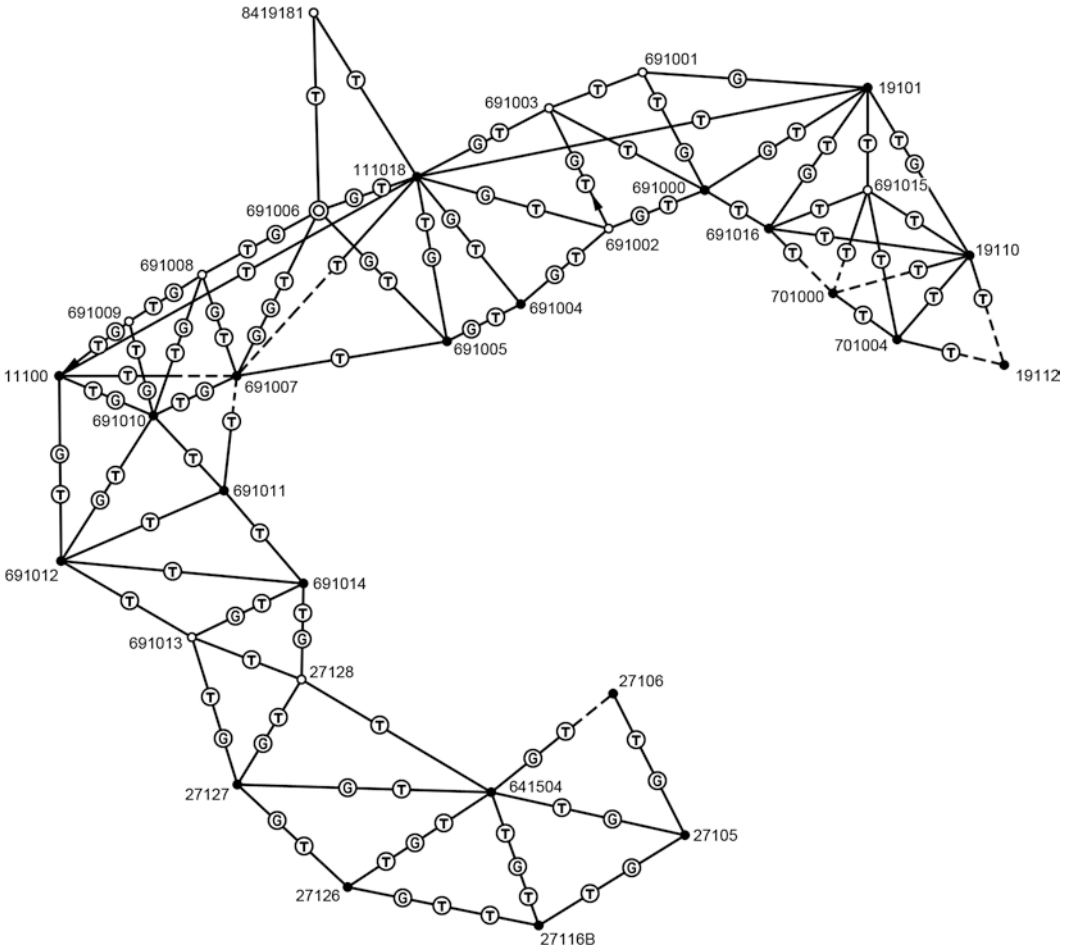


Figure 2

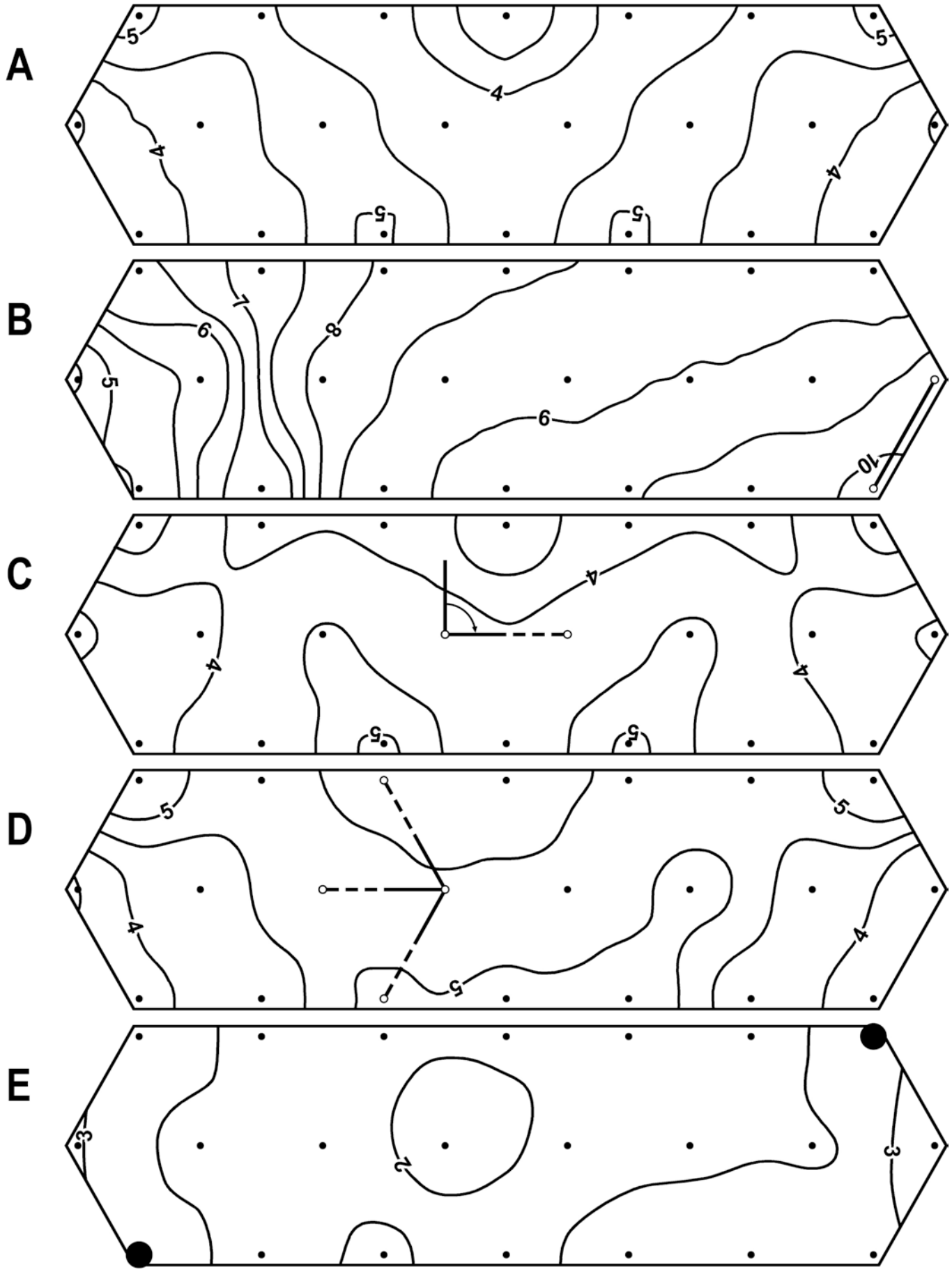


Figure 3

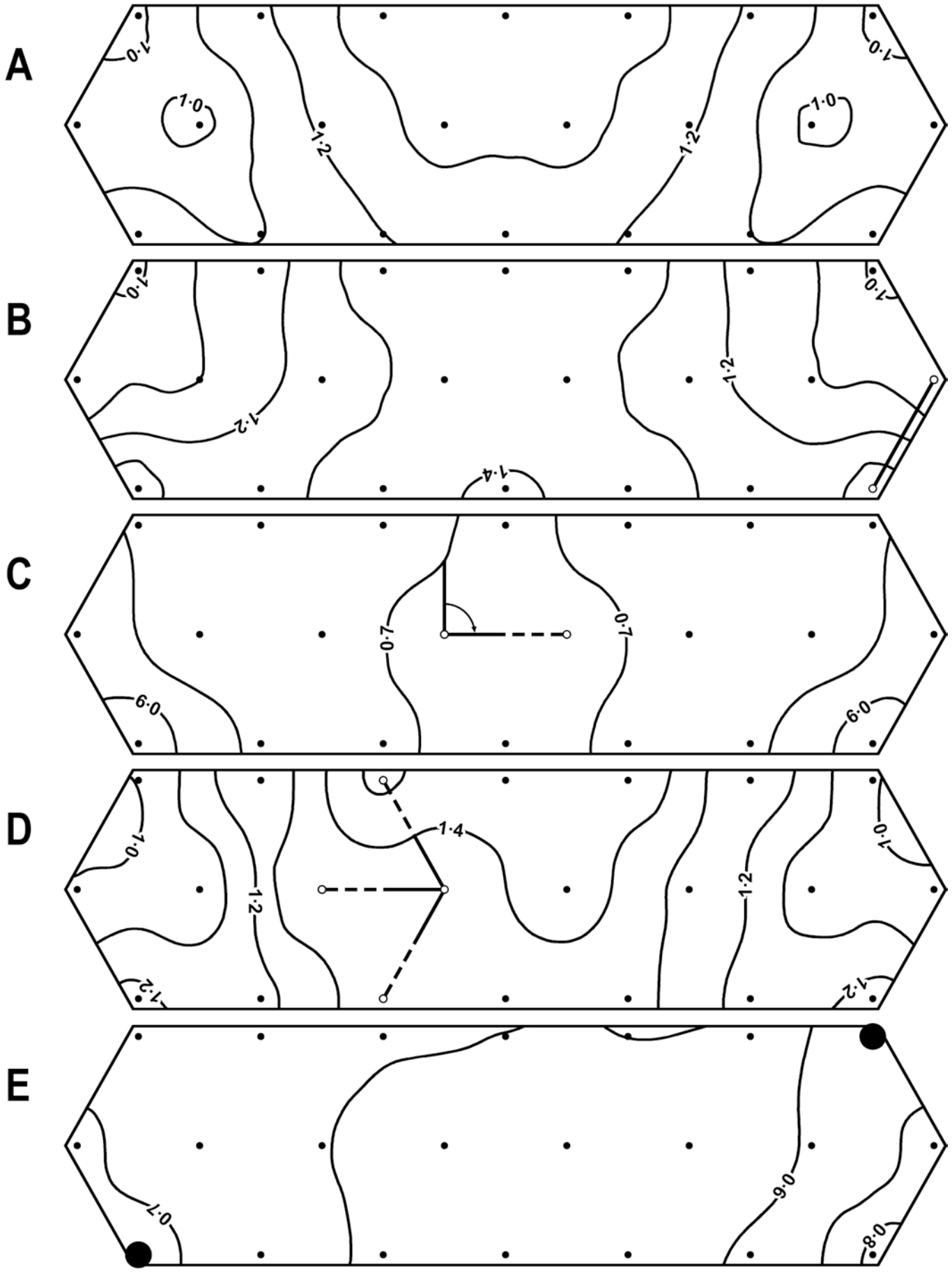

A
1·0
1·0
1·0
1·2
1·2
B
1·0
1·0
1·2
1·2
1·4
C
0·7
0·7
6·0
6·0
D
1·0
1·4
1·0
1·2
1·2
1·2
E
0·7
0·6
0·8

Figure 4

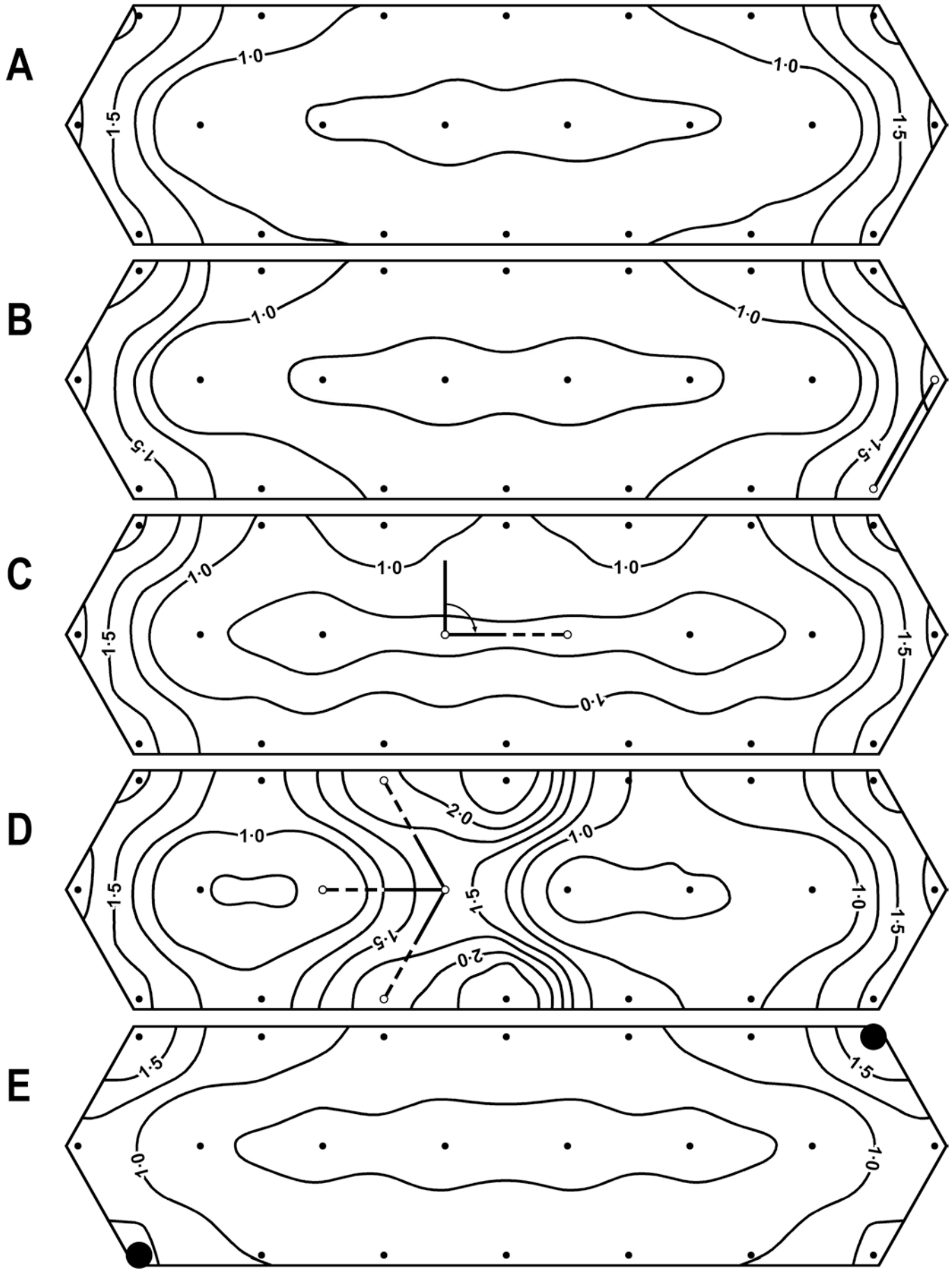

A
1·0
1·0
1·5
1·5
B
1·0
1·0
1·5
1·5
C
1·0
1·0
1·0
1·5
1·5
1·0
D
2·0
1·0
1·0
1·5
1·5
1·5
1·0
2·0
1·5
E
1·5
1·5
1·0
1·0

Figure 5

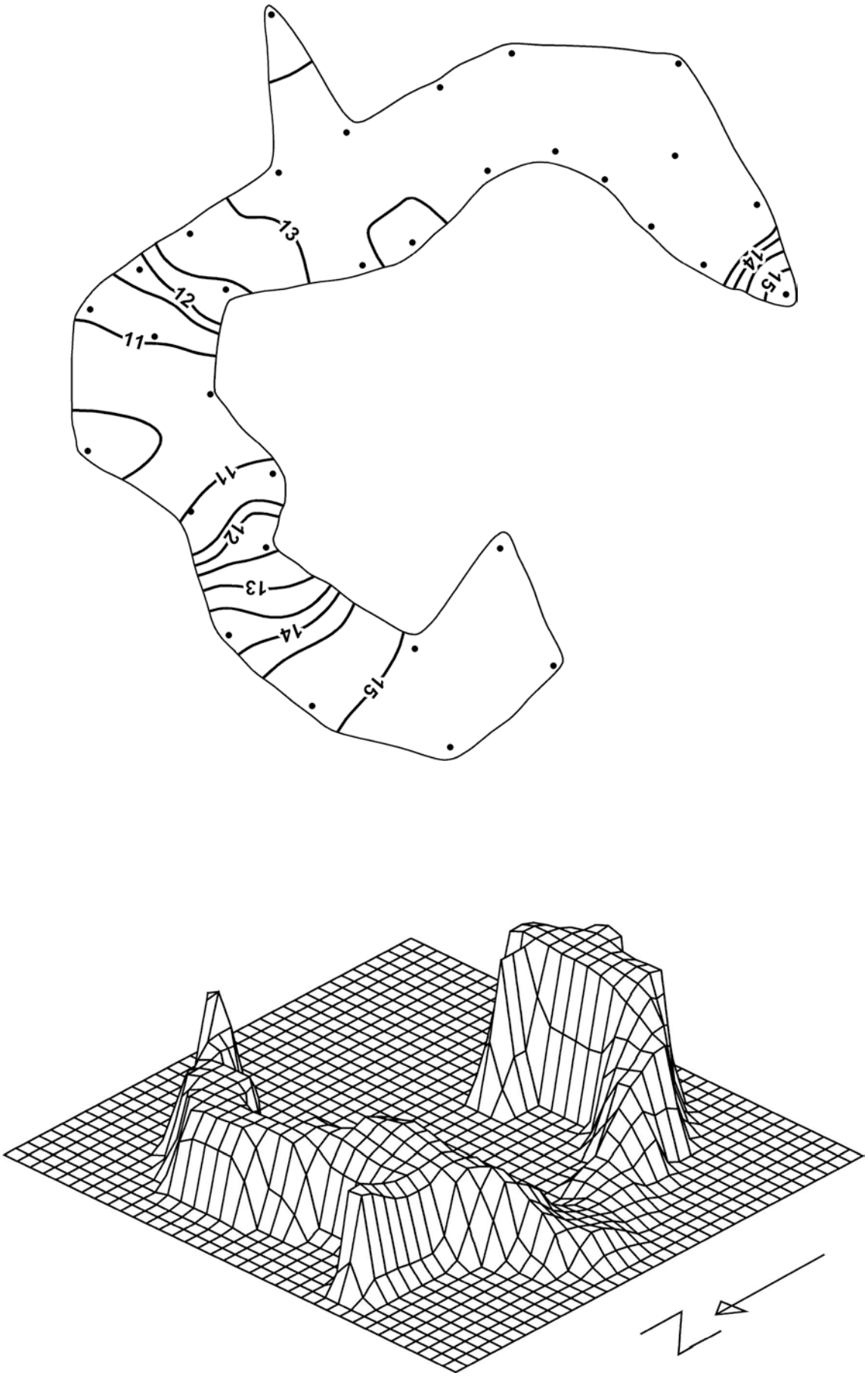


Figure 6

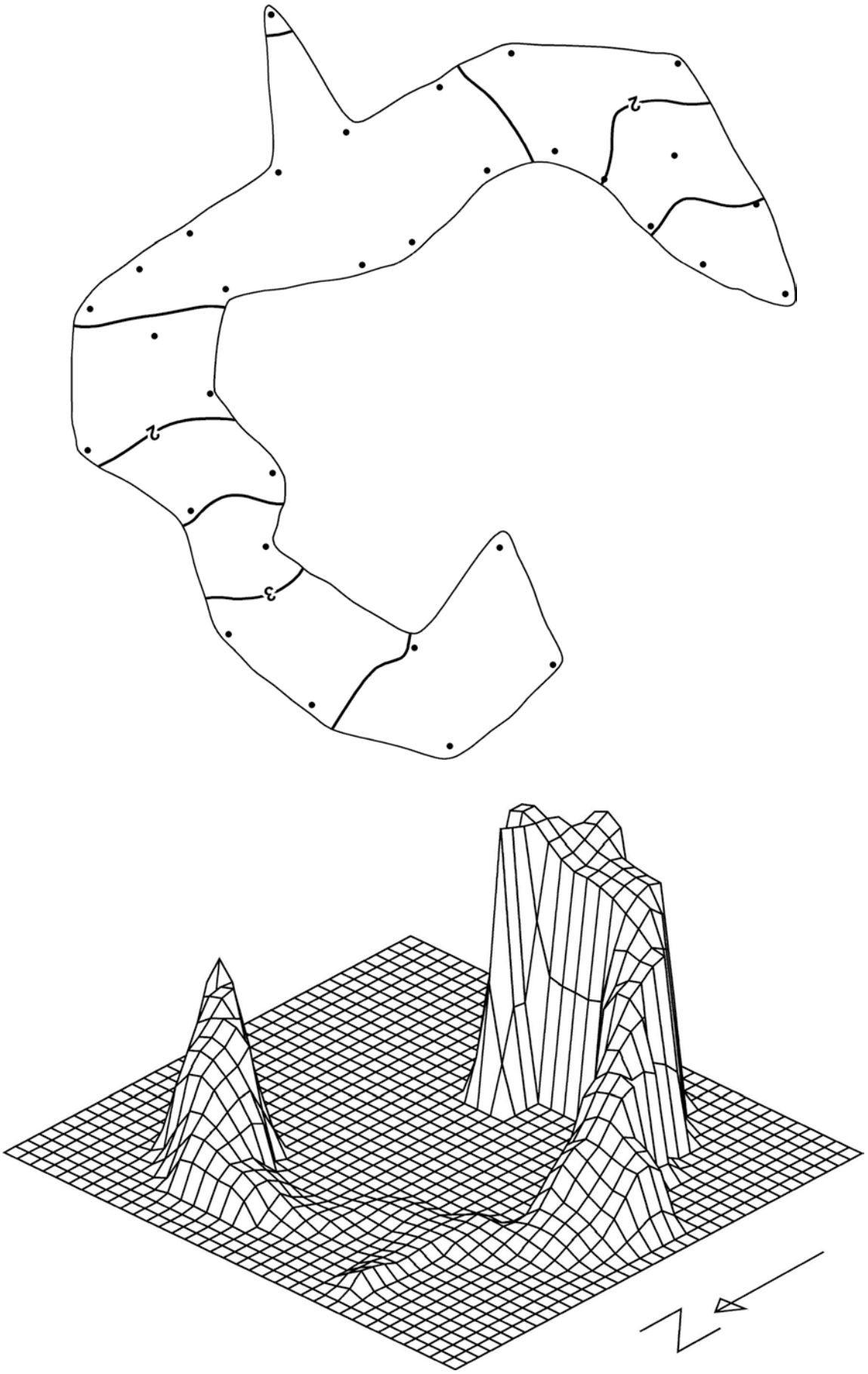


Figure 7

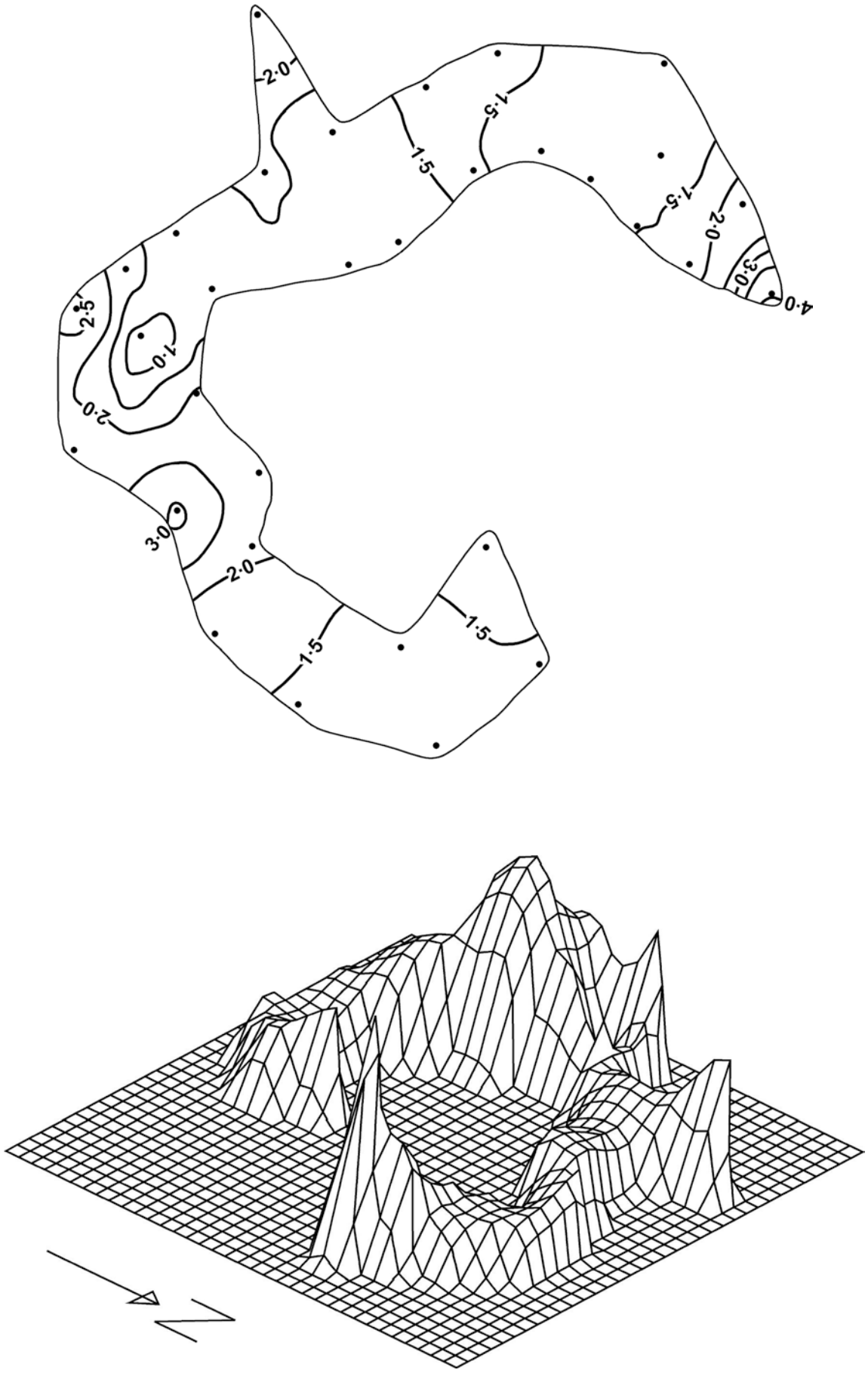


Figure 8

SYSTEMATIC ERRORS IN PHOTOGRAMMETRIC POINT DETERMINATION

Wolfgang FÖRSTNER
Stuttgart, Federal Republic of Germany

ABSTRACT

The refinement of the functional model used for photogrammetric point determination has lead to a significant increase of the accuracy, being about 3-8 µm at photoscale. It is discussed how the functional or the stochastical model may be further refined to compensate for varying, systematic effects and for local distortions which are caused by time-dependent changes of the flight and measuring conditions.

1. Introduction

1.1 During the last 15 years aerial triangulation has increasingly been applied for point determination as a basic for subsequent mapping and for network densification. This is mainly due to the use of rigorous adjustment procedures exploiting the potential of the highly developed hardware components especially films, cameras and measuring equipment. If a number of prerequisites are fulfilled (4-fold overlap, selfcalibration, targeted points, statistical test procedures) one can reach a precision of about 3-8 µm at image scale in all three coordinates and also with reliability values $\bar{\delta} \leq 5$ is able to guarantee for the quality of the result.

1.2 Photogrammetric and geodetic point determination have comparable characteristica. As an image represents a bundle of rays photogrammetry is a 3-dimensional and purely geometrical method using angular information only. Thus there is no separation of planimetry and height (as long as the bundle method is used) and no assumptions about the geoid are necessary. As a method for the densification of point fields it essentially depends on control points or at least some scale information, especially in close range applications. The feature of the image conserving the metrical information may be a reason to prefer photogrammetry in deformation analysis, provided the relative precision of 3-10 ppm is sufficient.

1.3 The underlying mathematical model in most cases simply is the perspective relation between the image and the terrain points. The introduction of additional unknowns, i.e. application of the so-called selfcalibration technique, for the compensation of systematic errors is widely applied and has proven to be effective. It has leads to a rather good coincidence between empirical results and theoretical prediction. This is astonishing, as the stochastical model is still oversimplified: In most

cases the photogrammetric observations are assumed to be uncorrelated and of equal precision.

1.4 Though further refinements of the functional model by using different groups of parameters for each strip have lead to an increase of the precision, this approach is not satisfying, primarily because it is an ad hoc solution, which just argues analytically and does not reflect reality. But also the control of the stability of the system have caused serious problems demonstrating the imperfection of the attempt.

On the other hand any refinement of the stochastical model has to cope with numerical difficulties, which however can only be a short termed argument considering the future computer facilities.

But in both cases the justification of further refinements require comprehensive empirical tests which themselves have to be justified by the theoretically founded formulation of a group of competing hypothesis.

1.5 This paper is supposed to discuss the possibilities of refining the mathematical model of photogrammetric point determination. Section 2 gives a motivation, classifying the error sources within the photogrammetric measuring process with respect to the treatment in the mathematical model. Based on the theoretical influence of systematic errors onto the adjustment result section 3 deals with limitations of the estimation procedures in order to check the necessity of establishing certain types of hypothesis. In section 4 some empirical results about the effect of refined models onto the accuracy are compiled, especially considering the duality of functional and stochastical models. These results are used in section 5 for the formulation of two equivalent refinements of the mathematical model.

2. Error sources in photogrammetric point determination

2.1 The photogrammetric measuring process consists of several distinct steps each being influenced by physical effects which disturb the ideal geometry of the "perspective" model.

Table 1 lists the different stages and gives a classification of the type of error with respect to a possible subsequent refinement of the mathematical model discriminating

- constant systematic errors
- variable systematic errors

- correlations between different points
- correlations between the coordinates of one point
- variations of variance.

Table 1: Classification of types of error sources

	systematic errors		correlation		variance
	const.	var.	global	local	
object (point definition, illumination)		x		x	x
atmosphere (refraction)		x	x	x	
aeroplane (turbulence of atmosphere)		x	x	x	
objective (lens distortion)	x	(x)			
pressure plate (moving part)	x	x			
film (emulsion)			x	x	x
image motion	x			x	x
film development		x	x	x	x
copy		x	x	x	
measuring (comparator, contrast)	x			x	x
corrections	x	x			

Without going into detail table 1 demonstrates that most of the errors are varying with time, at least cannot be treated as constant, lead to correlations and to variations of the variance. Main effects are caused by

- the difference between calibration and real disposition
- the influence of the atmosphere and the film development
- the instability of the instruments, especially the pressure plate, which is the moving part in the camera.

Of course one must keep in mind the absolute size of the variations being only a few micron, but significantly larger than the pure measuring error.

2.2 Though most of the effects have been investigated (cf. the comprehensive report by Schilcher, 1980) no general physical model is available. This is due to be difficulty to study the interaction of the different effects under realistic conditions and even if this would be possible the calibration had to be performed for each project, requiring a reduction of the number of free parameters, which then scarcely would be separable.

2.3 This is the reason why test field calibration only aims at the combined constant part of all possible effects. Up to a great extent this also holds for the selfcalibration technique. But this attempt without much additional field work can be extended towards a more general analy-

tical model for use in practice. The refinement does not necessarily have to reflect the physics of the photogrammetric process completely, but rather can be set up considering aspects of performance as numerical stability, ability to estimate parameters or to evaluate the result in a simple manner. Of course the actual reasons for the deviations from the idealized model have to be investigated, e.g. in order to improve the instruments. But the development of practical procedures has to be done independently.

3. Theoretical considerations

The possibilities of refining the mathematical model cover a great range of special alternatives. In order to get information, which extensions can be checked, e.g. by statistical means, and which type of hypothesis are not discernable we investigate the influence of errors in the mathematical model onto the result. Only those systematic errors are necessary to be modeled which really may distort the result.

Notation: Small letters designate scalars and vectors, capital letters matrices, stochastical variables are underscored. Model errors are designated with a ∇ (nabla) in front of the variable.

3.1 Let the linearized model be given

$$\underline{l} = A\,x + \underline{\varepsilon},\quad \underline{\varepsilon} \sim N(0, C_{\varepsilon\varepsilon}) \qquad (1)$$

with the observation vector $\underline{l} = (\underline{l}_i)$, the design matrix A, the u unknowns x and the model errors $\underline{\varepsilon}$. Their covariance matrix $C_{\varepsilon\varepsilon}$ is assumed to be known. In order to compensate for systematic errors this model can be extended to the mixed model

$$\underline{l} = A\,x + H\,\underline{s} + \underline{\varepsilon},$$
$$\underline{\varepsilon} \sim N(0, C_{\varepsilon\varepsilon}),\ \underline{s} \sim N(s_0, C_{ss}). \qquad (2)$$

The additional parameters $\underline{s}$ are treated as stochastical variables with an unknown mean s_0 and a known covariance matrix C_{ss}. It is well known that this model has two equivalent formulations which can be used for practical parameter estimation.

The first one is useful, if $C_{\varepsilon\varepsilon}$ and C_{ss} are diagonal matrices:

$$\underline{l} = A\,x + H\,s_0 + H\,t + \underline{\varepsilon},\qquad \underline{\varepsilon} \sim N(0, C_{\varepsilon\varepsilon})$$
$$\underline{0} = \qquad\quad t + \underline{\varepsilon}_t,\qquad \underline{\varepsilon}_t \sim N(0, C_{ss}). \qquad (3)$$

The second one in general leads to a full covariance matrix of the observations:

$$\underline{l} = A\,x + H\,s_o + \underline{\eta}, \qquad \underline{\eta} \sim N(0, C_{\varepsilon\varepsilon} + H\,C_{ss}\,H') \tag{4}$$

As all three formulations are statistically identical thus lead to the same results, it is possible to replace the extended functional model eq. (3) by an extended stochastical model eq. (4) and vice versa.

Model eq. (3) is frequently used in photogrammetric block adjustment with $s_o = 0$; often the second group of observation equations is omitted treating the additional parameters as free unknowns. In some cases at least very low weights are given to these fictitious observations just in order to achieve a stable solution (cf. Kilpelä, 1980).

We will now discuss the influence of not detected errors in the mathematical model onto the estimated unknowns $\underline{\hat{x}}$, starting from the original model eq. (1).

3.2 Systematic errors $\nabla l = H\,\nabla s$ in the observations cause changes of the unknown $\underline{\hat{x}}$

$$\nabla x = (A'\,P\,A)^{-1} A'\,P\,H\,\nabla s \tag{5}$$

with $P = \sigma^{-2} C_{\varepsilon\varepsilon}^{-1}$. A scalar measure is

$$\bar{\delta}^2 = \nabla x'\,C_{xx}^{-1}\,\nabla x = \nabla s'\,H'\,P\,A\,(A'\,P\,A)^{-1} A'\,P\,H\,\nabla s/\sigma^2 \tag{6}$$

which describes the total deformation of the network (including orientation parameters). This deformation is zero only if $A'\,P\,H = 0$:

$$\boxed{\bar{\delta}(\nabla l) = 0, \qquad \nabla l = H\,\nabla s \leftrightarrow A'\,P\,H = 0} \tag{7}$$

thus if the parameters are orthogonal to the unknowns.

3.3 Errors in the stochastical model do not influence the unbiasedness of the estimate $\underline{\hat{x}}$. But a wrong weight coefficient matrix

$$\bar{Q} = \sigma^{-2}\,\bar{C} = Q + \nabla Q = Q + H\,\nabla Q_{ss}\,H' \tag{8}$$

leads to a change ($\nabla Q << Q$, cf. Koch, 1980, p. 167)

$$\underline{\nabla x} = (A'\,P\,A)^{-1}\,A'\,P\,\nabla Q\,P\left(I - A\,(A'\,P\,A)^{-1}\,A'\,P\right)\underline{l} \tag{9}$$

in the estimates $\underline{\hat{x}}$. The expectation $E\left(\underline{\nabla x}'\,C_{xx}^{-1}\,\underline{\nabla x}\right)$ of the total deformation then results in

$$\bar{\delta}^2 = E\left(\underline{\nabla x}'\,C_{xx}^{-1}\,\underline{\nabla x}\right) = \mathrm{trace}(P\,\nabla Q\,P\,Q_{\bar{l}\bar{l}}\,P\,\nabla Q\,P\,Q_{vv}) \tag{10}$$

with the weight coefficient matrices $Q_{\bar{l}\bar{l}} = A\,(A'\,P\,A)^{-1} A'$ and $Q_{vv} = Q - Q_{\bar{l}\bar{l}}$ of the adjusted observations $\underline{\bar{l}} = A\,\underline{\hat{x}}$ and the residuals $\underline{v} = \underline{\bar{l}} - \underline{l}$ resp..

With $\nabla P = P\,\nabla Q\,P$, the vec operator, which maps a matrix into a vector and the Kronecker product $\otimes$, eq. (10) can be written as a quadratic form

$$\bar{\delta}^2 = (\mathrm{vec}\nabla P)'(Q_{\overline{ll}} \otimes Q_{vv})\,\mathrm{vec}\nabla P \qquad (10a)$$

If furthermore ∇P is a diagonal matrix this expression can be simplified using the Hadamard product $A * B = (a_{ij}\, b_{ij})$ (cf. Pukelsheim, 1977)

$$\bar{\delta}^2 = (\mathrm{diag}\nabla P)'\,(Q_{\overline{ll}} * Q_{vv})\,\mathrm{diag}\nabla P \qquad (10b)$$

where ∇P is a vector only containing the diagonal elements of ∇P.

Example: Starting from eq. (10b) with $Q = I$ the low influence of single weight errors ∇p_i onto the coordinates is proved, as in this case $\bar{\delta}_i = \sqrt{r_i\,(1-r_i)}\,\nabla p_i < \nabla p_i/2$, $r_i = (Q_{vv}\,P)_{ii}$ being the redundancy number of the observation $\underline{l}_i$.

From eq. (10) we now again derive conditions for $\bar{\delta} = 0$ using the decomposition $\nabla Q = H\,\nabla Q_{ss}\,H'$, which may be interpreted as a neglected set of additional stochastical parameters $H\underline{s}$. The influence of errors in the covariance matrix onto the coordinates is zero if $\nabla Q\,P\,Q_{\overline{ll}} = 0$ or $\nabla Q\,P\,Q_{vv} = 0$. The interpretation of the second condition is simplified if we write the functional model in terms of condition equations, i.e. according to standard problem I in the terminology of Tienstra: $U'\underline{l} = \underline{w}$. Then $U'A = 0$ and $Q_{vv} = Q\,U\,(U'\,Q\,U)^{-1}\,U'\,Q$. The conditions for $\bar{\delta} = 0$ then read as

$$\boxed{\bar{\delta}\,(\nabla Q) = 0,\ \nabla Q = H\,\nabla Q_{ss}H' \quad \leftrightarrow \quad \begin{array}{l} 1.\ A'\,P\,H = 0 \ \text{ or} \\ 2.\ U'\,H = 0 \end{array}} \qquad (11)$$

This result already has been found by Rao (1967), here however it is derived from the general expression eq. (10).

The conditions eq. (11) have a geometric meaning: Neglected stochastic parameters have no influence onto the estimation of x, if they are already contained in the functional model as a linear combination either of the condition equations or of the parameters. This is because $A'\,P\,H = 0$ is equivalent to $H \in \mathrm{col}(U)$ and $U'H$ is equivalent to $H \in \mathrm{col}(A)$, as $A'\,U = 0$ and the matrix $(A\ U)$ has full rank. col(A) designates the column space of A.

Example: The arithmetic mean is invariant with respect to (equal) correlations between the observations. Here $A' = (1,1,\ldots,1) = H'$, $Q = I$ and $\bar{Q} = I + \rho/(1-\rho)\,H\,H'$ and $H \in \mathrm{col}(A)$.

The second condition eq. (11.2) could not be found for errors in the functional model (cf. eq. (6)) as they would have lead to a singular normal equation matrix. But practically additional parameters which are very similar to already existing ones do not deteriorate the result, if the solution is stable enough to avoid rounding errors.

Thus eq. (11) gives complete conditions for additional parameters, fix and stochastical ones, to have no influence on the result. Additional parameters meeting these conditions then are not estimable in an extended model. The normal equation matrix $(AH)' P (A' H')$ for x and s_o is singular. Moreover, if the covariance matrix C_{ss} is to be estimated, the equation system for the unknown variance components describing C_{ss} will be singular showing that the variance components are either not estimable or not discernable.

The common conditions may form a basis for a joint evaluation of simultaneous refinements of the functional and the stochastical model.

4. Empirical results

The last section has provided some tools to evaluate possible extensions of the mathematical model. The following results of practical investigations want to show how far a mathematical model is able to represent reality and which further increase of the final accuracy one might expect.

4.1 The first example deals with the bundle block adjustment of the test block Appenweier (Klein, 1980). Table 2 gives the estimated precision $\hat{\sigma}_o = \sqrt{v' P v / r}$ of the image coordinates and the r.m.s. and the maximum errors at the 85 check points which were not used in the adjustment, both for single blocks with 20 % sidelap and for double blocks with 60 % sidelap. The adjustment has been performed 1. without any refinement of the model, 2. with 12 additional parameters common for all images of each block to compensate for systematic image errors and 3. with 12 parameters for each strip in order to consider possible differences of the deformations between the strips. Not determinable parameters were excluded to obtain a stable solution. We are only concerned with the planimetric results here.

The table shows clearly:
- The accuracy increases with increasing refinement of the model. This proves that the additional parameters really compensate for varying systematic errors.
- The maximum errors significantly decrease in the single blocks, which is of utmost importance for practical applications.

Table 2: Accuracy of bundle block adjustment Testblock Appenweier scale 1 : 7 800, area 9.1 x 10.4 km², estimated precision of terrestrial control and check point coordinates 1.2 cm

version	single blocks (sidelap 20 %)				double blocks (sidelap 60 %)			
	$\hat{\sigma}_o$	μ_{xy}	ε_{max}	$\frac{\mu_{xy}}{\hat{\sigma}_o}$	$\hat{\sigma}_o$	μ_{xy}	ε_{max}	$\frac{\mu_{xy}}{\hat{\sigma}_o}$
	µm	-	cm	-	µm	-	cm	-
1 no parameters	3.0	5.7	42.9	2.4	3.6	3.4	10.4	1.2
2 12 parameters blockwise	2.4	3.8	18.6	2.0	2.7	2.6	9.3	1.2
3 12 parameters stripwise	2.3	3.4	13.8	1.9	2.6	2.0	7.0	1.0

$\mu_{xy} = \sqrt{\mu_x^2 + \mu_y^2}$; μ_x, μ_y = r.m.s. residuals at 85 check points

ε_{max} = maximum residuals at check points

- The results, though extremely good, are not quite in accordance with theory as the ratio $\mu_{xy}(\mu m)/\hat{\sigma}_o(\mu m)$ should be 0.9 for single and 0.6 for double blocks. This discrepancy may be explained by neglections in the mathematical model, neglected correlations between the image coordinates and certainly also unrealistic assumptions about the precision of the control and check points, which have an average precision of 1.2 cm.

A plot of the parameter values (not shown here) reveals them to vary significantly from strip to strip. As there is no justification for this type of splitting the block with respect to the setup of the additional parameters the systematic errors of the individual images were investigated.

4.2 The systematic errors of time series of up to 76 images were derived from flights with reseau cameras (Schroth, 1982). The deviations from the ideal reseau reflect the deformation of the film caused by film transport, pressure plate, film development, temperature, humidity during the measurement etc.. The time dependency of the deformations is described by the time series of 18 orthogonal parameters, namely the 6 parameters usually needed for the orientation and 12 additional parameters. Fig. 1 - 3 give some representative examples.

The parameters significantly differ from zero. The mean value is given by a straight line. The dotted lines indicate the 3-fold standard deviation of the estimated parameter. Obviously the variation of the parameters cannot be explained by random errors only.

Also the type of the time series varies. Most of the time series show no correlation between the images (cf. fig. 1). Some time series seem to have

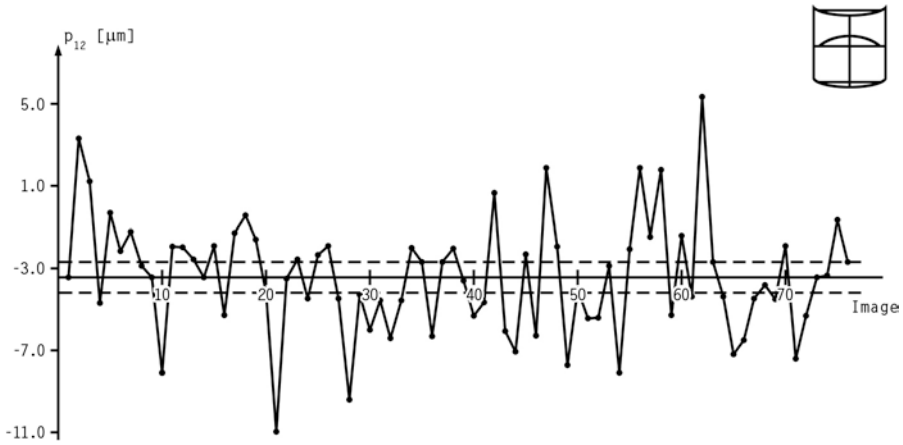


Fig. 1 Standardized values of parameter p_{12}

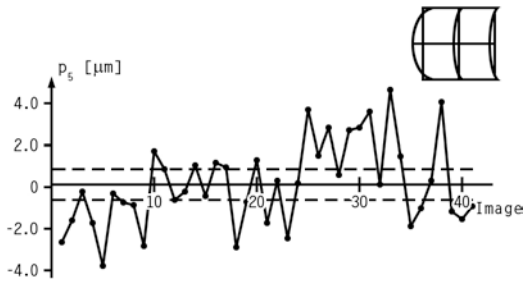


Fig. 2 Standardized values of parameter p_5

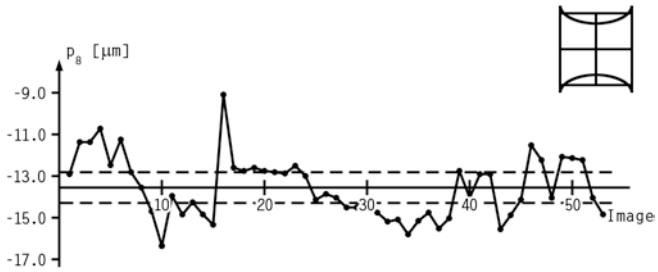


Fig. 3 Standardized values of parameter p_8

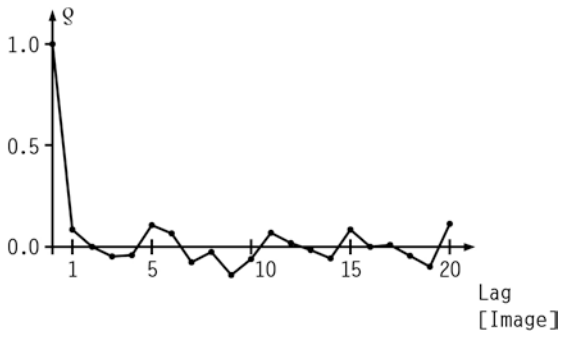


Fig. 4 Autocorrelation function of time series fig. 1

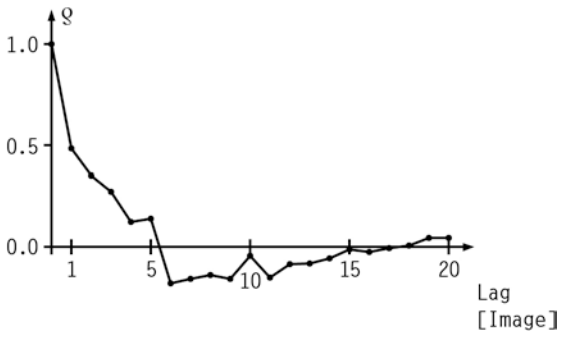


Fig. 5 Autocorrelation function of time series fig. 3

(from Schroth, 1982)

a trend which cannot be described by a constant value (cf. fig. 2). Some time series show rather large correlations between the parameters of adjacent images (cf. fig. 3). The autocorrelation functions of the time series fig. 1 and 3 are given in fig. 4 and 5. The autocorrelation function fig. 5 can be approximated by an exponential function exp(-c·Lag), being characteristic for a discrete 1st order Markov-process.

This investigations shows that the additional parameters may be modelled as stochastic variables.

Also time series for the variation of the scale of the images have been obtained showing high variations, which mainly are caused by the humidity during the measuring process. But this parameter only has to be modelled if geodetic measurements (e.g. with inertial systems) are used to get information about the position of the camera platform. Otherwise scale variations are absorbed by the z-coordinate of the projection centre. Thus the scale parameter of an image is fulfilling one of the conditions eq. (11), namely (11.2).

4.3 In section 3 it was shown, that the refinement of the mathematical model can be achieved by either extending the functional or the stochastical model. Of course one can also think of mixtures. Schilcher (1980) has proved this empirically.

He analysed 120 images taken with two different cameras, a wide angle and a super wide angle camera, flown over the test field Rheidt. He distinguished 3 functional models of different quality shown in table 3.
A is the most simple, C the most refined functional model.

Table 3 Theoretical precision of checkpoints for different functional and stochastical models (after Schilcher (1908)).

	model				
	A	B	C	C'	
a priori corrections: lens distortion, earth curvature, refraction	no	yes	yes	yes	
systematic errors common to all images	no	no	yes	yes	
covariance matrix	$\hat{C}_A$	$\hat{C}_B$	$\hat{C}_C$	$1\cdot\hat{\sigma}_o^2$	
precision wide angle	3.7	3.8	3.7	4.3	µm
precision super wide angle	6.7	5.5	5.1	5.8	µm

From the residuals after a spatial resection he estimates covariance matrices $\hat{C}_A$, $\hat{C}_B$ and $\hat{C}_C$. The theoretical precision of check points in terms of standard errors are derived by error propagation for the different levels

of the functional model and the corresponding estimated covariance matrices. For the most refined functional model C also the results using a diagonal matrix $C = l\,\hat{\sigma}_0^2$ are given.

The standard errors for the first three models confirm that the neglections caused by a very simple functional model can be compensated by a refined stochastical model and vice verca. The coincidence of the standard deviations for the wide angle camera can be said to be excellent. The moderate deviations of the values for the super wide angle between the models A and B can be explained by the large constant a priori corrections. Their neglection can not be fully compensated by an appropriate covariance matrix. In this case the assumption for the equivalence of the models eq. (3) and (4), namely the variability of the parameters, is not met.

The model C' using an extended functional model and an oversimplified stochastical model obviously gives worse results, again demonstrating that the introduction of constant common parameters leaves significant correlations between the observations.

The investigation clearly shows that the freedom of choosing between a refined functional and a refined stochastical model is not only a theoretical statement but can be realized leading to a further increase of the final accuracy.

5. A refined mathematical model for photogrammetric point determination

5.1 The structure of the photogrammetric measuring process and the empirical results suggest to treat the images as a time series, whose deformations may be modelled by a Markov-process.

The deformations thus can be described by the following 1st order autoregressive scheme

$$\begin{aligned} \underline{t}_1 &= \underline{\eta}_1, \qquad \underline{\eta}_1 \sim N(0, C_{ss}) \\ \underline{t}_k &= a\,\underline{t}_{k-1} + \underline{\eta}_k, \qquad \underline{\eta}_k \sim N\left(0, (1-a^2)\,C_{ss}\right),\ |a| \le 1,\ k \ge 2 \end{aligned} \tag{12}$$

The parameter vector $\underline{t}_1$ is the starting point. The parameter a controls the degree of correlation between the parameters $\underline{t}_k$ and $\underline{t}_{k'}$ of different images k and k', being $a^{|k-k'|}$. Thus if a = 1 all parameters are equal ($\underline{\eta}_k = 0$), whereas if a = 0 the parameters are independent.

The stochastic process is observed by measuring the coordinates, contained in the vector $\underline{l}_k$ for each image:

$$\underline{l}_k = A_k\, x + U_k\, t_o + U_k\, \underline{t}_k + \underline{\varepsilon}_k, \quad \underline{\varepsilon}_k \sim N\left(0, C_{ll}^{(k)}\right) . \qquad (13)$$

The unknown parameters x contain the coordinates of the new points and the orientation parameters of each image, t_o is a constant vector of additional parameters describing the mean deformation of the images.

5.2 This mixed model eqs. (12) and (13) has a similar structure as eq. (2). Thus a first way to estimate the unknowns x and t_k (k = (0,1,...) is solving the equation system (cf. eq. (3))

$$\begin{aligned} \underline{l}_k &= A_k\, x + U_k\, t_o + U_k\, t_k + \underline{\varepsilon}_k, & \underline{\varepsilon}_k &\sim N\left(0, C_{ll}^{(k)}\right) && (14a)\\ \underline{0}_1 &= -t_1 + \underline{\eta}_1, & \underline{\eta}_1 &\sim N(0, C_{ss}) && (14b)\\ \underline{0}_k &= a\, t_{k-1} - t_k + \underline{\eta}_k, & \underline{\eta}_k &\sim N\left(0, (1-a^2)\, C_{ss}\right) && (14c) \end{aligned}$$

using least squares. Eliminating t_k from eqs. (14b) and (14c) leads to the equivalent model (assuming $\mathrm{Cov}\left(\underline{\varepsilon}_k, \underline{\varepsilon}_{k'}\right)$, cf. eq. (4)).

$$\begin{aligned} &\underline{l}_k = A_k\, x + U_k\, t_o + \underline{\xi}_k,\\ &V\left(\underline{\xi}_k\right) = C_{ll}^{(k)} + U_k\, C_{ss}\, U'_k, \quad \mathrm{Cov}\left(\underline{\xi}_k, \underline{\xi}_{k'}\right) = U_k\, C_{ss}\, U'_{k'}\, a^{|k-k'|} \end{aligned} \qquad (15)$$

Both forms of the refinement have their advantages:

The extension eq. (14) of the functional model is favourable in large systems, especially in cases with high point density in the images. The number of parameters increases roughly proportional to the number of images. The banded structure of the normal equation system can be preserved. Main advantages is the ease of evaluation the extension, e.g. by testing the $\hat{\underline{\eta}}_k$. Estimating the parameter a is simple, estimating the covariance matrix C_{ss} is feasible. The reduction of the normal equations onto the orientation and additional parameters easily is possible, if the observations can be treated uncorrelated, at least if no correlations between points can be assumed.

Otherwise an extension of the functional model is unavoidable. This especially holds if local film distortions or similar effects cause correlations between points over short (< 1 cm) distances within an image. These deformations can not be compensated extending the functional model but rather have to be described by covariance functions. Thus $C_{ll}^{(k)}$ may be split into two additive components representing measuring errors and local film distortions. Refining the stochastical model so has the advantage of allowing all kinds of correlations without increasing the numerical effort.

5.3 Though the proposed extension of the mathematical model for photogrammetric point determination could be motivated by the results of quite a number of experiments several problems are open for future investigations:

1. Empirical tests with extended models (eqs. (14) and (15)) have to prove the efficiency of the refinement. Special effort has to be layed upon the question whether the compensation of local distortions really leads to a significant increase of the final accuracy.

2. The numerical effort for the adjustment which is heavily increased by the proposed extensions has to be limited. Therefore it seems necessary
   - to find an optimal set of additional parameters which is at the same time as small as possible,
   - to find a strategy for a stepwise refinement of the mathematical model, if that is of any advantages and
   - to compare the numerical properties of the two approaches eq. (14) and (15) with respect to computing time and stability of the system.

3. The increase of unknown parameters may weaken the whole system if the geometry of the block is not chosen properly. Therefore the quality of the result has at least to be checked. These checks should be as efficient and at the same time simple as possible and may be used to give recommendations for the design of blocks with the aim to reach a high reliability of the result.

It can be hoped that these investigations will lead to a further increase of the accuracy of photogrammetric point determination.

Literature

Kilpelä, E.: Compensation of Systematic Errors of Image and Model Coordinates. Int. Arch. of Photogr., Vol. B9, XIVth Congress of ISP, Hamburg 1980, p. 407-427

Klein, H.: New Results of Bundle Block Adjustments with Additional Parameters. Schriftenreihe des Inst. f. Photogr., Stuttgart, Heft 6, 1980, p. 13-30

Koch, K.-R.: Parameterschätzung und Hypothesentests. Dümmler, Bonn, 1980

Pukelsheim, F.: On Hsu's Model in Regression Analysis. Math. Operationsforsch. Statist., Ser. Statistics, Vol. 8 (1977), no. 3, p. 323-331

Rao, C. R.: Least Squares Theory using a Estimated Dispersion Matrix and its Application to Measurement of Signals. Fifth Berkely Symposium, Vol. I (1967), p. 355-372

Schilcher, M.: Empirisch-statistische Untersuchung zu Genauigkeitsstruktur des photogrammetrischen Luftbilds. DGK C, Heft 262, München 1980

Schroth, R.: On the Stochastic Properties of Image Coordinates. Pres. Paper to Symposium of Comm. III of ISP, Helsinki 1982

DATA-SNOOPING IN CONTROL NETWORKS

Henk DE HEUS
Delft, Netherlands

ABSTRACT

In this paper, the testing of observations according to the B-method is explained. Attention is paid to the reliability achieved by this way of data-snooping and the possible influence of non-detected errors.
The choice of the testing parameters is discussed. Finally the practical application of the testing procedure is discussed and illustrated by some examples.

1. Introduction

Testing should be an essential part in any dataprocessing procedure of geodetic observations, provided that we are not only interested in the numbers of the resulting qualities (often adjusted coordinates), but in their quality as well. This quality is not only given by the precision, as described in the variance-covariance matrix, for the precision indicates the possible variations in the results, provided that the assumptions about the stochastic and mathematical model for the observations (the so-called null-hypothesis) are valid. If not, when for instance standard deviations of observations are adopted wrongly, or because of gross errors in the observations, the results can differ much more than the variations allowed by the precision. That is why reliability, describing the possibilities to detect gross errors by testing and the influence of non-detected errors on the results is at least as important as precision is.
Later in this paper it will be discussed how to describe and characterize the reliability.
We now first turn to the testing procedure, for the possibility to detect gross-errors depends on the applied tests. In the extreme case where no tests are carried out one cannot detect any error at all and reliability is even imaginary.

2. Tests

The aim of testing is to check the null-hypothesis Ho:
- the chosen variance-covariance matrix of the observations describes the real stochastic deviations of the observations,

- the mathematical model gives a right description of the relations between the observations (and unknowns),
- there are no gross errors made in the observations.

The common test to check these assumptions is:

$$\frac{E}{\sigma^2} \leq \chi^2_{1-\alpha;b}$$

or

$$\frac{\hat{\sigma}^2}{\sigma^2} = \frac{E}{b\,\sigma^2} \leq F_{1-\alpha;b,\infty}$$

where:

E:	shifting variate from the least-squares adjustment
σ^2:	a priori variance factor
$\hat{\sigma}^2$:	a posteriori variance factor
b:	number of condition equations (degrees of freedom)
$\chi^2_{1-\alpha;b}$:	critical value of the χ^2-probability distribution
$F_{1-\alpha;b,\infty}$:	critical value of the F-probability distribution
α:	probability Ho will be rejected wrongly

In case the test variate is smaller than the critical value, the null-hypothesis is accepted and testing is ready. If not, the null-hypothesis is rejected and one has to find the cause of the rejection and errors in observations or assumptions have to be improved.

In the latter case there is always a (small) probability α the null-hypothesis is wrongly rejected. On the other hand, accepting the null-hypothesis does not necessarily mean all assumptions in the null-hypothesis are right. The corrections of the observations just do not indicate the null-hypothesis is wrong, but, in spite of this, depending on the power of the test, there might still be something wrong (for instance some errors in the observations). Because the null-hypothesis then should be rejected, this is a wrong decision too.

With respect to the quality of the network it is interesting to know how large possible non-detected errors might be. We now assume such an error can be denoted by a so-called alternative hypothesis Ha_p as:

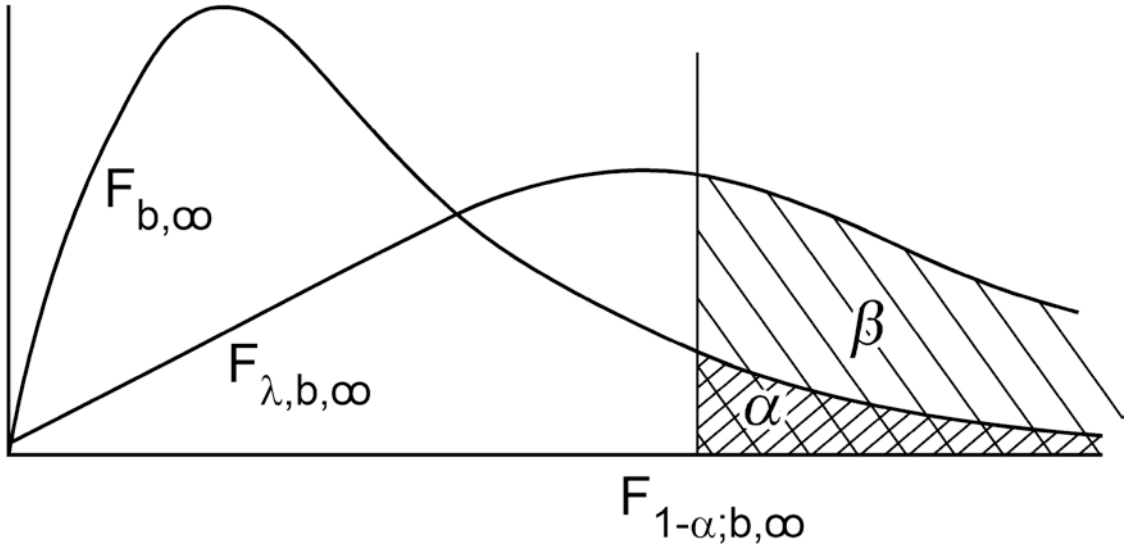
$$Ha_p : \nabla_p \tilde{l} = c_p \cdot \nabla_p$$

The vector of errors in the observations $\nabla_p \tilde{l}$ is supposed to be expressed by a chosen vector c_p and a scale-parameter ∇_p; for instance: c_p contains

the coefficients of a (linearized) refraction model, or, assuming one observation is wrong, c_p is an unit-vector.

By assuming a value for ∇_p, a particular error according to the alternative hypothesis Ha_p is assumed.

Let the influence of such an error on the shifting variate be λ_p. Then the testing variate $\hat{\sigma}^2/\sigma^2$ now has a non-central F-probability distribution $F_{(\lambda),b,\infty}$ with expectation $E\left\{\hat{\sigma}^2/\sigma^2 | Ha_p\right\} = 1 + \lambda_p/b$.



Testing $\hat{\sigma}^2/\sigma^2$ now has a probability β (the power of the test) to reject the null-hypothesis, which would be the right decision, but also a probability $1-\beta$ that the null-hypothesis is accepted, although the error is made.

A good reliability requires a β as large as possible. An alternative, more common way to describe the reliability starts with the choice of a fixed power β, typically 0.80. In this way for each alternative hypothesis Ha_p the influence on the shifting variate is fixed at λ, to be computed from

$$\lambda = \lambda(\alpha, \beta, b, \infty)$$

and the corresponding error $\nabla\tilde{}_p l$ can be computed:

$$\nabla\tilde{}_p l = c_p \sqrt{\frac{\lambda}{N_p}}$$

$$N_p = c_p^T \cdot Q_{ll}^{-1} \cdot (Q_{ll} - Q_{\hat{l}\hat{l}}) \cdot Q_{ll}^{-1} \cdot c_p$$

$Q_{ll} = P^{-1}$: covariance matrix of the observations

$Q_{\hat{l}\hat{l}}$: covariance matrix of the least squares estimates of the observations

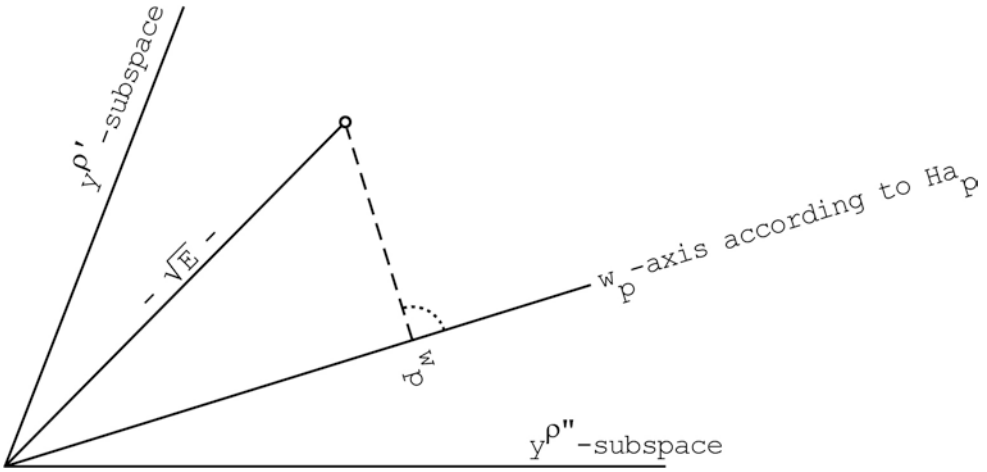
This value, of the error that will be detected with probability β by the test, is called marginally detectable error or boundary value. The smaller the marginally detectable error, the better the reliability according to an alternative hypothesis.

Let us now return to the case the null-hypothesis is rejected. For the above mentioned boundary values are only meaningful provided that the test variate is accepted, we have to find the error(s) and make the null-hypothesis accepted to get a workable description of the reliability.

In case a specific alternative hypothesis Ha_p is true the most sensitive test variate in testing that hypothesis is:

$$w_p = \frac{1}{\sigma\sqrt{N_p}}\, c_p^T \cdot P \cdot v$$

with a standard normal probability distribution.
w_p can be seen as the projection of the shifting variate E on the direction according to the alternative hypothesis in the y^ρ-condition subspace.



w_p is to be tested as follows:

If $|w_p| \le \sqrt{F_{1-\alpha_0;1,\infty}}$ then reject Ha_p
else accept Ha_p

This w-test now can be used to select erroneous observations by testing the group of so-called conventional alternative hypotheses. This testing procedure is called "data-snooping" and is realized by $c_{i'} = \delta_{i'}$. The c-vector in the alternative hypothesis is a unit vector $\delta_{i'}$ successively corresponding to each observation $l_{i'}$.

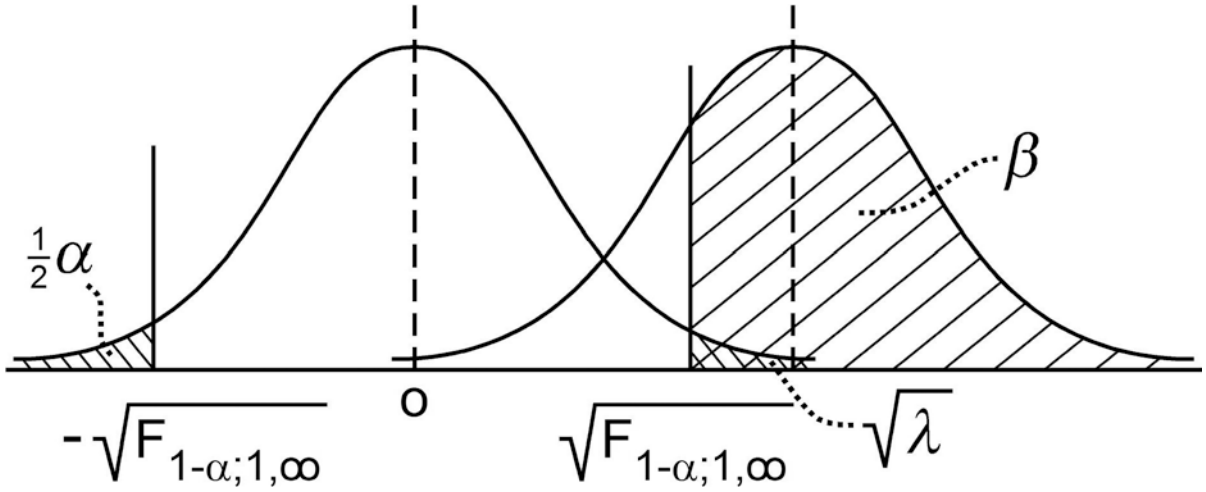
In case the observations are not correlated the test-variate w_i according to a conventional alternative hypothesis becomes simply:

$$w_i = \frac{v_i}{\sigma_{v_i}}$$

the correction of the observation divided by its standard deviation.

Data-snooping is a special application of the more general w-test. More complicated alternative hypothesis can be tested, by a w-test variate, although in practice it is often difficult to specify other alternative hypothesis (to chose the c-vector), especially when the assumption in the null-hypothesis are formulated carefully.
In general one needs external indications about the probability a particular alternative hypothesis will occur, for instance geophysical assumptions about possible deformations. Then these assumptions, translated in an alternative hypothesis (c-vector), can be tested from the deformation measurements using the w-test variate.

Just like the $\hat{\sigma}^2/\sigma^2$-test variate, the w_p-test variate can be smaller than the critical value, although an error as formulated by the alternative hypothesis Ha_p has been made.



After the choice of the power β, the boundary value for such an error can be computed again from

$$\nabla_{\tilde{p}} l = c_p \sqrt{\frac{\lambda}{N_p}}$$

where in this case λ is determined by the testing parameters α_o and β used in the w-test from:

$$\lambda = \lambda(\alpha_o, \beta, 1, \infty)$$

Because the w-test is a one-dimensional test the difference with the λ according to the $\hat{\sigma}^2$-test is the value 1 in the λ -function instead of the value b for the b-dimensional $\hat{\sigma}^2$-test.

3. B-method of testing

The $\hat{\sigma}^2$-test and the w_p-test are not necessarily of equal power in testing an alternative hypothesis Ha_p. The boundary value of each test depends on the values of the testing parameters α and β used in the tests.
In the B-method of testing both tests are related to each other by a special choice of their testing parameters:

$$\lambda_o = \lambda(\alpha_o, \beta_o, 1, \infty) = (\alpha, \beta_o, b, \infty)$$

$$\langle w\text{-test} \rangle \qquad \langle \hat{\sigma}^2\text{-test} \rangle$$

This choice of equal values for λ and β in both tests means that a certain error is detected with the same probability both by the $\hat{\sigma}^2$-test as the w_p-test on that alternative hypothesis. In other words: both tests lead to the same reliability, the same boundary values:

$$\nabla_{p,\tilde{o}} l = c_p \sqrt{\frac{\lambda_o}{N_p}}$$

This choice leads in principle to a very simple testing procedure:
First the $\hat{\sigma}^2$-test is carried out. If Ho is accepted, no further tests are necessary and the reliability for any alternative hypothesis is given by $\nabla_{p,\tilde{o}} l$. Only in case Ho is rejected the data-snooping procedure is carried out to detect an error in one or more observations.

The consequence of connecting both types of tests according to

$$\lambda_o = \lambda(\alpha_o, \beta_o, 1, \infty) = \lambda(\alpha, \beta_o, b, \infty)$$

is that after the choice of α_o and β_o (usually $\alpha_o = 0.001$ and $\beta_o = 0.80$) the value α of the $\hat{\sigma}^2$-test computed from this formula increases with the number of condition equations b. In larger adjustments (b large) this may lead to an unworkable large value for α, for the probability Ho is wrongly rejected should not be too large. In that case the adjustment and the testing procedure should be carried out in steps.

4. Measures for the reliability

In the previous section is already explained that application of the B-method of testing leads to the same reliability for both types of tests, expressed by the boundary value

$$\nabla_{p,o}\tilde{}\, l = c_p \sqrt{\frac{\lambda_o}{N_p}}$$

for any alternative hypothesis Ha_p.
However, it is possible to suppose an unlimited number of alternative hypotheses. Therefore, one should agree upon which alternative hypotheses are used to describe the reliability. Baarda suggested a ***convention***, namely to characterize the reliability of each network by the reliability of the conventional alternative hypotheses used in the data-snooping (this explains the name conventional alternative hypotheses). In this way reliability is defined in a clear and workable way.

The boundary values are called "internal reliability", expressing the size of the error in an observation that can be detected with probability β by testing. Since the final results of the data processing are not the observations, but derived quantities, it is interesting to know how the final results are influenced by possible non-detected errors: the "external reliability".
Let those final results (specific functions for special purpose networks, and coordinates in control networks) be denoted by X. Then, if X is derived from the observations l by

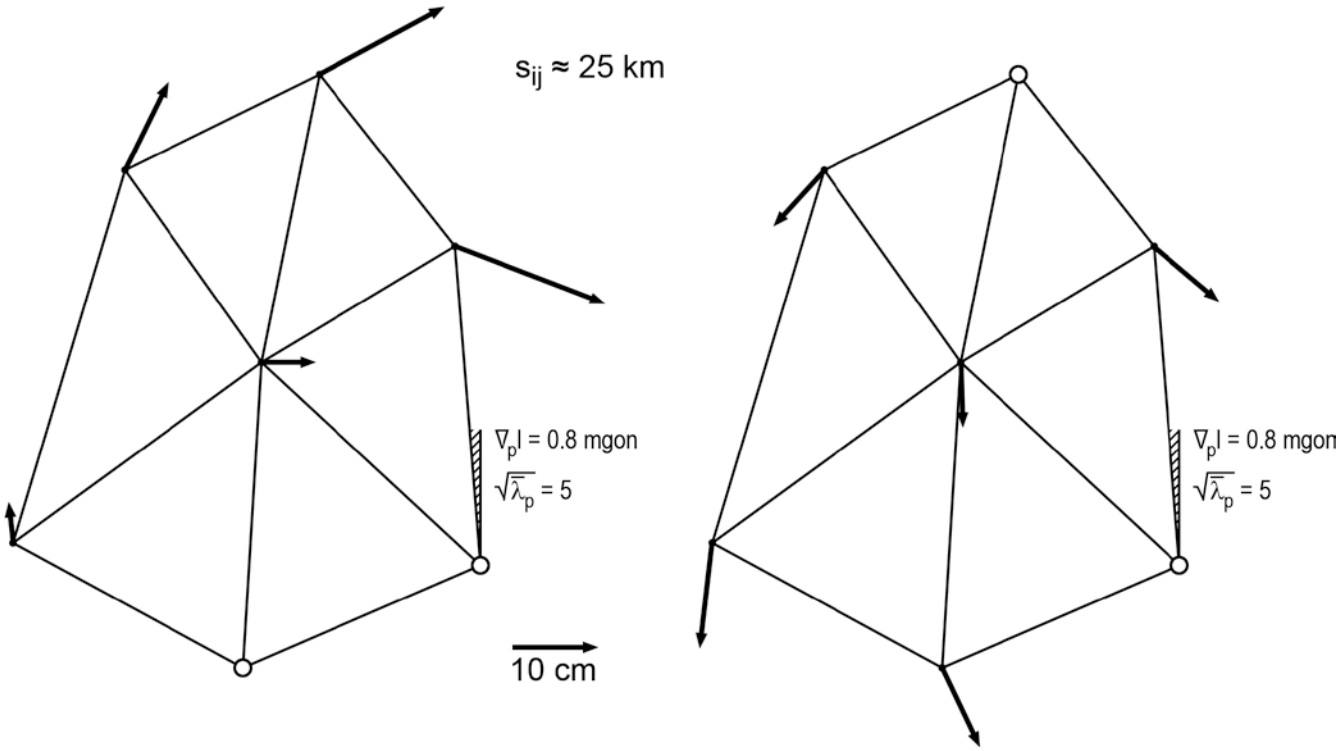
$$X = F \cdot l$$

the influence on X of an undetected error with the size of the boundary value is computed from

$$\nabla_{p,o}\tilde{}\, X = F \cdot \nabla_{p,o}\tilde{}\, l$$

In case X stands for coordinates the vectors $\nabla_{p,o}\tilde{}\, X$ can be visualized as in the picture on the next page is done for a small triangulation network.

The interpretation of those vectors $\nabla_{p,o}\tilde{}\, X$ is rather difficult. Besides the fact that one gets such a vector for each alternative hypothesis (= each observation), the vectors are depending on the coordinate definition (the S-system) too, as illustrated in the pictures.



Therefore the norm $\bar{\lambda}_{p,o}$ of the vector $\nabla_{p,\tilde{o}} X$ is introduced as a measure for the external reliability according to an alternative hypothesis Ha_p:

$$\bar{\lambda}_{p,o} = \nabla_{p,\tilde{o}} X^T \cdot Q_{xx}^{-1} \cdot \nabla_{p,\tilde{o}} X$$

In this way the external reliability for each alternative hypothesis is characterized by one number, which is independent of the coordinate definition.

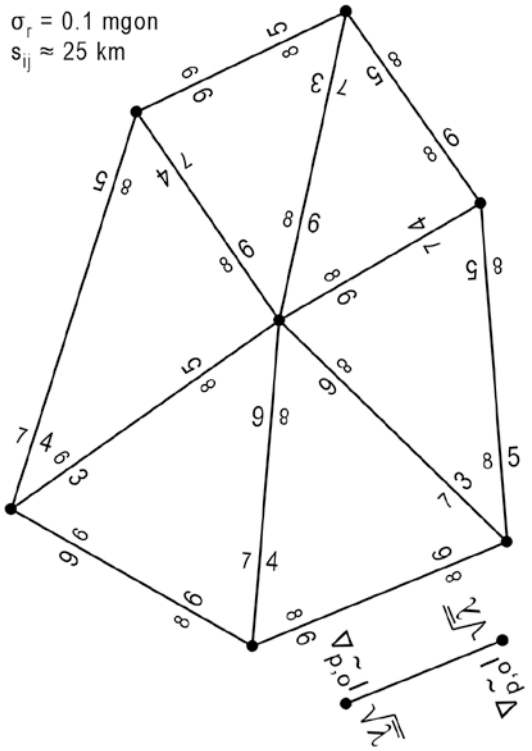
To have a criterion for the value $\bar{\lambda}_{p,o}$ the next relation may be helpful:

$$\nabla_{p,\tilde{o}} F \leq \sqrt{\bar{\lambda}_{p,o}} \cdot \sigma_F$$

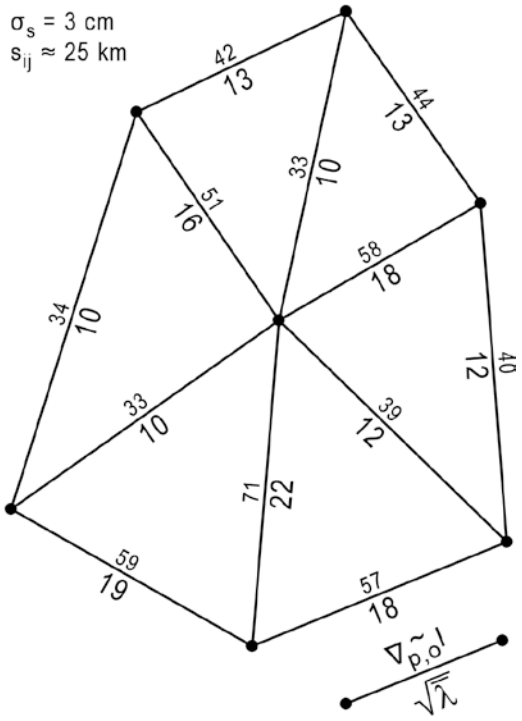
which means: the influence of a possible undetected error with size $\nabla_{p,\tilde{o}} l$ Ha_p on any function F of X is smaller than $\sqrt{\bar{\lambda}_{p,o}}$ times the standard deviation of that function. In other words: $\sqrt{\bar{\lambda}}$ is the ratio between reliability and precision. For instance, when the standard ellipses are multiplied by $\sqrt{\bar{\lambda}}$ one gets the areas to which the influence of possible non-detected errors on the coordinates is restricted (with power β).

For general purpose networks one would like to have a homogeneous reliability, that means, values of $\sqrt{\bar{\lambda}_{p,o}}$ as equal as possible. In that case the reliability of the whole network can be characterized by the

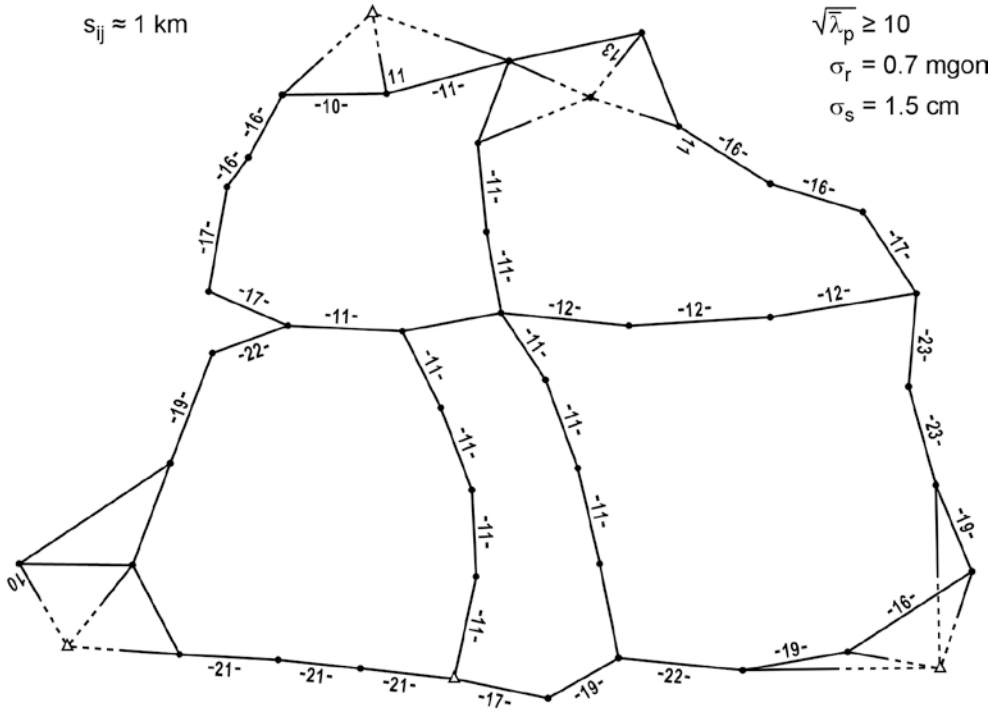
largest $\sqrt{\bar{\lambda}_{p,o}}$-value, denoted by $\sqrt{\bar{\lambda}_o}$. The following examples give an impression of the reliability of several networks, expressed by the $\sqrt{\bar{\lambda}_{p,o}}$-values resulting from the boundary values of the observations.



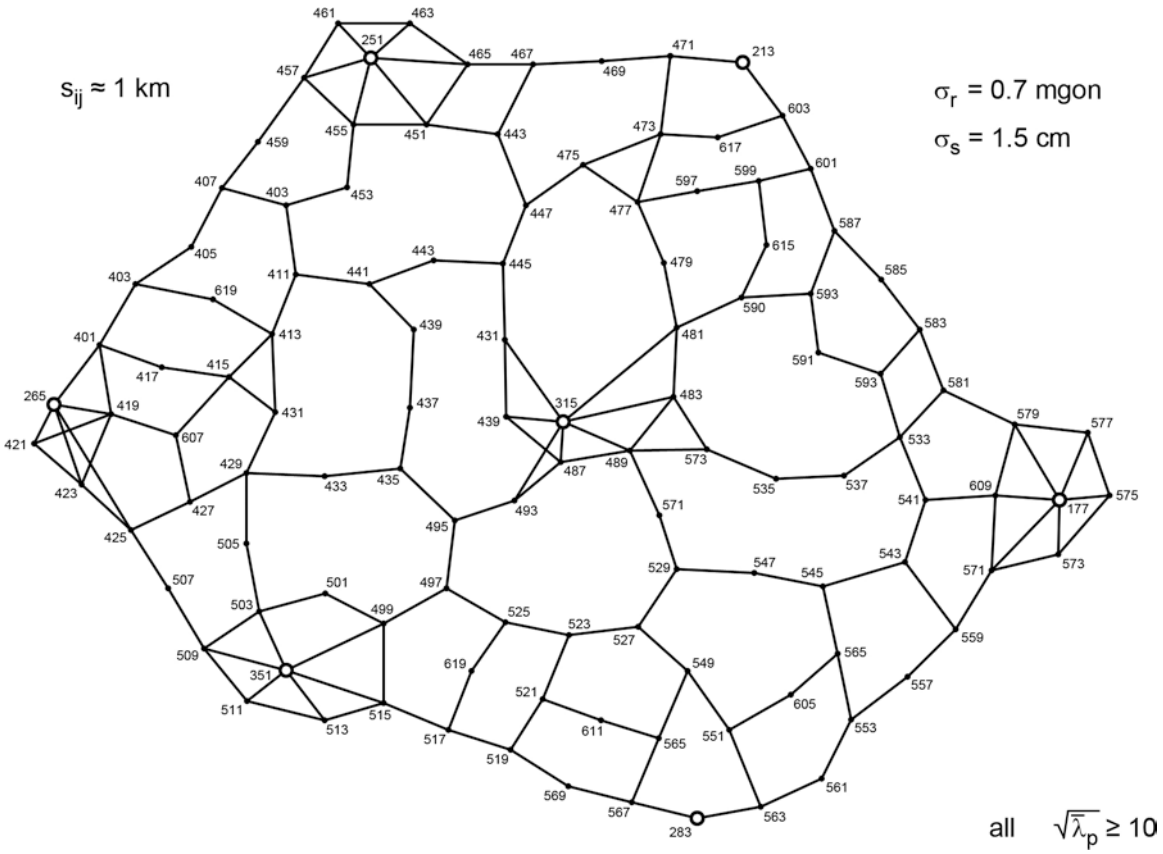
TRIANGULATION



TRILATERATION



TRAVERSE NETWORK



5. On the choice of the testing parameters

α gives the probability to reject the null-hypothesis, although it is valid. The consequence of such a wrong conclusion is that one needlessly tries to find a cause and fruitlessly spends money for instance by thinking what might be wrong and by remeasuring some observations. Since in addition, this may result in a decreasing trust in the testing procedure applied, α should be small.
But a smaller α, results in a larger value of λ, and consequently in larger boundary values; that means, larger errors, which have a probability $1-\beta$ to remain undetected. There seems to be evidence that about 1 in every 100 observations could be erroneous. Because of the distribution of the size of the errors less than 1 in every 1000 observations is expected still to contain an undetected error of the size of about the boundary value or smaller, after the testing procedure has been carried out. Although this indeed very rough estimate gives a rather small probability for undetected errors, they may cause enormous costs afterwards when they are detected later.

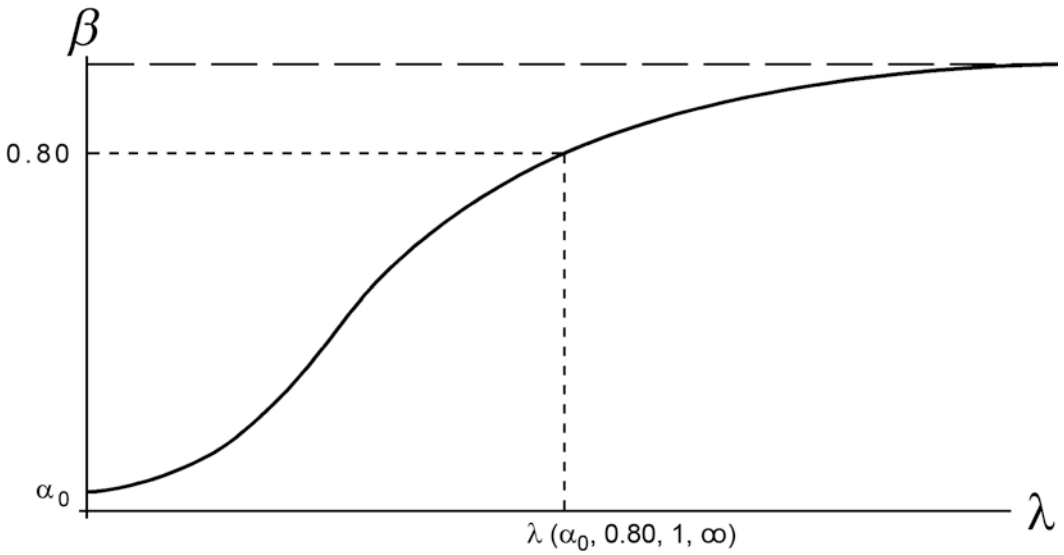
Beside by increasing α, the boundary values can be lowered by strengthening the network, by which of course the design and measurements of the network get more expensive, too.
So there are three influences on the costs of a network in optimizing the reliability:
- because of wrongly rejecting the null-hypothesis
- because of undetected errors which may cause trouble in the future
- because of strengthening the network
for which an optimum has to be found.
Theoretically such a problem can be solved, but application in practice is not easy, mainly due to the very difficult estimation of the costs of non-detected errors.
In The Netherlands, for 15 years, a value $\alpha_0 = 0.001$ is applied in the data-snooping. From this experience this choice seems to be workable. It means one false rejection in every 1000 observations tested. Since in practice, as mentioned before, for different reasons about 10 in every 1000 observations seems to be wrong, the trust in this testing procedure is very large.
In special purpose networks the optimization of the reliability is often easier. The costs of for instance a bridge not fitting the piers are easier to calculate and one only has to concentrate on the reliability of one quantity, in this example the distance between the two piers.

The choice of the power β_0 is of less interest, it is only fixing a power-level to judge the boundary values. About always the value $\beta_0 = 0.80$ is chosen, following statistical convention. A larger power leads to larger boundary values according the known relation

$$\lambda = \lambda(\alpha, \beta, 1, \infty)$$

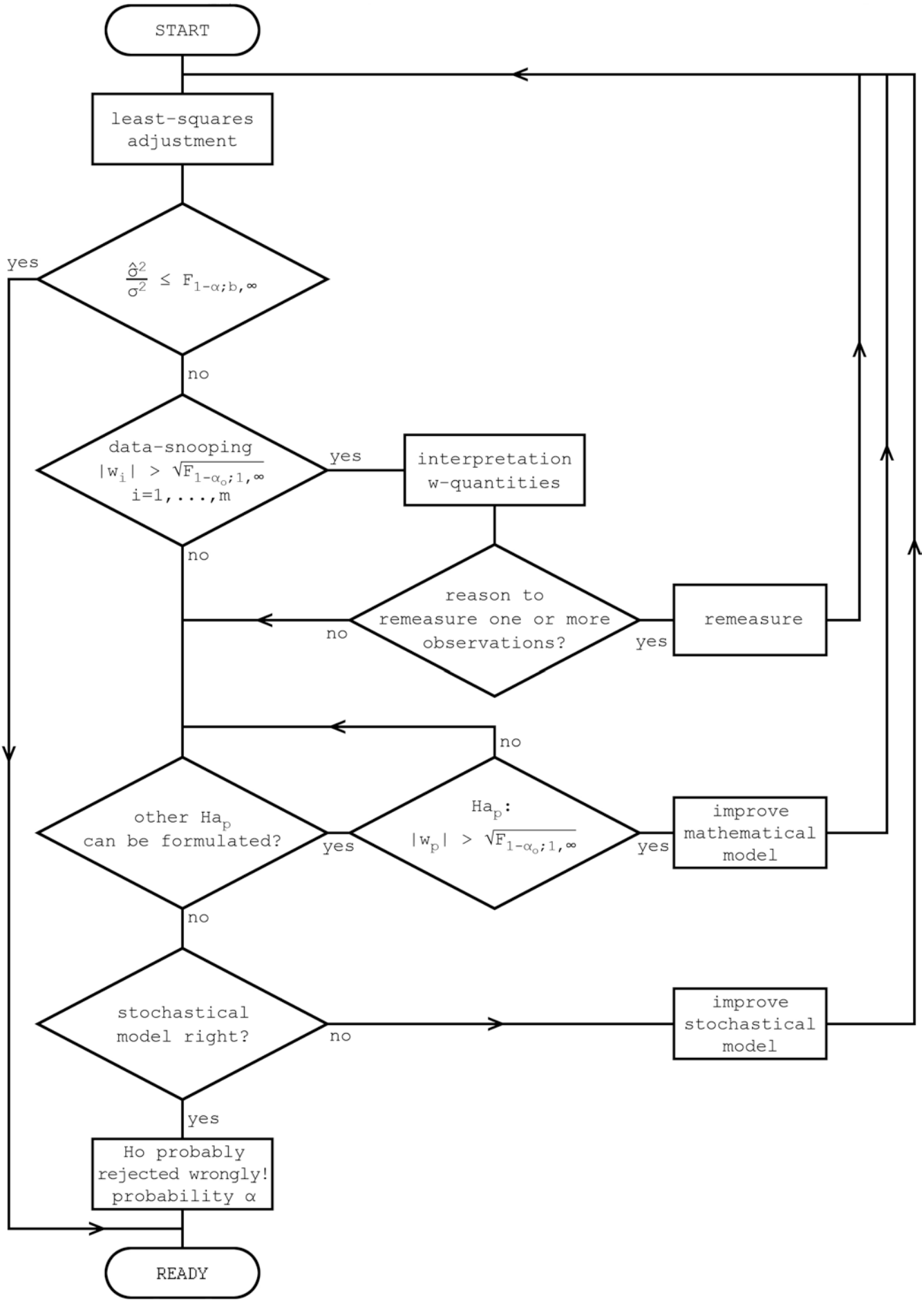
but in fact the reliability is not really different.

Actually, the curve of the relation between β and λ in the picture below is fixed by one point at $\beta = 0.80$ and boundary values for other values of the power β can easily be computed.



6. Practical application of the B-method of testing

The practical application of the testing procedure is described on the flow-chart on page 223.
If the $\hat{\sigma}^2$-test and one or more w_i-tests from the data-snooping are rejected the w-quantities have to be interpreted to decide which observations have to be remeasured. In almost all cases the largest w_i-value corresponds to the observation with the largest errors. However, because of the correlation between the w-quantities, interpretation can be difficult in case there are more errors and in case of very large errors. Then location of errors is easier in combination with testing the misclosures of the condition equations: Observations in accepted condition equations are considered to be right. After remeasuring the testing procedure should be repeated to be sure Ho is accepted now.
In case no measuring errors in the observations can be found the question arises whether other, special alternative hypotheses should be tested: influences of refraction, a wrong or too much simplified mathematical model, systematical errors in the observations and so on. For instance, when a part of the second order triangulation of The Netherlands was adjusted, all w_i-quantities were accepted, but the $\hat{\sigma}^2$-test was definitely rejected, because of not talking into account the curvature of the earth in the mathematical model.
If not, the mathematical model is probably right and the stochastical model is still to be checked: the level of the standard deviations of the several kinds of observations, neglected correlations and so on. Variance-analyses provide several tests to check the validity of the used stochastical model.
If the stochastical model is considered to be right too, the only conclusion left reads: Ho is probably wrongly rejected.


START
least-squares adjustment
$\frac{\hat{\sigma}^2}{\sigma^2} \leq F_{1-\alpha;b,\infty}$
yes
no
data-snooping
$|w_i| > \sqrt{F_{1-\alpha_o;1,\infty}}$
i=1,...,m
yes
interpretation w-quantities
no
reason to remeasure one or more observations?
no
yes
remeasure
no
other Ha_p can be formulated?
yes
Ha_p:
$|w_p| > \sqrt{F_{1-\alpha_o;1,\infty}}$
yes
improve mathematical model
no
stochastical model right?
no
improve stochastical model
yes
Ho probably rejected wrongly! probability α
READY

REFERENCES

BAARDA, W., 1967: Statistical Concepts in Geodesy. Math. Geod. Comm. Publ. on Geodesy, new series, Vol. 2, No. 4, Delft.

BAARDA, W., 1968: A Testing Procedure for Use in Geodetic Networks. Neth. Geod. Comm. Publ. on Geodesy, new series, Vol. 2, No. 5, Delft.

BAARDA, W., 1980: Measures for the Accuracy of Geodetic Networks. Int. Symposium on Optimization of Design and Computation of Control Networks, Sopron, Hungary.

DE HEUS, H.M., 1980: De huidige stand van de Delftse Toetsingsmethode. Geodesia Nederlands geodetisch tijdschrift 22, No. 9, pp. 307-314.

QUALITY RELATED PROBLEMS OF DENSIFICATION NETWORKS

Henk DE HEUS
Delft, Netherlands

ABSTRACT

The paper describes how the precision of networks can be judged by using a criterion matrix. In addition, a measure for the reliability of a network is introduced.
The practical application of these theories is shown by some examples. Some practical rules are given for the planning of networks of acceptable quality. The last part of the paper deals with the connection of free networks to fixed points with known coordinates, and the implications for precision and reliability.

1. Introduction

By "quality of a network" is indicated how well the coordinates of the network are estimated, in other words: the possible size of the differences in the coordinates in case the network would be measured and calculated a second time.
There are two reasons for different results:
First, the observations, from which the coordinates are computed, are ***stochastic*** quantities. Therefore, the coordinates are stochastic quantities, too. Their (normal) probability distribution, as described by their variance-covariance matrix indicates the possible variations around the estimated values of the coordinates (confidence regions). The possible variation because of the stochastic properties of the coordinates is called ***precision*** of the network.
The second reason for possible different results is a wrong assumption of the null-hypothesis Ho. The assumptions about the stochastical and mathematical model of the observations in Ho have to be realistic, otherwise one gets wrong results or coordinates.
To check Ho a testing procedure has to be carried out. Then the power of the tests applied can be expressed by means of marginally detectable errors and their influence on the coordinates. The possible differences because of measuring errors and other wrong assumptions in Ho are called ***reliability*** of the network. So "quality" implies both the precision and the reliability of the network.

The reliability has already been treated in my paper: "Data-snooping in Control Networks". So in this paper we will mainly restrict ourself to the precision of a network.

2. The precision of coordinates

The precision of the coordinates depends on the precision of the observations and the way the coordinates are derived from those observations. If the variance-covariance matrix of the observations is denoted by:

$$Q_{ll} = P^{-1}$$

and the coordinates X are calculated from the observations as:

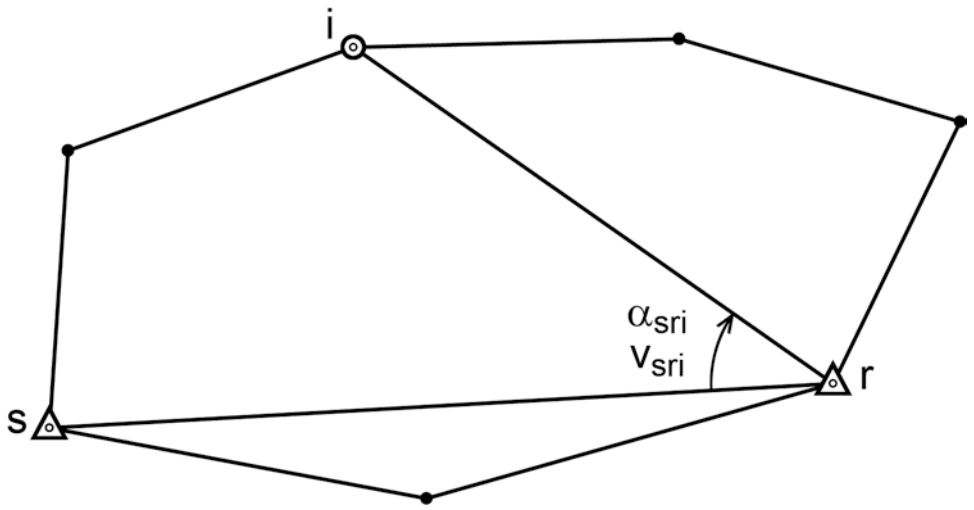
$$X = \Lambda l$$

the corresponding variance-covariance matrix of the coordinates follows from the law of propagation of covariances:

$$Q_{XX} = \Lambda Q_{ll} \Lambda^T$$

In case the coordinates result from a least-squares adjustment $\left[\Lambda = \left(A^T P A\right)^{-1} A^T P\right]$ the estimators X are the Best Linear Unbiased Estimators (BLUE), which means that the estimators X have the best possible precision. However, for any Λ, expressing a linear unbiased estimator X of the coordinates, the precision of the coordinates can be calculated from the above formula, although such estimators give approximate solutions for the coordinates.

Since location, orientation (and scale) of the coordinate system cannot be estimated from the observations, we can only calculate coordinates in a so-called S-system, which defines the above properties of the coordinate system. The coordinate system can be tied up to the network for instance by assuming the coordinates of two points non-stochastic. Then location, orientation and scale of the coordinate system are fixed by the approximate values of the coordinates of those two points P_r and P_s. The stochastic properties of the coordinates of a point P_i are in this case completely determined by the precision of the angle α_{sri} and the length-ratio v_{sri} which can be derived from the observations (picture on the next page).
In case coordinates are calculated as the minimum norm solution of the least-squares adjustment implicitly as S-system is chosen as well. The minimum norm solution can be interpreted as a fixation of the coordinate



system by means of a Helmert transformation of the (adjusted) network to the approximated values of the coordinates of ***all*** points of the network. The relation between this type of coordinate-quantities and the estimable angles and length-ratios of the network is now more complicated. On the other hand the precision of the coordinates seems to be better (on the average the standard ellipses are smaller). However, in any S-system angles and length-ratios derived from the coordinates have the same precision, so ***in fact the precision of the network is the same in any S-system.***
Because of their dependence on the (arbitrary) S-system standard ellipses are very hard to be used for examining the precision of a network. Anyway, in no S-system it is justified to use a fixed criterion value for the standard ellipses.

3. Criterion matrix

A better way to examine the precision of a network is by comparing the variance-covariance matrix of the coordinates with a criterion matrix.
In case of general purpose networks such a criterion matrix should meet the following demands:

a. point standard ellipses and relative standard ellipses are circles.
b. the relative precision between two points is equal in all directions, or the covariance function d_{ij}^2 only depends on the distance between the points: $d_{ij}^2 = f(l_{ij})$.
c. the criterion matrix is defined in the same S-system as the variance-covariance matrix of the coordinates is. Otherwise different functions of the angles and length-ratios would be compared.

In a kind of "absolute" system the following criterion matrix meets the demands a and b.

	x_i	x_j	x_r	x_s	y_i	y_j	y_r	y_s
x_i	d^2	$d^2-d_{ij}^2$	$d^2-d_{ir}^2$	$d^2-d_{is}^2$				
x_j	$d^2-d_{ji}^2$	d^2	$d^2-d_{jr}^2$	$d^2-d_{js}^2$			0	
x_r	$d^2-d_{ri}^2$	$d^2-d_{rj}^2$	d^2	$d^2-d_{rs}^2$				
x_s	$d^2-d_{si}^2$	$d^2-d_{sj}^2$	$d^2-d_{sr}^2$	d^2				
y_i								
y_j			0		same submatrix as left upper part			
y_r								
y_s								

To meet demand c, an S-transformation has to be carried out. d^2 is only a nuisance-parameter: by the S-transformation d^2 disappears from the formulas.
For a much more elegant derivation is referred to W. Baarda: "S-Trans-formations and Criterion Matrices".

As Baarda has shown in his Sopron paper the above matrix corresponds to a special case of the Taylor Karmann structure, introduced by E. Grafarend, namely the so-called chaotic structure. This restriction is necessary to get circular standard ellipses.
In The Netherlands the covariance function is chosen as:

$$d_{ij}^2 = c \cdot l_{ij}$$

K. Borre and P. Meissl have shown that d_{ij}^2 should be a logarithmic function of l_{ij}, but in practical applications the linear function has proven to be an acceptable simplification.
The criterion matrix can be used in two ways:

1. In case from the aim of the network a criterion value for c can be deducted the precision of the network can be compared with the precision demanded by the criterion matrix.
2. In case no criterion value for c is given one can determine the value c for which the criterion matrix gives a good representation of the precision of the network. Then, in the future, the real variance-covariance matrix of the coordinates can be replaced by this artificial matrix, for which only one parameter c has to be stored.

To give an idea of the meaning of a particular value c the following approximation might be helpful: the relative precision of two neighbouring points is roughly given (in cm) by $\sqrt{2 \cdot c \cdot l_{ij}\,(km)}$.

4. Examination of the precision

A first possibility to examine the precision of a network could be to compare the standard ellipses with the criterion circles computed from the criterion matrix. Although one can get a fair impression of the precision in this way, such a comparison is still depending on the S-system used, as is shown clearly in the two pictures on the next page. In the first example all standard ellipses are smaller than the criterion circles, but in the second some standard ellipses are too large.
The cause of these different results in examining the precision in this ways is, that in different S-systems different functions of the co-ordinates are examined. As explained in section 2 the point standard ellipse of for instance point 11 in the first example expressed the precision of $\alpha_{21,17,11}$ and $v_{21,17,11}$ of which the precision seems to be acceptable, but the same standard ellipse in the second example gives the precision of $\alpha_{3,21,11}$ and $v_{3,21,11}$ and these functions of the coordinates do not seem to meet the demands.
Because by examining standard ellipses only a limited number of functions of the coordinates are considered, only a rough impression of the precision of the network can be gathered.

To make any arbitrary function of the coordinates meet the criterion the following inequality should be valid:

$$\Lambda Q \Lambda^T \leq \Lambda H \Lambda^T$$

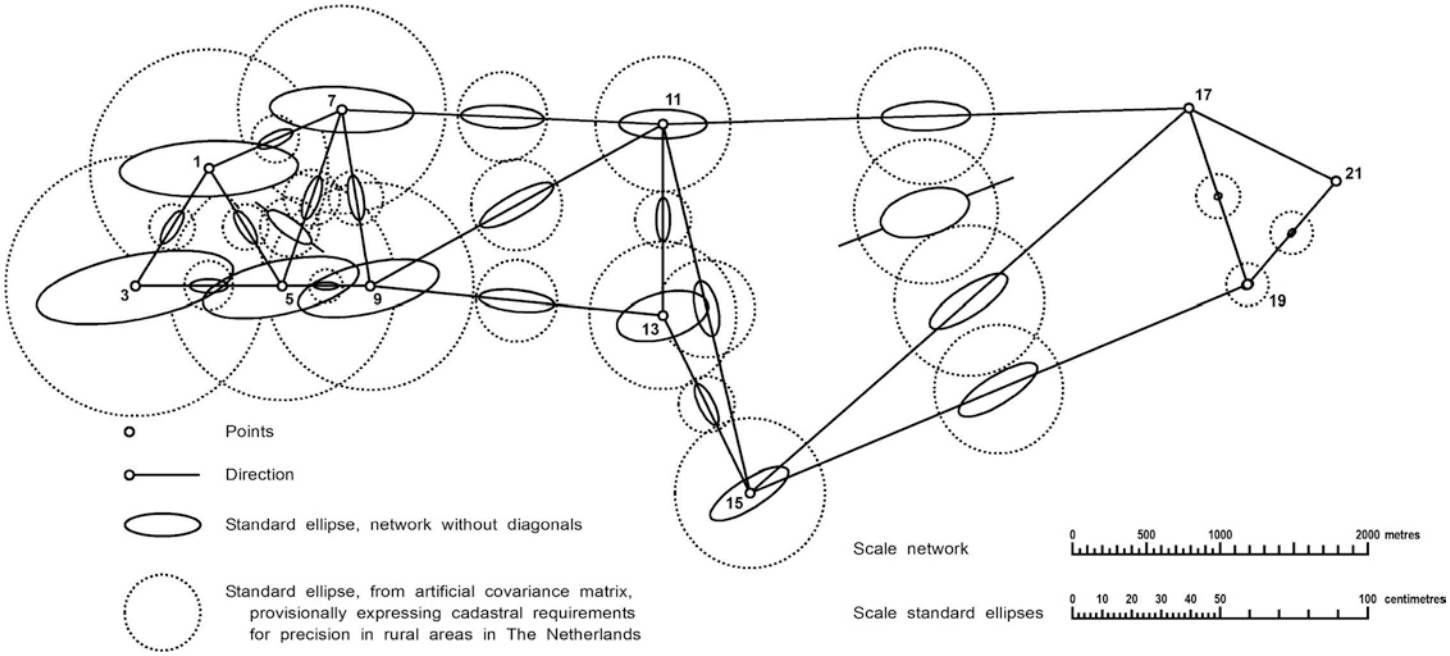
Q: variance-covariance matrix of the coordinates
H: artificial criterion matrix
Λ: vector of coefficients of an arbitrary linear(ized) function of the coordinates
This inequality can be written as

$$\Lambda (Q-H) \Lambda^T \leq 0$$

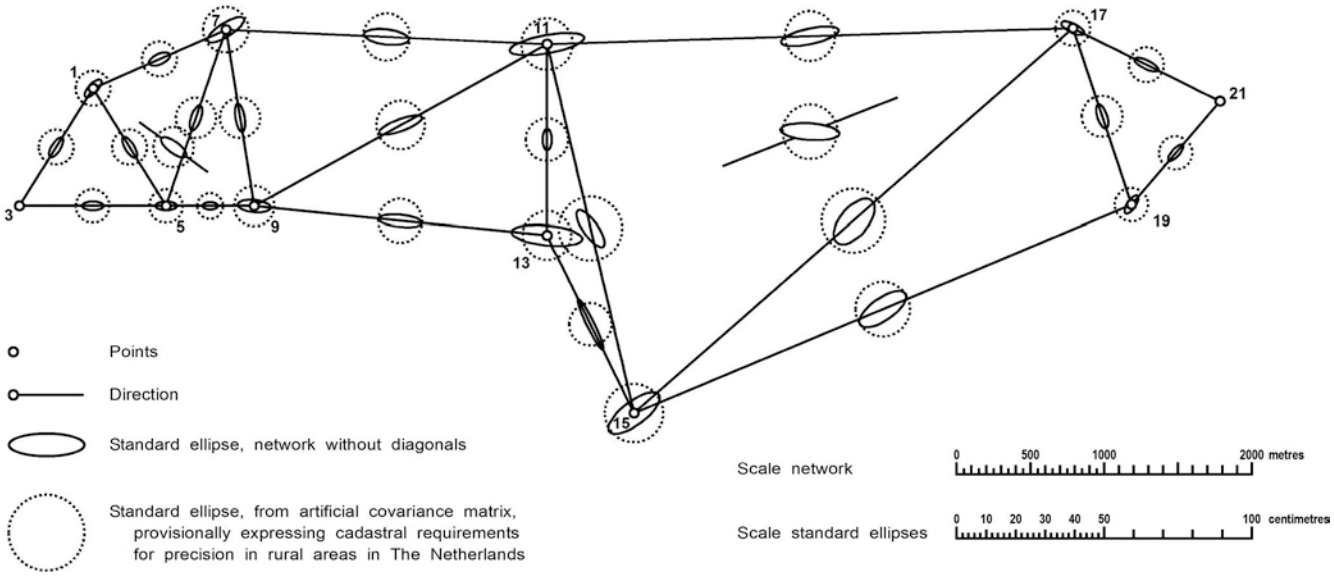
This means that the matrix (Q-H) must be negative semi-definite, which is true if all eigenvalues of the general eigenvalue problem

$$|Q - \lambda H| = 0$$

S - system
17,21



S - system
3,21



are less than or equal to 1, so the largest eigenvalue

$$\lambda_{max} \leq 1$$

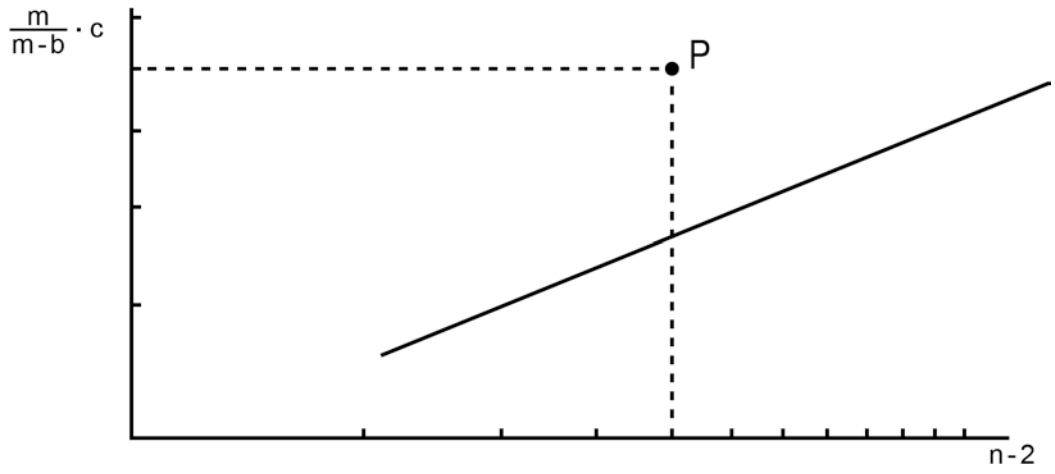
In case no criterion is given we can use the same eigenvalue problem to determine the matrix H, which is suitable to indicate and replace the precision given by the variance-covariance matrix Q. We just calculate λ_{max} of the eigenvalue problem with an arbitrary value c' in H and then we should choose the value $c = \lambda_{max} \cdot c'$ in the covariance function to get a largest eigenvalue equal to 1.

From many experiments in schematic triangulation and polygon networks the following functional relationship for not too oblong networks is deducted:

$$c \cong k\sigma^2 \frac{m-b}{m} (n-2)^q$$

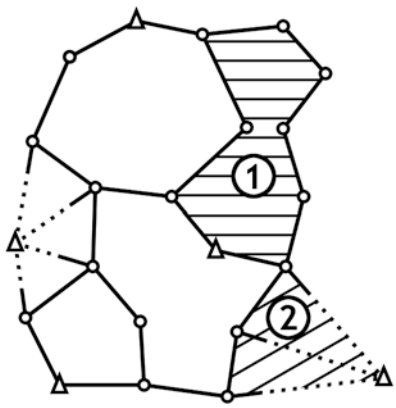
m: number of observations
b: number of condition equations
n: number of points in the network
σ^2: a priori covariance factor
k,q: parameters, depending on the type of network.

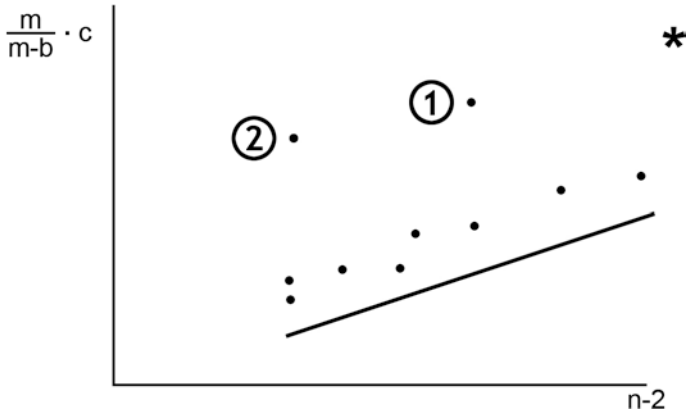
On logarithmic coordinate paper this relationship is represented by a straight line, as shown in the figure below.



Now the precision of practical networks can be computed with the precision of a schematic network. The calculated value c results in a point P, which should not be situated too much above the "ideal" line.
In case a value c is considered to be too large, the design of the network should be improved.

Then an indication of the weakest part of the network is given by the largest component of the eigen vector belonging to the largest eigenvalue (c). But, since the components of the eigen vector are depending on the chosen S-system, practical application is rather difficult.
A better way to locate parts of the network to be improved, is to calculate the value c of partial networks from the eigenvalue-problem of the corresponding part of the variance-covariance matrix. In this way the precision of parts of the network can be examined, provided that the S-system is defined only by points in that part of the network. If necessary, a S-transformation of the variance-covariance matrix has to be carried out.
The value c of each examined partial network is represented by a point in the figure as shown in the following example.



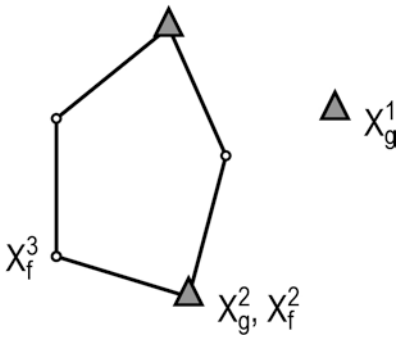


The plotted points should fluctuate around a line parallel to the ideal line and points situated too far to the upper side correspond to weak partial networks. In the example 1 indicates the weak relative precision between the two neighbouring, not connected points in the polygon concerned and 2 a weak intersection of the given point. When these two parts of the network are improved the value c of the whole network (*) will be smaller and all points will fairly well lie on a straight line.
Then the precision of the network can be considered to be homogeneous and the level of the precision is given by the position of the line.

5. Quality of densification networks

In view of a sharp detection of measuring errors by the testing procedure it is preferable to carry out the adjustment of densification networks in ***two*** steps.
Then possible errors in the observations can be well detected in the free network (the first step), because the mathematical and stochastical model of the free network are well defined; in case no phase-adjustment would be carried out, this data-snooping could be troubled because of possible complicated distortions of the given points and the generally only roughly known variance-covariance matrix of the given coordinates.
In the second phase the free network is linked up to the given points and only the coordinates of the given points are still to be tested.
The quality of the free network can be examined by the measure $\sqrt{\bar{\lambda}}$ for the reliability and c for the precision as explained in the previous section and in my paper "Data-snooping in Control Networks".
When both the coordinates of the free network and the given coordinates are defined in the same S-system the second phase of the adjustment can be carried out, if the variance-covariance matrix of the given coordinates is available or replaced by an artificial matrix as discussed in section 3.
Denoting:

X_g^1: given coordinates outside the network
X_g^2: given coordinates inside the network
X_f^2: coordinates of the given points in the free network
X_f^3: coordinates of the new points in the free network



and their corresponding covariance matrix as

	X_g^1	X_g^2	X_f^2	X_f^3
X_g^1	Q_g^{11}	Q_g^{12}		
X_g^2	Q_g^{21}	Q_g^{22}		
X_f^2			Q_f^{22}	Q_f^{23}
X_f^3			Q_f^{32}	Q_f^{33}

Q_g: artificial matrix
Q_f: variance-covariance matrix of the free network

Then the least-squares coordinates are calculated as:

$$\begin{pmatrix} \hat{X}^1 \\ \hat{X}^2 \\ \hat{X}^2 \\ \hat{X}^3 \end{pmatrix} = \begin{pmatrix} X_g^1 \\ X_g^2 \\ X_f^2 \\ X_f^3 \end{pmatrix} - \begin{pmatrix} Q_g^{12} \\ Q_g^{22} \\ Q_f^{22} \\ Q_f^{32} \end{pmatrix} \left(Q_g^{22} + Q_f^{22}\right)^{-1} \left(X_g^2 - X_f^2\right)$$

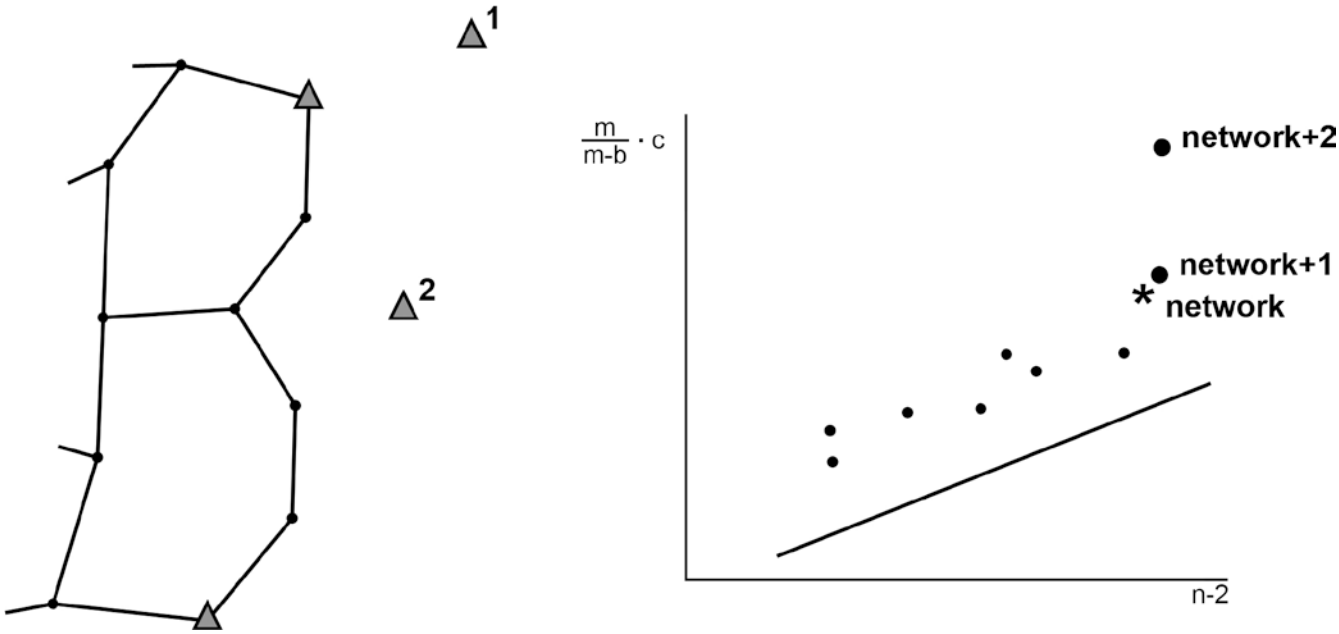
But, this least-squares adjustment to link up the free network to the given points causes corrections, not only to the given coordinates X_g^2 used, but to the correlated known points outside the network X_g^1 as well, which is very unpractical. So, there is a great practical advantage in keeping the coordinates of the given points fixed.
Then the effect of not correcting the given coordinates should be evenly distributed over the network by a "least-squares interpolation" of the free network between the given points. Coordinates according to such an interpolation can be computed using the algorithm of the least-squares adjustment, assuming the given coordinates ***not*** stochastical $\left(Q_g = Q\right)$.
This results in the following formula to calculate the coordinates:

$$\hat{X}^1 = X_g^1$$

$$\hat{X}^2 = X_g^2$$

$$\hat{X}^3 = X_f^3 + Q_f^{32}\, Q_f^{22\,-1}\left(X_g^2 - X_f^2\right)$$

By applying the law of propagation of covariances to this formula, using the correct covariance matrix Q_g of X_g the variance-covariance matrix $Q_{\hat{x}\hat{x}}$ of the estimators $\hat{X}$ of the interpolated coordinates is derived.
Then the precision of the network can be examined in exactly the same way as indicated in the previous section. Beside the precision of the partial networks the relative precision between the points X^3 of the network and given points X^1 outside the network has to be examined. The latter can be done by judging the difference between the value c of the covariance matrix of the network and the value c of the covariance matrix of the network including a given point outside the network as illustrated in the following example (picture on the next page).
Including point 1 hardly influences the c-value, but including point 2 clearly does and disturbs the homogeneity of the precision. In this example the designed network has to be improved by including point 2 in the network (for instance by intersection of point 2) or by enlarging the distance between point 2 and the network.



The reliability of the network after linking up the free network to the known points can be described again by the measure $\sqrt{\bar{\lambda}}$. The influence $\bar{\nabla}\hat{X}$ of marginally detectable errors on the coordinates $\hat{X}$ can be derived from the formula to calculate $\hat{X}$ on the previous page and further $\bar{\lambda}$ is calculated according to its definition:

$$\bar{\lambda}_p = \bar{\nabla}_p \hat{X}^T Q_{\hat{X}\hat{X}}^{-1} \bar{\nabla}_p \hat{X}$$

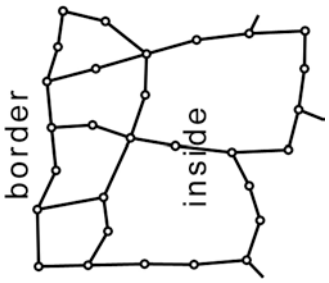
Beside the $\sqrt{\bar{\lambda}}$ according to the alternative hypotheses of the free network concerning errors in the observations, the values of $\sqrt{\bar{\lambda}}$ according to alternative hypotheses concerning distortion and errors in the coordinates of the given points now have to be examined.
For practical densification networks it has become obvious that the values $\sqrt{\bar{\lambda}}$ according to all alternative hypotheses always meet the demands if the reliability and the precision of the free network are acceptable.

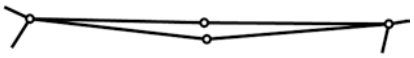
6. Rules for the design of densification networks

With respect to practical application all former theories concerning reliability and precision result in a few rules of thumb for the design of densification networks. It appears that meeting the criterion $\sqrt{\bar{\lambda}} \leq 10$ puts the strongest constraints on the design, namely:

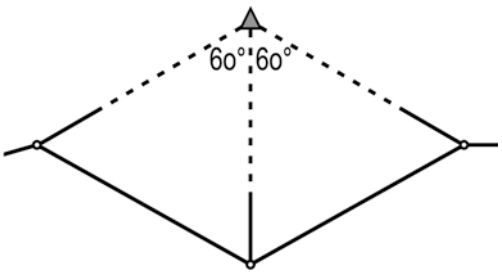
- closed polygons at the border of the network should not contain more

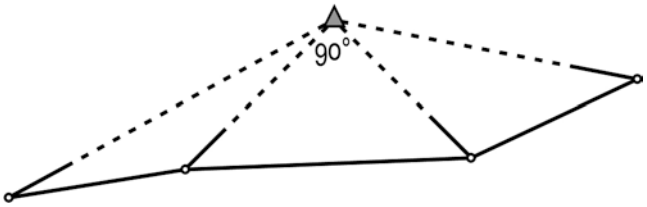
than 6 or 7 points; polygons inside the network up to about 12 to 14. In case the terrain compels to larger closed polygons a very handsome way to reach an acceptable reliability is by measuring directions and distances to an auxiliary point, which then should be only temporarily monumented.





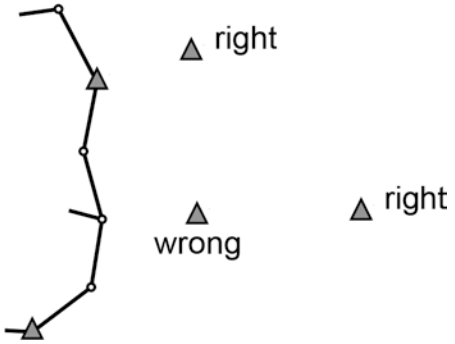
- every direction in an intersection of a given point is satisfactorily checked by another direction if the intersection angle is at most 60 degrees. A satisfactory precision is realized if one intersection angle is at least 60 degrees (optimal is 90°). The distances to the intersectional point should be at least 0.7 times the distances in the network. Remark: in case of only three intersectioning directions a possible error in one of them is indeed detectable, but cannot be located. Therefore at least four intersectioning directions are recommended.

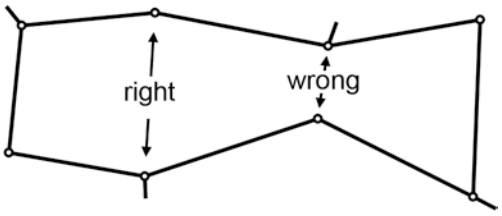




Two other rules, although not satisfactorily specified yet, concern

- the relative precision between a point of the network and a given point outside the network. Decisive is the position of the latter given point relative to the point of the network and to given points included in the network.
- the relative precision between points, not connected by measurements. Such points should not be too close.





REFERENCES

BAARDA, W., 1973: S-Transformations and Criterion Matrices. Neth. Geod. Comm. Publ. on Geodesy, new series, Vol. 5, No. 1, Delft

BAARDA, W., 1980: Measures for the Accuracy of Geodetic Networks. Int. Symposium on Optimization of Design and Computation of Control Networks, Sopron, Hungary

BORRE, K. and P. MEISSL, 1974: Strength Analysis of Levelling-type Networks. Mittlg. Geod. Inst. TH Graz, No. 15

GRAFAREND, E.W., 1971: Isotropietests von Lotabweichungsverteilungen in Westdeutschland. Z. Geophysik, 37, pp. 719-733

HORIZONTAL AND VERTICAL SURVEY CONTROL BY THREE-DIMENSIONAL TRIANGULATION

Ludvík HRADILEK
Praha, Czechoslovakia

ABSTRACT

In high mountains, the problem of establishing both the horizontal and vertical survey controls can be solved by a three-dimensional terrestrial triangulation

1. Introduction

The traditional procedures for establishing of horizontal and vertical survey controls fail in rough mountain areas: the spirit leveling loses its high accuracy, sometimes cannot be performed at all, and, consequently, the measured distances cannot be exactly reduced to the ellipsoid of reference. These problems can be solved by a method of three-dimensional triangulation which may be developed in the form of an exact combination of traditional trigonometric leveling a two-dimensional triangulation. In a three-dimensional adjustment we evaluate all the observed data, i.e. horizontal and vertical angles, measured distances, deflections of the vertical and spirit leveling differences in one step, without any transformations and without reduction of distances to the ellipsoid of reference. The accuracy obtained promotes the method to establish horizontal and vertical survey controls in high mountain areas.

2. Conditions for obtaining high accuracy

2.1 Solution for terrestrial refraction and evaluation of precisely measured distances, inclined as much as possible, are basic conditions for obtaining high accuracy especially in elevations. One may expect the effect of refraction to be practically negligible only with some of the very accurate rangefinders. In measuring angles with theodolites the situation is more serious; none of the methods determining/eliminating refraction in practical geodesy is perfect; most of them, however, are an aid to some extent. Trying to keep down the expenses for the field work we developed a method determining refraction from measured geodetic data, and especially from vertical angles in the adjustment of a three-dimensional net. The observational procedure for vertical angles must, of course,

be adopted to the refraction model used in the adjustment. When determining the main value of refraction common to all the lines of sight at a station of observation, it is necessary for all the vertical angles at the station to be measured quickly one after the other and to be processed as one observation unit. The changes of refraction deduced from changes of elevation angles, as determined by repeated measurements, play an important role in checking the refraction model designed for the adjustment, or for deducing a new improved model (HRADILEK, 1980).

The refraction problems are reduced appreciably if the observation stations can be situated on sharp mountain peaks or triangulation towers at least 15 - 20 m above the ground. Accurately measured and significantly inclined distances (> 15°), as well as spirit-leveled height differences, provided they were measured between some points of the net, support the elimination of systematic refraction errors and contribute substantially to the accuracy of elevation coordinates.

2.2 Deflections of the vertical should never be neglected when processing the geodetic data obtained by measuring in mountain areas, except when adjusting only distances and spatial angles. There is a lot of methods determining deflections of the vertical by interpolation and/or prediction; the optimum procedure depends on the quantity and quality of the astronomical and gravimetric observations. If neither gravimetric observations nor topographic maps are available, it is sufficient to determine the deflections of the vertical by astronomical measurements at about one quarter of the stations and to calculate the remaining deflections of the vertical it is necessary to increase the number of vertical angles measured to five to eight angles. Their lines of sight should differ substantially both in the lengths and in their azimuths.

All the problems with refraction and deflections of the vertical can be avoided by drawing all information necessary for establishing of a three-dimensional network from distance measurements of a top precision.

3. Mathematical and stochastical models

With regard to the values of parameters being determined, neither the method of adjustment, nor the choice of the coordinate system should play an important role in the three-dimensional adjustment. However, the data processing will be much simpler and less expensive if we choose the method of adjustment and the coordinate system according to the purpose and the extend of the net.

3.1 The three-dimensional geodetic coordinate system φ, λ, h is suitable for treating terrestrial three-dimensional nets, the main purpose of which is to establish horizontal and vertical geodetic controls in high mountain areas. Using elevation system designed by MOLODENSKI (1948) it is easy to transform the results of spirit leveling into the ellipsoid heights h and vice versa. Therefore, the mathematical models for the adjustment of three-dimensional nets may be deduced in a form of an exact combination of traditional models for adjusting triangulated heights and two-dimensional horizontal networks. This approach confirms a close relation between the three-dimensional and traditional procedures and can be used to an advantage in modernizing older trigonometric nets by three-dimensional triangulation.

3.2 The local Cartesian coordinate system (also referred to as topocentric) will prove particularly useful in treating single-purpose nets of smaller extent. If nets with a maximum diameter up to 40 km are involved, the transformation of the observed data into the local Cartesian system, as well as the inverse transformation of the adjusted data can be expressed by simple formulas. The transformation of the variance-covariance matrix is practically unnecessary in nets of such an extent.

3.3 Since it is generally simpler to set up observation equation than condition equations, the adjustment by variation of parameters is frequently preferred to the adjustment of conditioned observations. Even though the latter method also has some advantages such as more thorough checks and, sometimes, a considerably smaller number of unknowns. The observation equations for a three-dimensional adjustment deduced as an exact combination of traditional methods were given by HRADILEK (1980). The condition equations are discussed in the section 3.4.

A large part of the condition equations for a three-dimensional adjustment can be adopted from the two-dimensional adjustment of the horizontal networks, at least as far as the form of the condition equations is concerned. When compiling the condition equations it is useful to transform the observed horizontal angles α_{ijk} and corresponding zenith distances z_{ij}, z_{ik} into the spatial angles ω_{ijk} according to the formula

$$\cos \omega_{ijk} = \cos \alpha_{ijk} \sin z_{ij} \sin z_{ik} + \cos z_{ij} \cos z_{ik} \ . \qquad (1)$$

This transformation, together with the transformation of corresponding variances, is a disadvantage of the conditioned adjustment in comparison with the parametric adjustment. An advantage of the transformation (1) is

elimination of unknown deflections of the vertical.

3.4 To render the procedure of establishing the various variants of condition equations sufficiently lucid, we shall demonstrate it on a tetrahedron, which is a three-dimensional analogy of the equilateral triangle in the two-dimensional net. Assuming we have observed six distances, twelve horizontal directions with four unknown station corrections and twelve zenith distances with eight orientation parameters, i.e. deflections of the vertical, we have twelve redundant observations in the tetrahedron $P_1P_2P_3P_4$. By the transformation (1) of the horizontal directions and zenith distances we obtain twelve spatial angles without any unknown orientation parameters so that the number of degrees of freedom remains twelve. The twelve condition equations may be deduced according to any twelve independent geometric conditions, as given e.g. in the following examples which again illustrate a close similarity of our three-dimensional adjustment with traditional methods.

In the first Example we introduce:

4 condition equations for the sum of spatial angles in each planar triangle forming the wall of the tetrahedron; e.g. for triangle $P_1P_4P_3$ we write

$$\omega_{143} + \omega_{314} + \omega_{431} = 180^\circ \ ; \tag{2}$$

3 side conditions, which are formally identical with the side conditions in the adjustment of a two-dimensional triangulation net. For the vertex P_1 as a pole and for the basic triangle $P_2P_3P_4$ we write the side condition in the form

$$\frac{\sin\omega_{231}\sin\omega_{314}\sin\omega_{412}}{\sin\omega_{241}\sin\omega_{312}\sin\omega_{431}} = 1 \ ; \tag{3}$$

5 conditions in the form of Sine Law for planar triangle, e.g. for the triangle $P_1P_4P_3$:

$$l_{13} : l_{14} = \sin\omega_{431} : \sin\omega_{314} \ . \tag{4}$$

Each condition of the type (4) must contain at least one new distance that has not yet been used in setting up the condition equations.

In the second Example we have adjusted:

4 condition equations of the type (2),
8 Sine Law conditions (4), i.e. two conditions for each triangle

forming the wall of the tetrahedron.

In the third Example we use:

12 conditions in the form of Cosine Law, i.e. one condition for each spatial angle, e.g. for angle ω_{143} in the form

$$l_{34}^2 = l_{13}^2 + l_{14}^2 - 2\, l_{13}\, l_{14} \cos \omega_{143} \ . \qquad (5)$$

The results of the adjustments according to the examples No 1-3 were equivalent. However, there are many other possibilities. Better results were obtained when introducing into the adjustment only a part of the total number of spatial angles, namely those replacing measured horizontal angles or horizontal directions. For the spatial angles we could then write

8 mutually independent conditions of the type (2) - (5), completed with

12 condition equations for zenith distances containing eight orientation parameters as deduced by HRADILEK (1958).

3.5 The stochastic model for measured data should include both accidental and systematic sources of errors. Therefore, the formula expressing variances of vertical angles depends on the functional model evaluating refraction by adjustment (HRADILEK, 1980). The stochastic model for spatial angles is a variance-covariance matrix deduced using the transformation equations (1).

4. Practical results. Further Developments

The adjustments determining refraction at every observation station were applied in nine trigonometric and three-dimensional nets, and checked by independent methods as e.g. spirit leveling and astronomic measurements. The refraction coefficients were estimated with average standard errors of 0.011 and 0.012 in nets with average distances of 7 and 5 km, respectively. The first three-dimensional local net was established in the High Tatra Mts. in 1961/62. Its accuracy of about 1.2 cm in both horizontal and vertical coordinates was sufficient to yield, by repeated measurements in 1975, the tectonic crustal movements of the mountain peaks reaching 6 - 8 mm/year with a standard deviation of 1.9 mm/year. This precision corresponds to the equipment used in 1961/62. Investigations on models corresponding to the terrain conditions of the High Tatra have shown that a new equipment such

as multicolored rangefinders, can improve the above accuracy more than twice (HRADILEK and LOULOVÁ, 1982).

REFERENCES

HRADILEK, L., 1958: Bestimmung der relativen Lotabweichungen und des Refraktionskoeffizienten beim Ausgleich trigonometrisch gemessener Höhennetze. Studia geoph. et geod. 2 (1958), pp. 101-121

HRADILEK, L., 1980: Mathematische Modelle für die trigonometrische Höhenmessung und die dreidimensionale Triangulation. Mitt. Inst. Theor. Geodäsie Nr. 61, Bonn 1980

HRADILEK, L. and LOULOVÁ, A., 1982: Some problems of Three-Dimensional Triangulation (in Czech). Geodetický a kartografický obzor 28 (1982), pp. 85-88

MOLODENSKI, M.S., 1948: Vneshneye gravitatsionnoye pole i figura poverkhnosti Zemli. Izd. A.N. USSR Ser. geogr.-geof. XII, No 3, Moscow 1948

YEREMEYEV, V.F., 1951: Teoriya ortometritsheskikh, dinamitsheskikh i normalnykh vysot. Trudy CNIIGAiK, vyp. 42, Moscow 1951

CRITERION MATRICES FOR ESTIMABLE QUANTITIES

F. KRUMM
Stuttgart, Federal Republic of Germany

ABSTRACT

Within the concept of the analysis of geodetic networks often use is made of idealized variance-covariance matrices (criterion matrices). According to that, criterion matrices for estimable quantities can be developed as a starting point for the procedure of analyzing. The paper deals with such criterion matrices in two- and three-dimensional Euclidean space. A definition of estimable quantities is given.

Introduction

The point-coordinates of a geodetic network are determined by geodetic measurements and parameters, which define a coordinate system (datum parameters). Following the formulation of the design of the network and the performance of the measurements, the question of the precision of the network always arises. Usually the 2 x 2 subblocks in a two-dimensional respectively the 3 x 3 subblocks of the variance-covariance matrix of the coordinates in a three-dimensional network, act as a gauge for the point precision expressed in standard errors. Scalar functions of the above mentioned matrix, e.g. its trace or determinant, serve as a criterion for the precision of the complete network. But sometimes it's not desirable to express all informations of precision through only one number; it would be more beneficial to compare the complete variance-covariance with a given ideal one. In that sense an ideal(ized) variance-covariance matrix - criterion matrix - is a basic tool for the design and the diagnosis of a geodetic network. The question on which way criterion matrices can be established and which conditions must be fulfilled, arises. Some necessary definitions of subsequently used terms are presented in the first part of the paper. An attempt to motivate them is made. The second part deals with the derivation of the variance-covariance matrix of estimable quantities from the variance-covariance matrix of absolute or relative coordinates whereas the third part treats the contrary, that is the way from the criterion matrix of absolute or relative coordinates to the variance-covariance matrix of estimable quantities. In the last section the general establishment of artificial covariance matrices based on the problems of

correlation functions and stochastic processes is discussed. General remarks on the (geometrical) interpretation of homogeneous and isotropic criterion matrices are included.

I. Mathematical / statistical fundamentals

Homogeneity and isotropy of scalar- and vector-valued functions play an important role within the optimization of geodetic nets and the formulation of criterion matrices. So it is useful to give a definition of homogeneity and isotropy without getting too deep into the extensive theory of scalar and vector fields. Examples for scalar-valued functions are the distance of two points (2-point-function) and the angle as the difference of two directions (3-point-function); the position vector of a geodetic netpoint is an example for a vector-valued 1-point-function, the difference-vector of two geodetic netpoints an example for a vector-valued 2-point-function (Grafarend/Schaffrin 1979). Two additional definitions for measurable and estimable quantities have to be given subsequently (Molenaar 1981).

1. Scalar-valued functions

Let $s(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha)$ be a scalar-valued function of position-vectors $(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha)$. The variance-covariance of this signal referring to the points $(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha)$ and $\underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta$ is defined by:

$$\sum\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha,\ \underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta\right) =$$
$$E\left\{\left[s\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha\right) - E\left\{s\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha\right)\right\}\right]\ \left[s\left(\underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta\right) - E\left\{s\left(\underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta\right)\right\}\right]\right\}$$

$E\{\cdot\}$ stands for the expectation operator.

Def. 1:

A scalar-valued function is homogeneous and isotropic, if

$$\sum\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha,\ \underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta\right) =$$
$$\sum\left(\left|\underset{\sim}{r}_2-\underset{\sim}{r}_1\right|,\left|\underset{\sim}{r}_3-\underset{\sim}{r}_1\right|,\ldots,\left|\underset{\sim}{r}_\beta-\underset{\sim}{r}_{\beta-1}\right|\right)$$

holds, which means that $\Sigma\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha,\ \underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta\right)$ is only a function of the length of all difference-vectors $\underset{\sim}{r}_\gamma-\underset{\sim}{r}_\delta$ $(1\le\delta<\gamma\le\beta)$.

2. Vector-valued functions

Let $\underset{\sim}{s}(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha)$ be a vector-valued function of position vectors $\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha$. The variance-covariance of this signal referring to $\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha,\ \underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta$ is defined by:

$$\tilde{\Sigma}_{ij}\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha,\underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta\right) =$$
$$E\left\{\left[\underset{\sim}{s}_i\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha\right)-E\left\{\underset{\sim}{s}_i\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha\right)\right\}\right]\ \left[\underset{\sim}{s}_j\left(\underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta\right)-E\left\{\underset{\sim}{s}_j\left(\underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta\right)\right\}\right]\right\}$$

Def. 2:

A variance-covariance matrix of vector-signals is homogeneous and isotropic, if

$$\tilde{\Sigma}\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\alpha,\underset{\sim}{r}_{\alpha+1},\ldots,\underset{\sim}{r}_\beta\right) =$$
$$\sigma_o\left(\left|\underset{\sim}{r}_2-\underset{\sim}{r}_1\right|,\ldots,\left|\underset{\sim}{r}_\beta-\underset{\sim}{r}_{\beta-1}\right|\right)\cdot\delta_{ij}\left(\underset{\sim}{e}_i\otimes\underset{\sim}{e}_j\right)+$$
$$\sum_{\gamma=2}^{\beta}\sum_{\mu=\gamma}^{\beta}\sigma_{\gamma\mu}\left(\left|\underset{\sim}{r}_2-\underset{\sim}{r}_1\right|,\ldots,\left|\underset{\sim}{r}_\beta-\underset{\sim}{r}_{\beta-1}\right|\right)\left[\left(\underset{\sim}{r}_\gamma-\underset{\sim}{r}_{\gamma-1}\right)\otimes\left(\underset{\sim}{r}_\mu-\underset{\sim}{r}_{\mu-1}\right)+\left(\underset{\sim}{r}_\mu-\underset{\sim}{r}_{\mu-1}\right)\otimes\left(\underset{\sim}{r}_\mu-\underset{\sim}{r}_{\mu-1}\right)\right],$$

where σ_o and $\sigma_{\gamma\mu}(2\le\gamma<\mu\le\beta)$ are certain homogeneous and isotropic scalar-valued functions and the basic vectors $\underset{\sim}{e}_i$ span the corresponding Euclidean space. More easily expressed, homogeneity resp. isotropy means the invariance against translations res. rotations in the sense of

$$\Sigma\left(\underset{\sim}{r}_1+\underset{\sim}{t},\ldots,\underset{\sim}{r}_\beta+\underset{\sim}{t}\right)=\Sigma\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\beta\right)\ \text{resp.}\ \Sigma\left(R\underset{\sim}{r}_1,\ldots,R\underset{\sim}{r}_\beta\right)=\Sigma\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\beta\right)$$

for the variance-covariance of scalar signals and

$$\tilde{\Sigma}\left(\underset{\sim}{r}_1+\underset{\sim}{t},\ldots,\underset{\sim}{r}_\beta+\underset{\sim}{t}\right)=\tilde{\Sigma}\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\beta\right)\ \text{resp.}\ \tilde{\Sigma}\left(R\underset{\sim}{r}_1,\ldots,R\underset{\sim}{r}_\beta\right)=R\tilde{\Sigma}\left(\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_\beta\right)R'$$

for the variance-covariance of vector-signals. $\underset{\sim}{t}$ denotes a vector of translation-parameters, R a matrix of rotation-parameters. This fact can be shown if the variance-covariance matrix of Cartesian coordinates in a two-dimensional space is considered. The 2x2 subblocks characterize the local error ellipses, which are identical in size for homogeneity. The error ellipses degenerate to circles due to isotropy. For homogeneity and isotropy of the variance-covariance matrix of Cartesian coordinates the error ellipses degenerate to circles of the same size (→ Tylor-Karman-structure). It must be stressed that homogeneity and isotropy do not result into error ellipses being circles of the same size in general. This is only true for coordinates. On the other hand there is no geometrical interpretation for homogeneous and isotropic variance-covariance matrices for observables of type distance, angle, etc. The only way to prove homogeneity and isotropy is by using the above formulated definitions. The variance-covariance matrix of e.g. angles is homogeneous and isotropic, if it is a pure function of the length of all difference vectors. Finally it must be mentioned that the postulate of homogeneity and isotropy is only plausible for networks

where great importance is attached to relative accuracy and equal precision in all parts of the net. For special types of networks e.g. tunnel networks or deformation nets the postulate of homogeneity and especially isotropy is undoubtedly not the optimal one, because e.g. for the detection of movements certain directions may be preferred.

Def. 3:
Let y be a quantity which is represented by some stochastic variables y_i^*, defined by the i-times repetition of a measuring process. The variable y_i is said to be measurable if $E\{y^*\} = y$ holds. $E\{\cdot\}$ is the expectation operator.

Def. 4:
Functions are called (unbiasedly) estimable, if they can be represented as pure functions of measurable quantities without the introduction of extra parameters (datum parameters).

Corollary 1:
Measurable quantities are estimable.

Example 1: distance-measurements
Let $s^1_{\alpha\beta}$, $s^1_{\alpha\gamma}$ and $s^1_{\beta\gamma}$ represent the distances between three netpoints α,β,γ observed by a first instrument (or at a first epoch) and $s^2_{\alpha\beta}$, ... the distances between the same netpoints observed by a second instrument (or at a second epoch). Because of different length scale factors of the instruments (or of a time-varying of the scale between epoch 1 and epoch 2) it is clear that

$$E\{s^1_{\alpha\beta}\} \neq E\{s^2_{\alpha\beta}\},\ E\{s^1_{\alpha\gamma}\} \neq E\{s^2_{\alpha\gamma}\} \text{ and } E\{s^1_{\beta\gamma}\} \neq E\{s^2_{\beta\gamma}\}$$

To avoid this fact, one should proceed to distance ratios.

If we set $\omega^1_\alpha =: \frac{s^1_{\alpha\beta}}{s^1_{\alpha\gamma}}$, $\omega^1_\beta =: \frac{s^1_{\beta\gamma}}{s^1_{\alpha\beta}}$, $\omega^1_\gamma =: \frac{s^1_{\alpha\gamma}}{s^1_{\beta\gamma}}$

and $\omega^2_\alpha =: \frac{s^2_{\alpha\beta}}{s^2_{\alpha\gamma}}$, $\omega^2_\beta =: \frac{s^2_{\beta\gamma}}{s^2_{\alpha\beta}}$, $\omega^2_\gamma =: \frac{s^2_{\alpha\gamma}}{s^2_{\beta\gamma}}$

it is valid that $E\{\omega^1_j\} = E\{\omega^2_j\}$ for $j = \alpha,\beta,\gamma$

Example 2: theodolite measurements
Let $r^1_{\alpha\beta}$ and $r^1_{\alpha\gamma}$ represent the directions from a point α to two other points β and γ and $z^1_{\alpha\beta}$ resp. $z^1_{\alpha\gamma}$ the corresponding zenith angles measured with a theodolite in a first setup. Talking up position in a second setup, we will get the directions $r^2_{\alpha\beta}$ and $r^2_{\alpha\gamma}$ and the zenith angles $z^2_{\alpha\beta}$ resp. $z^2_{\alpha\gamma}$.

Because of different reference directions for the bearings and the zenith angles, it will hold that

$$E\{r^1_{\alpha\beta}\} \neq E\{r^2_{\alpha\beta}\} \;,\; E\{r^1_{\alpha\gamma}\} \neq E\{r^2_{\alpha\gamma}\} \;,\; E\{z^1_{\alpha\beta}\} \neq E\{z^2_{\alpha\beta}\} \;,\; E\{z^1_{\alpha\gamma}\} \neq E\{z^2_{\alpha\gamma}\} \;.$$

From the above observations the following expressions can be derived:

$$\cos\psi^1_{\gamma\alpha\beta} = \cos\left(r^1_{\alpha\beta} - r^1_{\alpha\gamma}\right)\sin z^1_{\alpha\beta}\sin z^1_{\alpha\gamma} + \cos z^1_{\alpha\beta}\cos z^1_{\alpha\gamma}$$

$$\cos\psi^2_{\gamma\alpha\beta} = \cos\left(r^2_{\alpha\beta} - r^2_{\alpha\gamma}\right)\sin z^2_{\alpha\beta}\sin z^2_{\alpha\gamma} + \cos z^2_{\alpha\beta}\cos z^2_{\alpha\gamma} \;.$$

Now it can be stated that $E\{\psi^1_{\gamma\alpha\beta}\} = E\{\psi^2_{\gamma\alpha\beta}\}$ is valid.

Corollary 2:

Distance ratios and angles are measurable and estimable quantities.
The reason for the use of estimable quantities is their applicability without any establishment of a datum for origin, orientation and scale of a coordinate system. Thus their variance-covariance matrix can be specified independently of those parameters.

II. The variance-covariance matrix of estimable quantities derived from the variance-covariance matrix of absolute coordinates

1. Distance ratios

Let ω_1 be the ratio of two distances s_1, s_2 and ω_2 the ratio of two other distances s_3 and s_4 (the quantities $\underline{r}$ denote the position vector of the corresponding net point):

$$\omega_1 = \omega\left(\underline{r}_1, \underline{r}_2, \underline{r}_3, \underline{r}_4\right) = \frac{s_1}{s_2} = \frac{s\left(\underline{r}_1, \underline{r}_2\right)}{s\left(\underline{r}_3, \underline{r}_4\right)} = \frac{\left[(x_1 - x_2)^2 + (y_1 - y_2)^2\right]^{1/2}}{\left[(x_3 - x_4)^2 + (y_3 - y_4)^2\right]^{1/2}}$$

$$\omega_2 = \omega\left(\underline{r}_5, \underline{r}_6, \underline{r}_7, \underline{r}_8\right) = \frac{s_3}{s_4} = \frac{s\left(\underline{r}_5, \underline{r}_6\right)}{s\left(\underline{r}_7, \underline{r}_8\right)} = \frac{\left[(x_5 - x_6)^2 + (y_5 - y_6)^2\right]^{1/2}}{\left[(x_7 - x_8)^2 + (y_7 - y_8)^2\right]^{1/2}}$$

Introducing approximate values, the expressions for $d\omega_1$ and $d\omega_2$ can be derived:

$$\begin{bmatrix} d\omega_1 \\ d\omega_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial\omega_1}{\partial x_1} & \frac{\partial\omega_1}{\partial x_2} & \frac{\partial\omega_1}{\partial y_1} & \frac{\partial\omega_1}{\partial y_2} & \frac{\partial\omega_1}{\partial x_3} & \frac{\partial\omega_1}{\partial x_4} & \frac{\partial\omega_1}{\partial y_3} & \frac{\partial\omega_1}{\partial y_4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\partial\omega_2}{\partial x_5} & \frac{\partial\omega_2}{\partial x_6} & \frac{\partial\omega_2}{\partial y_5} & \frac{\partial\omega_2}{\partial y_6} & \frac{\partial\omega_2}{\partial x_7} & \frac{\partial\omega_2}{\partial x_8} & \frac{\partial\omega_2}{\partial y_7} & \frac{\partial\omega_2}{\partial y_8} \end{bmatrix}$$

$$\left[dx_1\ dx_2\ dy_1\ dy_2\ dx_3\ dx_4\ dy_3\ dy_4\ dx_5\ dx_6\ dy_5\ dy_6\ dx_7\ dx_8\ dy_7\ dy_8\right] =$$

$$= \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \begin{bmatrix} d'_1 \\ d'_2 \end{bmatrix}$$

The use of the law of propagation of errors leads to the variance-covariance matrix Ω of the two distance ratios. Ω is a so-called eight-point function because of its dependency on eight position vectors representing the netpoints.
The two variance-covariance matrices are defined by

$$\Omega\left(\underline{r}_1,\underline{r}_2,\underline{r}_3,\underline{r}_4,\underline{r}_5,\underline{r}_6,\underline{r}_7,\underline{r}_8\right) = E\{[d\omega_1 - E\{d\omega_1\}]\,[d\omega_2 - E\{d\omega_2\}]\}$$

$$X\left(\underline{r}_1,\underline{r}_2,\underline{r}_3,\underline{r}_4,\underline{r}_5,\underline{r}_6,\underline{r}_7,\underline{r}_8\right) = E\{[d_1 - E\{d_1\}]' \; [d_2 - E\{d_2\}]\}$$

$$\Omega = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} X \begin{bmatrix} b'_1 & 0' \\ 0' & b'_2 \end{bmatrix}$$

$$= \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \begin{bmatrix} Q_{d_1d_1} & Q_{d_1d_2} \\ Q_{d_2d_1} & Q_{d_2d_2} \end{bmatrix} \begin{bmatrix} b'_1 & 0' \\ 0' & b'_2 \end{bmatrix}$$

$$= \begin{bmatrix} b_1Q_{d_1d_1}b'_1 & b_1Q_{d_1d_2}b'_2 \\ b_2Q_{d_2d_1}b'_1 & b_2Q_{d_2d_2}b'_2 \end{bmatrix}$$

It can be easily shown that the variance-covariance matrix $\Omega(\underline{r}_1,\ldots,\underline{r}_8)$ for distance ratios is a homogeneous and isotropic scalar-valued function, because it is only a function of the length of all difference vectors and the homogeneous and isotropic variance-covariance matrix of Cartesian coordinates X.

2. Angles

Let $A\left(\underline{r}_1,\underline{r}_2\right)$ and $A\left(\underline{r}_1,\underline{r}_3\right)$ be the azimuths of two network sides - $\underline{r}_1$ the position vector of the station, $\underline{r}_2$ and $\underline{r}_3$ the position vectors of the target points. $A\left(\underline{r}_4,\underline{r}_5\right)$ and $A\left(\underline{r}_4,\underline{r}_6\right)$ are two other azimuths respectively.
To simplify matters, we confine ourselves again to a two dimensional space.
To receive estimable quantities, angles are developed from azimuths:

$$\alpha_1 := \alpha\left(\underline{r}_1,\underline{r}_2,\underline{r}_3\right) = A\left(\underline{r}_1,\underline{r}_2\right) - A\left(\underline{r}_1,\underline{r}_3\right)$$

$$\alpha_2 := \alpha\left(\underline{r}_4,\underline{r}_5,\underline{r}_6\right) = A\left(\underline{r}_4,\underline{r}_5\right) - A\left(\underline{r}_4,\underline{r}_6\right)$$

The variance-covariance matrix of α_1 and α_2 is given through

$$\Lambda\left(\underline{r}_1,\underline{r}_2,\underline{r}_3,\underline{r}_4,\underline{r}_5,\underline{r}_6\right) = E\{[d\alpha_1 - E\{d\alpha_1\}]\,[d\alpha_2 - E\{d\alpha_2\}]\}$$

where it is assumed that the relations between angles and coordinates could have been linearized with the introduction of approximate values.

$$\begin{bmatrix} d\alpha_1 \\ d\alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial\alpha_1}{\partial x_1} & \frac{\partial\alpha_1}{\partial x_2} & \frac{\partial\alpha_1}{\partial x_3} & \frac{\partial\alpha_1}{\partial y_1} & \frac{\partial\alpha_1}{\partial y_2} & \frac{\partial\alpha_1}{\partial y_3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{\partial\alpha_2}{\partial x_4} & \frac{\partial\alpha_2}{\partial x_5} & \frac{\partial\alpha_2}{\partial x_6} & \frac{\partial\alpha_2}{\partial y_4} & \frac{\partial\alpha_2}{\partial y_5} & \frac{\partial\alpha_2}{\partial y_6} \end{bmatrix}$$

$$\begin{bmatrix} dx_1 & dx_2 & dx_3 & dy_1 & dy_2 & dy_3 & dx_4 & dx_5 & dx_6 & dy_4 & dy_5 & dy_6 \end{bmatrix} =$$

$$= \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \begin{bmatrix} d'_1 \\ d'_2 \end{bmatrix}$$

$A(\underset{\sim}{r}_1, \underset{\sim}{r}_2) - A(\underset{\sim}{r}_1, \underset{\sim}{r}_3)$ may be defined to

$\arctan\frac{y_2 - y_1}{x_2 - x_1} - \arctan\frac{y_3 - y_1}{x_3 - x_1}$, so that (after linearization) the partial derivatives with respect to the coordinate corrections can be calculated. With

$$X(\underset{\sim}{r}_1, \underset{\sim}{r}_2, \underset{\sim}{r}_3, \underset{\sim}{r}_4, \underset{\sim}{r}_5, \underset{\sim}{r}_6) = E\{[d_1 - E\{d_1\}]'[d_2 - E\{d_2\}]\}$$

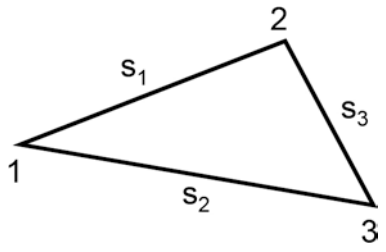
the variance-covariance matrix Λ of angles can be determined as before.

$$\Lambda = \begin{bmatrix} b_1 Q_{d_1 d_1} b'_1 & b_1 Q_{d_1 d_2} b'_2 \\ b_2 Q_{d_2 d_1} b'_1 & b_2 Q_{d_2 d_2} b'_2 \end{bmatrix}$$

And again it can be easily verified that Λ is a homogeneous and isotropic scalar-valued six-point function because of its dependency on only the length of all difference vectors.

III. The variance-covariance matrix of absolute coordinates derived from the variance-covariance matrix of estimable quantities

For a three-point-network in two dimensional space the variance-covariance matrix of absolute coordinates is allocated to the variance-covariance matrix of distance ratios.



The network configurations is given through two distance ratios

$$\omega_1 = \frac{s_1}{s_2} \quad \text{and} \quad \omega_2 = \frac{s_1}{s_3} \ ,$$

so the starting point will be the homogeneous and isotropic variance-covariance matrix Ω of ω_1 and ω_2:

$$\Omega = \begin{bmatrix} E\{[d\omega_1 - E\{d\omega_1\}]\ [d\omega_1 - E\{d\omega_1\}]\} & E\{[d\omega_1 - E\{d\omega_1\}]\ [d\omega_2 - E\{d\omega_2\}]\} \\ E\{[d\omega_2 - E\{d\omega_2\}]\ [d\omega_1 - E\{d\omega_1\}]\} & E\{[d\omega_2 - E\{d\omega_2\}]\ [d\omega_2 - E\{d\omega_2\}]\} \end{bmatrix}$$

with

$$\omega_1 = \frac{s\left(\underset{\sim}{r}_1, \underset{\sim}{r}_2\right)}{s\left(\underset{\sim}{r}_1, \underset{\sim}{r}_3\right)} = \frac{s_1}{s_2} = \frac{\left[(x_1 - x_2)^2 + (y_1 - y_2)^2\right]^{1/2}}{\left[(x_1 - x_3)^2 + (y_1 - y_3)^2\right]^{1/2}}$$

$$\omega_2 = \frac{s\left(\underset{\sim}{r}_1, \underset{\sim}{r}_2\right)}{s\left(\underset{\sim}{r}_2, \underset{\sim}{r}_3\right)} = \frac{s_1}{s_3} = \frac{\left[(x_1 - x_2)^2 + (y_1 - y_2)^2\right]^{1/2}}{\left[(x_2 - x_3)^2 + (y_2 - y_3)^2\right]^{1/2}}$$

and

$$d\omega_1 = \frac{\partial\omega_1}{\partial x_1}dx_1 + \frac{\partial\omega_1}{\partial x_2}dx_2 + \frac{\partial\omega_1}{\partial x_3}dx_3 + \frac{\partial\omega_1}{\partial y_1}dy_1 + \frac{\partial\omega_1}{\partial y_2}dy_2 + \frac{\partial\omega_1}{\partial y_3}dy_3$$

$$d\omega_2 = \frac{\partial\omega_2}{\partial x_1}dx_1 + \frac{\partial\omega_2}{\partial x_2}dx_2 + \frac{\partial\omega_2}{\partial x_3}dx_3 + \frac{\partial\omega_2}{\partial y_1}dy_1 + \frac{\partial\omega_2}{\partial y_2}dy_2 + \frac{\partial\omega_2}{\partial y_3}dy_3$$

or in matrix notation

$$\begin{bmatrix} d\omega_1 \\ d\omega_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \begin{bmatrix} dx' \\ dy' \end{bmatrix}$$

where b_1 resp. b_2 are the two row-vectors containing the partial derivatives corresponding to ω_1 resp. ω_2.
The use of the law of propagation of errors leads to

$$\Omega = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} X\ [b'_1, b'_2] = BXB'$$

with X the variance-covariance matrix of the coordinates. The above equation has to be inverted with respect to X. It holds that

$$o(\Omega) = 2x2,\ o(X) = 6x6,\ o(b_1) = o(b_2) = 1x6 \ ,$$

which means that we are concerned with the solution of a system of equations with a rank deficiency. This problem arises because absolute coordinates and their variance-covariance matrix cannot be determined without a definition for orientation and origin of a coordinate system. One possibility to solve the problem is to fix the coordinates of two points. But this leads to the undesirable fact that these points will have variances of zero. Another way to solve $\Omega = BXB'$ is outlined in the following.

Using the Kronecker-product "$\otimes$", the above matrix equation can be rewritten as a vector equation

$$\operatorname{vec}\Omega = (B \otimes B)\operatorname{vec} X$$

where the vec of a (symmetric) matrix stacks columns of it one under another in a single column. Because Ω and X are symmetric matrices, it will be sufficient to confine oneself in the essential elements. This is done by the "vech"-operator, which does the same thing as the "vec"-operator, but starting each column at the matrix' diagonal element (Henderson/Searle 1979).

$$\operatorname{vech}\Omega = H(B \otimes B)G \operatorname{vech} X$$

where H and G are those matrices that define the transformations

$$H \operatorname{vec}\Omega = \operatorname{vech}\Omega$$

$$G \operatorname{vech} X = \operatorname{vec} X$$

$$o(\operatorname{vech}\Omega) = 3x1,\ o(\operatorname{vech} X) = 21x1,\ o(B \otimes B) = 4x36$$

$$o(H) = 3x4,\ o(G) = 36x21$$

The inversion of $\operatorname{vech}\Omega = H(B \otimes B)G \operatorname{vech} X$ can be performed using e.g. a right-inverse of $H(B \otimes B)G$, which corresponds to a minimum norm solution (Grafarend 1982)

$$\operatorname{vech} X = G'(B \otimes B)'H'[H(B \otimes B)GG'(B \otimes B)'H']^{-1} \operatorname{vech}\Omega$$

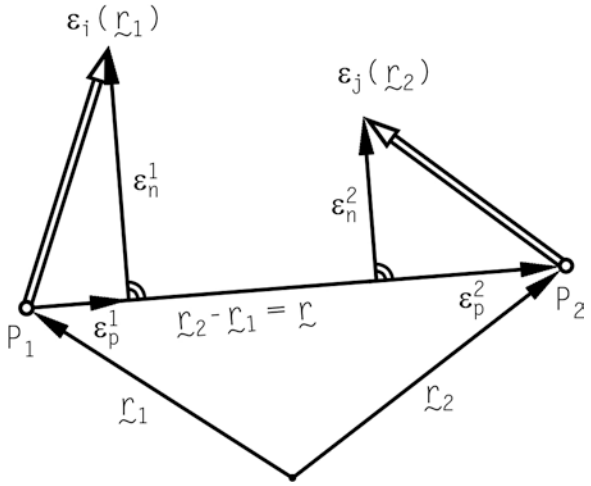
It can be shown that X as a variance-covariance matrix for absolute coordinates is not homogeneous and isotropic. The investigation of the way leading to homogeneity and isotropy (weighted generalized inverses, etc.) is left open for the future.

IV. Artificial covariance matrices, correlation function and stochastic process

In the last chapters use was made of idealized variance-covariance matrices (criterion matrices), which serve as a criterion for the precision of a geodetic network. They are applied to optimize a network with respect to the nets configuration (FOD = First Order Design) or with respect to the weights of the observations (SOD = Second Order Design). So the question arises how to generate such matrices as a starting point for the optimization. The basic idea for this is that the errors of position vectors constitute a stochastic process. Within that concept, a geodetic net is an inhomogeneous and unisotropic field of error-vectors and the error situation is described by the covariance function (Grafarend 1972). For the experts, who are concerned with the interpretation of the precision of general geodetic nets, the error situation of a new determined point is the best when its error ellipse is a circle (isotropy!) with minimal radius

(except for special network types such as deformation nets, etc.) Further on it would be very sufficient if the error situation is uniform over the complete network (homogeneity). That means that the idealized variance-covariance matrix of the net coordinates should be homogeneous and isotropic. A homogeneous and isotropic net is endowed with the so-called TAYLOR-KARMAN structure. The correlations between the error vectors of the netpoints are described by two characteristic functions, longitudinal and lateral correlation function (Grafarend/Schaffrin 1979), Schaffrin/Grafarend/Schmitt 1977).
Let $\varepsilon_i(\underset{\sim}{r}_1)$ be the error vector for a point P_1 defined by the position vector $\underset{\sim}{r}_1$, and $\varepsilon_j(\underset{\sim}{r}_2)$ the corresponding error vector for a point P_2. The covariance function ϕ_{ij} is the central moment of second order, describes completely the error situation of a geodetic net.

$$\phi_{ij} = E\left\{\varepsilon_i\left(\underset{\sim}{r}_1\right)\ \varepsilon_j\left(\underset{\sim}{r}_2\right)\right\}$$



For $\underset{\sim}{r}_2 = \underset{\sim}{r}_1$ follows the error situation for <u>one</u> point

$$\phi_{ij} = E\left\{\varepsilon_i\left(\underset{\sim}{r}_1\right)\ \varepsilon_j\left(\underset{\sim}{r}_2\right)\right\} = \sigma_{ij}\left(\underset{\sim}{r}_1\right)$$

and for $\underset{\sim}{r}_2 \neq \underset{\sim}{r}_1$, ϕ_{ij} describes the correlations between point P_1 and point P_2. Autocorrelations result for i = j, crosscorrelations for i ≠ j. For a homogeneous and isotropic situation, $\phi_{ij}\left(\underset{\sim}{r}_1, \underset{\sim}{r}_2\right)$ has the so-called TAYLOR-KARMAN structure, and is holds that

$$\phi_{ij}\left(\underset{\sim}{r}_1, \underset{\sim}{r}_2\right) = \phi_{ij}\left(\underset{\sim}{r}_2 - \underset{\sim}{r}_1\right) = \Sigma_m\left(\underset{\sim}{r}\right)\delta_{ij} + \left[\Sigma_l\left(\underset{\sim}{r}\right) - \Sigma_m\left(\underset{\sim}{r}\right)\right]\frac{\Delta x_i \Delta x_j}{r^2} \quad .$$

ϕ_{ij} is only a function of the difference vector $\underset{\sim}{r} = \underset{\sim}{r}_2 - \underset{\sim}{r}_1$ respectively a function of the above mentioned scalar-valued longitudinal and lateral function Σ_l and Σ_m.

$\Sigma_l(\underset{\sim}{r}) = E\{\varepsilon_p(\underset{\sim}{r}_1)\,\varepsilon_p(\underset{\sim}{r}_2)\}$ denotes the autocorrelation of the parallel component ε_p of ε, $\Sigma_m(\underset{\sim}{r}) = E\{\varepsilon_n(\underset{\sim}{r}_1)\,\varepsilon_n(\underset{\sim}{r}_2)\}$ the autocorrelation of the normal component ε_n of ε.

The correlations $E\{\varepsilon_p(\underset{\sim}{r}_1)\,\varepsilon_n(\underset{\sim}{r}_2)\}$ resp. $E\{\varepsilon_n(\underset{\sim}{r}_1)\,\varepsilon_p(\underset{\sim}{r}_2)\}$ are zero. δ_{ij} is the Kroneckersymbol with $\delta_{ij} = 1$ for $i=j$ and $\delta_{ij} = 0$ for $i \neq j$; Δx_i is the difference vector $x_i^{P_2} - x_i^{P_1}$ and r is the length of $\underset{\sim}{r}$. By the way $\Sigma_m(\underset{\sim}{r})$ and $\Sigma_l(\underset{\sim}{r})$ are the eigenvalues of $\phi_{ij}(\underset{\sim}{r}_1, \underset{\sim}{r}_2)$ which allow the construction of the error ellipses. The elements of a criterion matrix Q with Taylor-Karmann structure can now be formulated (Wimmer 1981)

$$Q_{x_i x_j} = \Sigma_m(\underset{\sim}{r}) + [\Sigma_l(\underset{\sim}{r}) - \Sigma_m(\underset{\sim}{r})]\,\frac{\Delta x^2}{r^2}$$
$$= \Sigma_m(\underset{\sim}{r}) + [\Sigma_l(\underset{\sim}{r}) - \Sigma_m(\underset{\sim}{r})]\,\cos^2 A_{ij}$$

$$Q_{x_i y_j} = Q_{y_i x_j} = [\Sigma_l(\underset{\sim}{r}) - \Sigma_m(\underset{\sim}{r})]\,\frac{\Delta x\,\Delta y}{r^2}$$
$$= [\Sigma_l(\underset{\sim}{r}) - \Sigma_m(\underset{\sim}{r})]\,\sin A_{ij} \cos A_{ij}$$

$$Q_{y_i y_j} = \Sigma_m(\underset{\sim}{r}) + [\Sigma_l(\underset{\sim}{r}) - \Sigma_m(\underset{\sim}{r})]\,\frac{\Delta y^2}{r^2}$$
$$= \Sigma_m(\underset{\sim}{r}) + [\Sigma_l(\underset{\sim}{r}) - \Sigma_m(\underset{\sim}{r})]\,\sin^2 A_{ij}$$

with $\Delta x = x_j - x_i$, $\Delta y = y_j - y_i$, $A_{ij} = \arctan \frac{\Delta y}{\Delta x}$. Because the error ellipses of absolute coordinates have to be circles, $Q_{x_i x_i} = Q_{y_i y_i} = 1$ and $Q_{x_i y_i} = 0$ for the normalized case. Therefore $\Sigma_m(o) = \Sigma_l(o) = 1$.

For the determination of $\Sigma_l(\underset{\sim}{r})$ and $\Sigma_m(\underset{\sim}{r})$ for $\underset{\sim}{r} \neq 0$ models of covariance functions have to be introduced. Because of the multiplicity of the used models, it is only mentioned that a covariance function C(r) should fulfill four postulates

a) $C(-r) = C(r)$

b) $C(=) = \sigma^2 > 0$

c) $|C(r)| \leq C(0)$

d) $\lim_{r \to \infty} C(r) = 0$

The corresponding correlation function results from a normalization of C(r) by dividing C(r) by $C(0) = \sigma^2$.

An important parameter the correlation functions depend on, is the so-called characteristic distance, which serves as a range of influence. The determination of the characteristic distance is done by empirical investigations of real correlations in model networks with homogeneous size (Schmitt 1981). An excellent overview of the different types of covariance functions is given in Wimmer (1981).

Conclusions

The main problem within the optimization of geodetic networks is the formulation of appropriate criterion matrices. A first step for their determination in a two-dimensional space is already done by using correlation functions, but the answer how the derived informations correspond to correlations in real networks is still outstanding. Because the choice of the longitudinal and lateral function Σ_l and Σ_m for the crosscorrelations is rather arbitrary (restrictions for them are only the above specified four postulates), the results of any optimization procedure (e.g. the result of the SOD) may be possibly fairly weak.
The introduction of estimable quantities extricates in the first step from the problems of orientation and origin of a coordinate system. Nevertheless one is concerned later on with the solution of a rank deficient system of equations. In this case, in order to get a solution we are free in the choice of restrictions or other techniques. However the chosen concept (e.g. homogeneity and isotropy) offers some tools to make a solution possible.

Acknowledgement

This report was sponsored by the German National Science Foundation (Deutsche Forschungsgemeinschaft) under grant Gr 323/12.
This support is gratefully acknowledged.

REFERENCES

GRAFAREND, E., 1972: Genauigkeitsmaße geodätischer Netze. Deutsche Geodätische Kommission, Report A 73, München, FRG

GRAFAREND, E., and SCHAFFRIN, B., 1979: Kriterion Matrizen I - zweidimensionale homogene und isotrope geodätische Netze. Zeitschrift für Vermessungswesen 104, 133-149

GRAFAREND, E., 1982: Optimal Design of Geodetic Networks. Deutsche Geodätische Kommission, Report 258/III, München, FRG

HENDERSON, H.V., and SEARLE, S.R., 1979: Vec and Vech operators for matrices, with some uses in Jacobians and multivariate statistics. Canad. J. Statist. 7, 65-81

MOLENAAR, M., 1981: A further inquiry into the theory of S-transformations and criterion matrices. Netherlands Geodetic Commission, Volume 7, Number 1

SCHAFFRIN, B., GRAFAREND, E., and SCHMITT, G., 1977: Kanonisches Design Geodätischer Netze I. Manuscripta Geodaetica 2, 263-306

SCHMITT, G., 1981: Optimal Design of Geodetic Networks. IAG-Symposium on geodetic networks and computations, Munich, FRG

WIMMER, H., 1981: Ein Beitrag zur Gewichtsoptimierung geodätischer Netze. Dissertation script, München, FRG

DIFFICULTIES IN DEFINING THE QUALITY OF GEODETIC NETWORKS

Jan van MIERLO

Karlsruhe, Federal Republic of Germany

ABSTRACT

The quality of a geodetic network is concerned with the precision and the reliability of the coordinates. The paper reviews various measures for precision and reliability which are related to the specification of the functional use of the network. The difficulties in using a criterion, which is not invariant with respect to S-transformation, are indicated. The use of a criterion matrix is outlined.

The sensitivity of a network for checking special hypothesis is discussed,

1. Introduction

For practical purposes the position of points in a free geodetic network are most conveniently expressed by coordinates. It is customary to express the precision of their determination by the covariance matrix of the coordinate variates. However, a covariance matrix consists of a mass of numbers of different values and signs. Hence it is very difficult to conclude anything about the structure of such a matrix by simply inspecting these numbers. In order to get an overall picture of the precision it is often usual to draw the point standard ellipses in two- and ellipsoids in three-dimensional networks. However, the standard ellipses, ellipsoids of the points of the network are dependent on the way the coordinates are computed. The same also applies to the relative standard ellipses (ellipsoids), pertaining to coordinate differences of points of the network. It is essential to represent the precision of a network with invariant quantities, invariant with respect to the coordinate definition. This aspect shall be discussed in this paper.

The quality of a network is related to the specification of its functional use. In a general purpose network the main requirement will usually be good overall precision e.g. all network points in a two-dimensional network have circular standard ellipses with the same radius. This requirement will be considered in more detail. In a network designed for special purposes like e.g. detection of

deformations often a special requirement for precision and reliability is given.

Often in geodetic practice redundant observations are measured in order to make checks for model errors possible. The goodness of the control for model errors is dependent on the design of the network and the precision of the observation variates. Undetected errors affect the computed coordinates, not the precision of the coordinates. The reliability of the network is determined by the goodness of the checking of the observations and the adjustment model. The reliability and the precision of a network are interrelated. They define the quality of a network. However, an optimum precision in a general purpose network doesn't guarantee an optimum reliability. The measures for quality of a network, precision and reliability, must be independent on the chosen coordinate definition, in terms of Baarda, independent with respect to an S-transformation, BAARDA (1973).

2. Measures for precision

A notation that has widespread usage is the following: $\hat{x}$ is used to denote an estimate of x (mean of x), and more generally $(\hat{x}_1,\ldots,\hat{x}_n)^T$ is a vector that estimates the vector $(\tilde{x}_1,\ldots,\tilde{x}_n)^T$. Now let C_x denote the regular covariance matrix of the estimators. Let $\varphi(x) = f^T x$ be a linear function of x with $f^T=(f_1,\ldots,f_n)$ and $x^T=(x_1,\ldots,x_n)$. By means of $\hat{x}$ we can estimate $\varphi(x)$:

$$\Phi = f^T \hat{x} \text{ : estimate of } \varphi(x).$$

The variance of Φ is given by

$$\sigma_\Phi^2 = f^T C_x f \ .$$

If we obtain for any linear function φ small variances, the quality of the estimated $\hat{x}$ can be regarded as "good". A maximum for σ_Φ^2 doesn't exist but

$$\max\left(\frac{f^T C_x f}{f^T f}\right) = \lambda_{max}$$

can be regarded as a measure for the quality of $\hat{x}$, where λ_{max} is the maximum eigenvalue of the covariance matrix C_x.

It can be proved that

$$\lambda_{min} f^T f < \sigma_\Phi^2 < \lambda_{max} f^T f$$

where λ_{min} is the minimum eigenvalue of C_x. The maximum eigenvalue indicates a tendency to expect the worst. A small value of λ_{max} indicates a good precision of $\hat{x}$.

If λ_{max} is an "outlier" with respect to the other eigenvalues, the "mean" quality of $\hat{x}$ can be better expressed by

1. the average of all the eigenvalues
2. the geometric mean of the eigenvalues.

These measures of quality are simply

$$\frac{1}{n} \sum_{i=1}^{n} \lambda_i = \frac{1}{n} \mathrm{Tr}\,(C_x)$$

$$\sqrt[n]{\lambda_1 \ldots \lambda_n} = \sqrt[n]{\mathrm{Det}\,(C_x)}$$

The determinant of the covariance matrix C_x is known as Wilks' generalized variance.

Another measure for the precision of $\hat{x}$ may be

$$\lambda_{max} - \lambda_{min} \ .$$

The smaller the value of $\lambda_{max} - \lambda_{min}$, the better is the precision of $\hat{x}$, see also GRAFAREND (1979). The standard hyperellipsoid is given by the equation

$$y^T C_x^{-1} y = 1 \ .$$

The axes of the standard ellipsoid are $\sqrt{\lambda_i}$, $i = 1 \ldots n$, where λ_i are the eigenvalues of C_x. If $\lambda_{max} - \lambda_{min} \to 0$ the hyperellipsoid corresponds closer to a hypersphere.

3. Comparison of covariance matrices

Let $\hat{x}_1^T = (\hat{x}_1, \ldots, \hat{x}_n)$ and $\hat{x}_2^T = (\hat{y}_1, \ldots, \hat{y}_n)$ are estimates of $\Theta^T = (\Theta_1, \ldots, \Theta_n)^T$. The corresponding covariance matrices are C_1 and C_2. The problem is to develop a criterion by which can be decided which estimate is better in the sense of precision. As long as no decision has been made on the use of the estimates we have to take into account all possible functions $\varphi(\Theta)$. For every function we require a good estimate. There are two ways in which we can estimate $\varphi(\Theta)$, i.e. $\varphi(\hat{x}_1)$ and $\varphi(\hat{x}_2)$. For linear functions we obtain

$$\Phi(\hat{x}_1) = f^T \hat{x}_1$$

$$\Phi(\hat{x}_2) = f^T \hat{x}_2$$

where $f^T = (f_1, \ldots, f_n)$.

We define $\hat{x}_1$ better than $\hat{x}_2$ or as good as $\hat{x}_2$ if for any function

$$\sigma^2_{\Phi(\hat{x}_1)} \leq \sigma^2_{\Phi(\hat{x}_2)}$$

hence $f^T C_1 f \leq f^T C_2 f \qquad \forall f$.

Then we have

$$f^T (C_1 - C_2) f \leq 0$$

or $C_1 - C_2$ negative semi-definite.

A matrix is negative semi-definite if all the non zero eigenvalues are negative. If all the diagonal elements of a matrix are negative, the matrix is not necessarily negative definite. It is a necessary but not a sufficient condition for a regular negative definite matrix. This means if

$$(C_1)_{ii} < (C_2)_{ii} \quad \forall i$$

and hence also $Tr(C_1) < Tr(C_2)$

the estimate $\hat{x}_1$ is according to our definition not necessarily better than $\hat{x}_2$. For geodetic applications it means that by only comparing the trace of covariance matrices of coordinates of two different network designs, one doesn't have a guarantee that the matrix with minimum trace is the best one.

Of course, it is not necessary to compute all the eigenvalues of C_1-C_2. If the maximum non zero eigenvalue is negative we can state that C_1 is better than C_2.

Instead of $f^T(C_1-C_2)f \leq 0$ we

can also write

$$f^T A^{-1} A (C_1-C_2) A^T A^{T^{-1}} f \leq 0$$

or with $f^T A^{-1} = \bar{f}^T$

$$\bar{f}^T A (C_1-C_2) A^T \bar{f} \leq 0 \ .$$

The eigenvalues of $A(C_1-C_2)A^T$ are in general different from the eigenvalues of C_1-C_2. Let us define new estimates $\hat{Y}_1$ and $\hat{Y}_2$

$$\hat{Y}_1 = A\hat{x}_1 \quad \text{with} \quad \bar{C}_1 = A C_1 A^T$$

$$\hat{Y}_2 = A\hat{x}_2 \quad \text{with} \quad \bar{C}_2 = A C_2 A^T \ .$$

Instead of comparing C_1 and C_2 one can also compare $\bar{C}_1$ and $\bar{C}_2$. Hence $\hat{Y}_1$ is better than $\hat{Y}_2$ if $\bar{C}_1$-$\bar{C}_2$ is negative definite. All the non zero eigenvalues $\bar{\lambda}_i$ of $\bar{C}_1$-$\bar{C}_2$ must be negative:

$$\bar{C}_1-\bar{C}_2 = A(C_1-C_2)A^T \ .$$

In general the eigenvalues $\bar{\lambda}_i$ are different from the eigenvalues λ_i from C_1-C_2. For that reason we cannot use the eigenvalues of C_1-C_2 as measures for the "goodness" of C_1 with respect to C_2.

A measure for quality or "goodness" may be the maximum of

$$\max \frac{\sigma^2_{\Phi(\hat{x}_1)}}{\sigma^2_{\Phi(\hat{x}_2)}} = \frac{f^T C_1 f}{f^T C_2 f} \qquad \forall f$$

If can be proved that this maximum is the maximum general eigenvalue of C_1 with respect to C_2. μ is a general eigenvalue of C_1 with respect to C_2 if for $x \neq 0$.

$$C_1 x = \mu C_2 x \ .$$

The vector x is the general eigenvector. Note that the general eigenvalues are the eigenvalues of $C_2^{-1}C_1$ (C_2 is regular). If μ_{max} denotes the maximum general eigenvalue of C_1 with respect to C_2 we have

$$\mu_{max} = \max \frac{f^T C_1 f}{f^T C_2 f} \ .$$

If $\mu_{max} \leq 1$, than C_1 is "better than or at least as good as" C_2. Instead of the estimators itselves we use the covariance matrices.

We can use μ_{max} as a measure for the quality of C_1 compared with C_2. It can be proved that the general eigenvalues are invariant with respect to a non-singular transformation. This property is a decisive argument to use the maximum general eigenvalue as a measure of the goodness of the estimate $\hat{x}_1$ and $\hat{x}_2$.

4. Scaling of covariance matrices

Now let C_2 be multiplied with a scalar k. We consider the problem to find values for k such that

$$C_1 - kC_2$$

is negative semi-definite. It turns out that the factor k must be greater than or equal to μ_{max} of C_1 with respect to C_2.

$$k > \mu_{max}$$

Having μ_{max}, we can "scale" the matrix C_2 with a factor $\geq \mu_{max}$. Hence if we multiply the estimate $\hat{x}_2$ with $\sqrt{k} \geq \sqrt{\mu_{max}}$ than the estimate $\hat{x}_1$ is better than $\sqrt{k}\hat{x}_2$ as $C_1 - k\,C_2$ is negative semi-definite.

Suppose $\hat{x}_2$ itself has not been given but only the covariance matrix C_2, which can now be used to test the goodness of $\hat{x}_1$. The matrix C_2 can now be called a criterion matrix. If the maximum general eigenvalue of C_1 with respect to C_2, $\mu_{max} \leq 1$, then the estimate $\hat{x}_1$ is at least as good as the unknown estimate which would follow from C_2. If apart from a scale factor the criterion matrix is known, the estimate $\hat{x}_1$ or C_2 is at least as good as $k\,C_2$ if

$$k \geq \mu_{max} \quad .$$

The requirement

$$\forall f \quad f^T C_1 f \leq f^T C_2 f$$

i.e. all general eigenvalues of C_1 with respect to C_2 are ≤ 1 also implies

$$\forall f \quad f^T C_2^{-1} f \leq f^T C_1^{-1} f \quad .$$

Now let $f^T C_1^{-1} f = 1$, then

$$f^T C_2^{-1} f \leq 1.$$

Hence if f is a point of the standard hyperellipsoid

$$f^T C_2^{-1} f = 1$$

this point is lying within the standard hyperellipsoid

$$f^T C_1^{-1} f \leq 1.$$

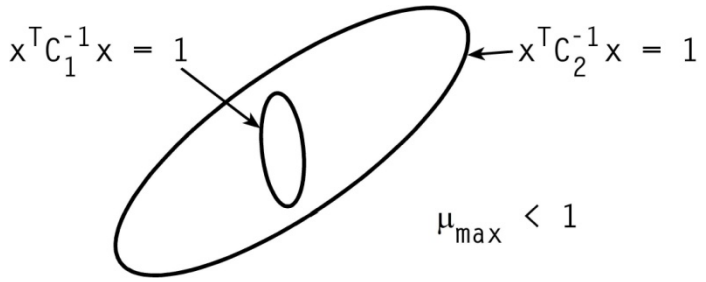


fig. 1

If the maximum eigenvalue of C_1 is less than the maximum eigenvalue of C_2, it is still possible that $\mu_{max} > 1$. This situation is sketched in fig. 2.

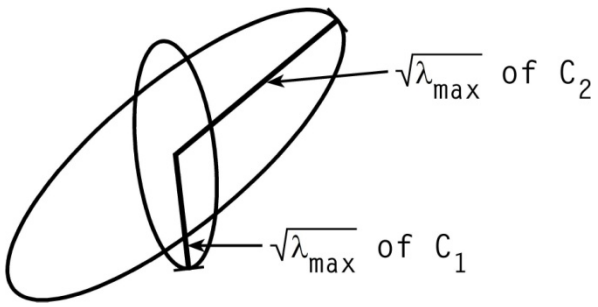


fig. 2

If $\mu_{max} \leq 1$, the form of the two standard hyperellipsoids will agree better if the difference between the maximum and the minimum general eigenvalue approaches zero. Considering only μ_{max} can give too pessimistic impression of the goodness or quality of the compared estimates. The computation of $\mu_{max} - \mu_{min}$ should not be omitted. If this difference is great the mean value of the general eigenvalues or its geometric mean can be used as an "average" measure of goodness of C_1 with respect to C_2.

$$\bar{\mu} = \frac{1}{n} \sum_{i=1}^{n} \mu_i$$

$$\bar{\bar{\mu}} = \sqrt[n]{\mu_1, \mu_2, \ldots, \mu_n}$$

Note that $\bar{\mu}$ can also be computed from the trace of the matrix $C_2^{-1}C_1$:

$$\bar{\mu} = \frac{1}{n} \mathrm{Tr}\left(C_2^{-1}C_1\right).$$

The geometric mean can be written as

$$\bar{\bar{\mu}} = \sqrt[n]{\mathrm{Det}\left(C_2^{-1}C_1\right)}.$$

The parameters $\bar{\mu}$ and $\bar{\bar{\mu}}$ are invariant with respect to a non-singular transformation. Note that if C_2 is the identity matrix we obtain the same measures for quality of estimates $\hat{x}_1$ which covariance matrix C_1, which has been discussed in the previous paragraph. Summarizing, the following quantities can be used to compare the goodness of $\hat{x}_1$ and $\hat{x}_2$ with respectively covariance matrices C_1 and C_2:

1. μ_{max} maximum general eigenvalue of C_1 with respect to C_2
2. $\mu_{max} - \mu_{min}$
3. $\frac{1}{n} \sum_{i=1}^{n} \mu_i = \frac{1}{n} \mathrm{Tr}\left(C_2^{-1}C_1\right)$ arithmetic mean
4. $\sqrt[n]{\mu_1, \mu_2, \ldots, \mu_n} = \sqrt[n]{\mathrm{Det}\left(C_2^{-1}C_1\right)}$ geometric mean

The eigenvalues of $C_2^{-1}C_1$ are equivalent with the general eigenvalues of C_2 with respect to C_1. So far we have outlined a technique to compare two different estimates of the mean values $\Theta = (\Theta_1, \ldots, \Theta_n)$ with known covariance matrices.

5. Geodetic applications

First, we shall consider free plane networks. By assuming approximate (or arbitrary) coordinates for two of its points, the remaining coordinates can be computed. The two points for which coordinates are assumed form the computational base, or according to Baarda an S-base, BAARDA (1973). The coordinates with respect to the S-base are the estimates $\hat{x}$ with covariance matrix $C_{\hat{x}}$. It is well known that the coordinates and its covariance matrix are not invariant with respect to the choice of the computational base. The eigenvalues of $C_{\hat{x}}$ are also not invariant. The introduced measure of goodness of estimates as λ_{max}, λ_{max}-λ_{min}, $Tr(C_{\hat{x}})$ and $Det(C_{\hat{x}})$ cannot be used. If we use the same S-base for another design of the free network we obtain a covariance matrix which can now be compared with $C_{\hat{x}}$.

If the trace and the determinant of the covariance matrix are used as criteria for comparison of two different designs, we meet the same difficulties as was stated in the previous paragraph: the matrix with smallest trace doesn't guarantee that the variances of all possible functions computed with it are smaller that the one derived from the covariance matrix with greater trace. The same holds for the determinant criterion. The computation of the invariant maximum eigenvalue of the two covariance matrices can now be used to decide which one of the matrices is the best one.

In practice often only one single design of a network is available. Nowadays it is possible to compute an artificial covariance matrix for the coordinates by means of a covariance function. Such a matrix can be used as a criterion matrix i.e. "ideal" covariance matrix with which the covariance matrix of a designed network can be compared. The criterion matrix introduced by Baarda is defined with respect to an S-base so that a comparison with a covariance matrix computed in the same S-system is possible, BAARDA (1973). If the criterion matrix has no specified computational base, we have to apply an S-transformation. By This S-transformation the criterion matrix can be based on the computational base of the covariance matrix of the designed network. Let us denote the criterion matrix K.

The designed network with covariance matrix C is "at least as good" as the criterion matrix K if the general maximum eigenvalue of C and K $\mu_{max} \leq 1$.
If $\mu_{max} \leq 1$ the standard ellipses lie all within the standard ellipses derived from the criterion matrix (the converse is not true). It is also advantageous to compute the minimum general eigenvalue of C and K. As an average measure for the precision can serve

$$\bar{\mu} = \frac{1}{n} \sum \mu_i = \frac{1}{n} \mathrm{Tr}(K^{-1}C) \ .$$

The geometric mean of the general eigenvalue can also be used. As the general eigenvalues are invariant with respect to an S-transformation, they are just the quantities which can be used to qualify the precision of a geodetic network. It should be once more noted that the trace of the covariance matrix itself should be no longer used as a criterion, if one has agreed a criterion matrix.

Several types of criterion matrices do exist. The criterion matrices introduced by Grafarend and P. Meissl and K. Borre have no specified computational base. The criterion matrix of Alberda is a special case of the criterion matrix having Taylor-Karman structure, BAARDA (1979). By means of an S-transformation we can transform these matrices in a specified S-base, otherwise these matrices cannot be used to make a comparison of covariance matrices of coordinates.
We can only compare matrices if they are defined in the same S-system.

6. Precision of densification networks

In densification networks we meet the problem that for practical reasons often the coordinates of the control points are not to be altered. This can be realized by taking these coordinates as errorless in the adjustment. The covariance matrix of the coordinates resulting from this adjustment may not be used to characterize the precision of the network as the "noise" of the given

points has been neglected. A solution for this problem was already given by Baarda in 1956, see also BAARDA (1967). In order to obtain a more realistic covariance matrix, a matrix must be added, which is a function of the covariance matrix of the control points, see also MIERLO (1982).

The relative precision of the control points and the densification points, which can be derived from its "corrected" covariance matrix is in conformity with the adjustment procedure. A comparison with a given criterion matrix is now possible. The main problem is the determination of the covariance matrix of the control points. This matrix is generally unknown, but it is possible to construct a substitute matrix for the real covariance matrix. This can be done by analysing the network of higher order i.e. the method with which the coordinates of the given points are derived.

Until now often this "correction matrix" is neglected with the argument that the covariance matrix of the control points is badly known. Nevertheless a corrected covariance matrix must be preferred so that it is not possible to represent the precision of the network as better than is true.

7. Reliability of networks

The quality of a network can be described by two components: the precision and the reliability of the coordinates. The reliability of the coordinates depends on the way the observations are tested and improved for made errors. If the observations are not tested, than one cannot detect errors. In that case the reliability of the network is as bad as possible.

In spite of statistical tests, errors in the observations may be remain undetected. The effect of undetected errors on the coordinates is defined as the external reliability of a network. As one doesn’t know which observation can be possible wrong, one has to arrive at a convention concerning statements about reliability of networks. Baarda suggest to test only the "conventional" alternative hypotheses i.e. one particular observation is wrong and all others are undisturbed. These so-called outlier tests have shown their usefulness in practical geodetic networks.

In practice little attention has been paid to the choice of the significance level and the power of the test. If we assign a value for the power of the outlier test one can determine a boundary value for the (gross) error in the corresponding observation. The boundary values are an indication how well the observations can be checked. They form a conventional measure for the internal reliability, BAARDA (1968). In reasonably good networks one usually finds values of about 5 to 8 times the standard deviation of the observation variate in question (level of significance 0.001, power 80 %).

In practice it is far more important to know the effect of undetected errors on coordinates. The effect of a boundary value on the coordinates has been termed the external reliability. The caused distortion of the network by a wrong observation, the size of the boundary value, has to be measured. In a free network the computed coordinate differences $\nabla_i x$ derived from an error in observation l_i, are dependent on the used S-system. However, the weighted norm $\bar{\lambda}_i$

$$\bar{\lambda}_i = (\nabla_i x)^T C_{xx}^{-1} (\nabla_i x)$$

is invariant with respect to an S-transformation. Therefore one can use the weighted norm as a measure for the external reliability of a geodetic network. Baarda proposes as a global measure of the external reliability of a network the maximum value of $\bar{\lambda}_i$.

It is also possible to derive values of $\bar{\lambda}$ for partial networks. The distortion has mainly a partial character. For practice, the reliability measure has an interesting property. It can be shown that the standard deviation of an arbitrary function of the coordinates σ_F multiplied with $\bar{\lambda}_{max}$ is greater than or equal the maximum effect of boundary values on it

$$|\nabla F| < \sigma_F \sqrt{\bar{\lambda}_{max}} \quad .$$

Strain analysis techniques have been applied to study the effect of inconsistent observations and constraints in geodetic networks, VANÍČEK et al. (1981). The examples which are given of the application of this interesting technique give

rise to questions, because the result indicates that the results are not invariant with respect to an S-transformation. It is also possible to apply this technique to study the (local) distortion of a network caused by a "gross" error the size of the boundary value. Only if the S-base is lying in the (local) distorted network the elements of the strain tensor are invariant with respect to S-transformations. It should be stated that these computations have not been done yet. A comparison with the measure for reliability of a partial network is being in preparation. If the strain analysis technique is useful to quantify the reliability of a geodetic network is still an open question. More research into this matter is required.

8. Sensitivity of a net

In the field of the analysis of geodetic deformation measurements the sensitivity of a given net to critical deformations plays an essential role. The deformation of an object is derived from the comparison of two or more sets of coordinates of the same network which is a representation for the object. Assuming only two epochs, with coordinate sets $\hat{x}_1$ and $\hat{x}_2$, the difference vector d is given by

$$d = \hat{x}_2 - \hat{x}_1$$

and the corresponding covariance matrix is

$$C_d = C_1 + C_2 \ .$$

Assuming $C_1 = C_2 = C$, we have $C_d = 2C$. If the deformation model can be written as

$$d = k \sqrt{\lambda_j}\ s_j$$

where s_j is the normed eigenvector of C corresponding to the eigenvalue λ_j, it can be proved that a boundary value for d increases in direct proportion to the eigenvalues of C. Expected deformation in the direction of the eigenvector of the maximum eigenvalue are therefore worst detectable. Among these very special

deformation models, the sensitivity of the net in the direction of the eigenvector of λ_{max} is very bad, PELZER (1980). However, in free networks the eigenvector and the eigenvalues are not invariant with respect to an S-transformation. Therefore, the use of the eigenvectors as a tool for the design of relative deformation measurements must be considered with care. Moreover, it can be possible, that the effect of undetected errors in the observations has an influence on the deformation model which is more serious than the "eigenvector deformation" model, MIERLO (1981).

9. Conclusion

The quality of a geodetic network must be characterized by invariant quantities with respect to S-transformations. For that reason the general eigenvalues of the covariance matrix of the coordinates and an adopted criterion matrix can be used to measure the precision. The reliability can be characterized by a distortion parameter which measures the distortion of the network due to a gross error which has the size of the boundary value. The level of significance and the adopted power of the test for the computation of the boundary values must be specified as well as the covariance function of the criterion matrix.

REFERENCES *

ALBERDA, J.E., 1980: A review of analysis techniques for engineering survey control schemes. Paper presented to the Industrial and Engineering Survey Conference, London, 2-4 Sept 1980.

BAARDA, W., 1967: Statistical Concepts in Geodesy. Netherlands Geodetic Commission. Publications on Geodesy, New Series, Vol. 2, No. 4, Delft.

BAARDA, W., 1968: A testing procedure for use in geodetic networks. Netherlands geodetic Commission. Publications on Geodesy, New Series, Vol. 2, No. 5, Delft.

* The references are kept to a minimum: they are only meant to give the reader an idea where to look for further information and more references.

BAARDA, W., 1973: S-Transformation and Criterion matrices. Netherlands Geodetic Commission. Publications on Geodesy, New Series, Vol. 5, No. 1, Delft.

BAARDA, W., 1979: Measures for the accuracy of geodetic networks. In "Optimization of design and computation of control networks" edited by F. Halmos and J. Somogyi, pp. 419-426, Akadémiai Kiadó, Budapest.

BORRE, K., 1978: Error propagation in absolute geodetic networks, a continuous approach. Studia Geoph. et Geod. 22 (1978), pp. 213-223.

GRAFAREND, E., HEISTER, H., KELM, R., KROPFF, H., SCHAFFRIN, B., 1979: Optimierung geodätischer Messoperationen. Herbert Wichmann, Karlsruhe 1979.

MIERLO, J. van, 1980: Free network adjustment and S-transformation. Veröff. d. DGK, Reihe B, Heft Nr. 252, München 1980.

MIERLO, J. van, 1981: A testing procedure for analysing geodetic deformation measurements. In "Proceedings of the II Int. Symp. of Deformation Measurements by Geodetic Methods", Bonn, Sept 25 to 28, 1978, pp. 321-354.

MIERLO, J. van, 1982: Design and adjustment of densification networks. In "Proceedings 7th Conference of Southern African Surveyors", Johannesburg, 25-29 January 1982, edited by I.B. Watt.

PELZER, H., 1980: Geodätische Netze in Landes- und Ingenieurvermessung. Konrad Wittwer, Stuttgart 1980.

VANÍČEK, P., THAPA, K., SCHNEIDER, D., 1981: The use of strain to identify incompatible observations and constraints in horizontal geodetic networks. Manuscripta Geodaetica, Vol. 6 (1981), pp. 257-281.

PRINCIPAL COMPONENT ANALYSIS AND GEODETIC NETWORKS
- SOME BASIC CONSIDERATIONS -

Wolfgang NIEMEIER *
Calgary, Alberta, Canada

ABSTRACT

As additional criteria for the diagnosis of geodetic networks, principal component analysis (PCA) is introduced and its geometrical and statistical properties are outlined. These criteria show geometrically the correlation between variates, i.e. information that cannot always be derived from other common criteria for precision. As the examples will show, PCA yields a geometrically clear insight into the network behaviour and it is therefore ideally suited to the description of certain global properties of a network, especially one designed for monitoring.

* On leave from the University of Hannover, Geodetic Institute, Nienburger Str. 1, 3000 Hannover, Federal Republic of Germany

1. Introduction

In the past two decades the quality of geodetic networks has been defined and discussed in numerous publications. It is nowadays a common procedure to refer to the goodness of a network in terms of its precision (quality of design) and reliability (quality of conformance), whereas the costs are usually considered as limiting factors only.

In this paper an additional and perhaps useful method, the principal component analysis (PCA), will be discussed. In addition to previous papers on this topic (Augath et al. 1979, Dupraz & Niemeier 1979, Leonhard & Niemeier 1980, Niemeier & Rohde 1981, Pelzer 1976, 1979), where only the applications of PCA were considered, here the fundamental ideas und principles of PCA are stressed. PCA is widely used in multivariate statistics to show certain properties of higher dimension covariance matrices in lower space. To familiarize the reader with this method the initial sections of this paper outline both geometrical and statistical aspects of PCA, and first applications and interpretations to the covariance matrix K_{xx} of coordinate variates, for simple networks, are discussed.

It turns out that the main feature of PCA is the portrayal of information not included in either the point confidence ellipses or the relative confidence regions. This additional information concerns the correlation of variates. It will be shown that for highly correlated variates the 1st principal component is a representation of most of the variance in a one-dimensional space, it points out the correlation of parameter estimates and by this may give a geometrically clear insight into weak zones of a network, an insight, which cannot be reached by commonly used analysis techniques.

In analyzing results of multiple measuring epochs, or in the definition of the quality of a net as a monitoring network, PCA is of further use. This will be shown by relating this method to sensitivity analysis. Especially the effect will be discussed, that results of multiple observations tend to exhibit its maximum variation in the direction of the main principle axis.

2. Geometrical and Statistical Aspects of Principle Component Analysis (PCA)

2.1 Introduction to PCA

All information concerning the precision of a geodetic net is contained in the covariance matrix K_{xx} of the parameter estimates $\hat{x}$. A graphical

representation for a (regular) (n,n)-covariance matrix K_{xx} can be found from the quadratic form

$$(x - \hat{x})^T K_{xx}^{-1} (x - \hat{x}) = c \qquad (1)$$

which is the equation of a group of ellipses (ellipsoids) centered in $\hat{x}$, where the group parameter is c and is related to the assumed level of significance by $c = \chi^2_{n,1-\alpha}$.

From linear algebra it is known that for any quadratic matrix a decomposition into eigenvalues λ_i and eigenvectors s_i can be found, which may be given for the covariance matrix by

$$K_{xx} = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ s_1 & s_2 & \cdots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \\ 0 & & & \lambda_n \end{bmatrix} \begin{bmatrix} & s_1^T & \\ \cdots & \cdots & \cdots \\ & s_2^T & \\ \cdots & \cdots & \cdots \\ & \vdots & \\ \cdots & \cdots & \cdots \\ & s_n^T & \end{bmatrix} \qquad (2)$$

The principle axes now are defined as the eigenvectors s_i and the 1st (main) principal component (PC) is given by

$$p_1 = s_1 \sqrt{\lambda_1} \qquad (3)$$

where λ_1 is considered to be the maximum eigenvalue of K_{xx} and s_1 its associated eigenvector.

Now it turns out that a one-dimensional representation of K_{xx} by the 1st PC is only meaningful, if the eigenvalue λ_1 accounts for an essential portion of the whole variance of K_{xx}, given by $\sum_{i=1}^{n} \lambda_i$. This means a PCA gives a geometrically clear insight into weak zones of the net only if λ_1 represents at least 40-60% of $\sum_{i=1}^{n} \lambda_i$. Many applications can be found in practice however, where this percentage is reached.

2.2 Geometric Properties of the Eigenstructure of Covariance Matrices

One line of thought, leading to the eigenstructure of the covariance matrix and hence to the PCA, comes from linear algebra (e.g. Green & Carroll 1976, Strang 1976): A vector space can be represented by a set of original orthogonal basis vectors e_i or any other – for simplicity also orthogonal - basis vectors f_i, where the transformation matrix L between the two basis vector systems is an orthogonal matrix, i.e.

$$LL^T = L^T L = I\,, \qquad (4)$$

and the rows of L consist of the direction cosines of the new basis vectors f_i in terms of the original basis vectors e_i, i.e. L represents a simple

rotation of the coordinate system. If a matrix T is given in the original basis, a matrix T^o, which is now referred to the new basis vectors f_i can be computed by

$$T^o = L\,T\,L^T \tag{5}$$

The objective now is to find a basis vector transformation such that T^o takes on a particular simple form, such as a stretch, i.e. involving differential stretching or contracting of points in the direction of the given axes. This is meaningful, as it is known in linear algebra (e.g. Green & Carroll 1976) that any nonsingular matrix transformation with real-valued entries can be uniquely decomposed into the composite transformation (a) rotation - stretch – rotation or (b) rotation – reflection – stretch – rotation.
To make these transformations more clear, in Figure 1 an arbitrary transformation by a (2,2)-matrix T is depicted, where x are the original coordinates and x* the image coordinates.

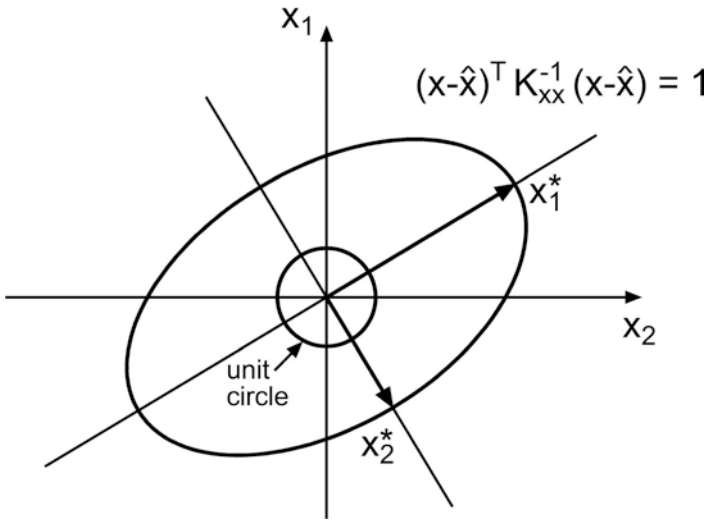


Fig. 1: Change of basis vectors of a matrix representing a quadratic form (after Green & Carroll 1976)

To find now the specific basis vector transformation, a requirement is, that for some preimage vectors

$$x = \begin{bmatrix} x_1 \\ \\ x_2 \end{bmatrix} \tag{6}$$

the image vector (see Figure 1)

$$x^* = \begin{bmatrix} x_1^* \\ \\ x_2^* \end{bmatrix} \tag{7}$$

has to have the same direction as the preimage vector. This means, that for some particular vectors x the direction is maintained, i.e. only the effect

of stretch exists:

$$x^* = \lambda x \tag{8}$$

In general the transformation T can be applied to x:

$$x^* = T x \tag{9}$$

The substitution of eq. (8) in eq. (9) then results in:

$$T x = \lambda x \tag{10}$$

which can be rearranged to

$$T x - \lambda I x = 0 \tag{11}$$

$$(T - \lambda I) x = 0 \tag{12}$$

To obtain solutions $x \neq 0$, the determinant of eq. (12) must be zero, which leads to the characteristic equation of T:

$$|T - \lambda I| = 0 \tag{13}$$

As is well-known, the roots of this equation are the eigenvalues λ_i of T, and their associated vectors from eq. (12) are the eigenvectors x_i.

Two central results can be obtained from this process, which are of some importance for the geometric interpretation of PCA (see section 2.1):

(i) the eigenvalues λ_i indicate the magnitude of the stretch (or stretch followed by refraction),

(ii) the eigenvectors indicate the new directions (basis vectors f_i) along which the stretching takes place.

Note that the elements of the eigenvectors are direction cosines for the new basis vectors, but they are expressed in terms of the original basis!

2.3 Maximizing Linear Composites

A second way that leads to the PCA is via the statistical point of view (e.g. Giri 1977), Gnanadesikan 1977, Kendall 1975). The error or confidence ellipses, given by eq. (1), can be interpreted as horizontal sections through the 2-dimensional normal distribution of the parameter vector x. This means, that if no external influences exist, repeated experiments (in geodetic networks, this means multiple observations epochs) will give estimates which are spread in a similar manner to the dots in Figure 2, and which are in correspondence to the confidence ellipse.

The statistical concept bow is to replace the two variables x_1 and x_2 by a linear composite

$$z = s_1 x_1 + s_2 x_2 \tag{14}$$

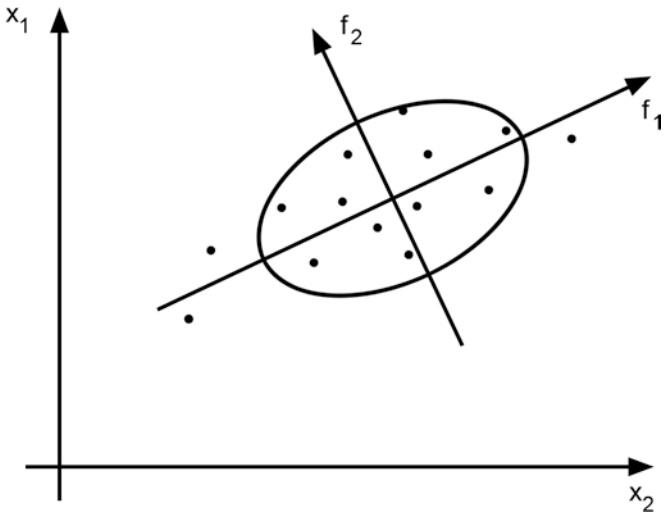


Fig. 2: Illustration of principal components with bivariate data (after Gnanadesikan 1977).

with the requirement that the variance of this linear composite is maximized, subject to $s^T s = 1$.

$$\mathrm{Var}(z) = \max \qquad \forall\, s \tag{15}$$

Geometrically, the different estimates are maximally separated along the linear composite.

The motivation for pursuing this line of thought is the desire to replace two or more correlated variables with a single linear composite that accounts for as much as possible of the variation shared by the component variables. Of course, this process will yield only satisfactory results, if a high correlation exists. For geodetic networks this implies that a somewhat inhomogeneous geometry is present.

To find a new basis vector s with the property of maximizing the variance requires the maximization of the expression

$$s^T K_{xx}\, s \tag{16}$$

subject to the normalization constraint $s^T s = 1$. This problem is standard in optimization and it can be stated as follows:

$$F = s^T K_{xx}\, s - \lambda(s^T s - 1) \tag{17}$$

where λ is an additional unknown, the Lagrange multiplier. To maximize F, the partial derivative

$$\frac{\partial F}{\partial s} = 2(K_{xx} s - \lambda s) \tag{18}$$

is set equal to zero. Dividing both sides by 2 and factoring out s leads to

$$(K_{xx} - \lambda\, I)\; s = 0 \tag{19}$$

which is equivalent to eq. (12). Again, a value for λ is sought, that satisfies the characteristic equation

$$|K_{xx} - \lambda I| = 0 \tag{20}$$

However, for a (n,n)-covariance matrix K_{xx} one obtains n eigenvalues λ_i. Only the largest eigenvalue λ_1 will maximize F. The eigenvalue s_1, corresponding to λ_1 is the desired linear composite (eq. (14)) and by this the new basis vector, which gives the maximum variance achievable along one dimension. The eigenvector s_2, corresponding to the second largest eigenvalue λ_2, is the direction of that linear composite, which gives the maximum achievable variance orthogonal to s_1.

2.4 Summary of Geometrical and Statistical Aspects of PCA

In geometric interpretation the 1st PC

$$p_1 = s_1 \sqrt{\lambda_1}$$

is a one-dimensional representation of the major semi-axis of the n-dimensional confidence hyperellipsoid. $\sqrt{\lambda_1}$ is the length of this major semi-axis and the elements of s_1 are the direction-cosines of this projection onto the original basis vectors.

In the statistical interpretation the maximum possible separation between the variables is found in the direction of s_1, and that the maximum variance λ_1 is in this direction. For geodetic networks this means, the composite with s_1 is worst determined among all possible functions of x and the differences of repeated coordinate determinations tend to lie in this direction!

3. Application and Interpretation of PCA in Geodetic Networks

3.1 Two-Dimensional Examples

For a simple and geometrically straight forward application of the PCA to geodetic networks this section starts with 2-dimensional parameter vectors.

In the first example a plane intersection from two reference points will be considered. As geometric representation of the (2,2)-covariance matrix K_{xx} the standard (Helmert) error ellipse

$$(x - \hat{x})^T K_{xx}^{-1} (x - \hat{x}) = 1 \tag{21}$$

for this example is depicted in Figure 3.

As discussed in the previous sections, the 1st principal component is given by

$$p_1 = s_1 \sqrt{\lambda_1} = \begin{bmatrix} s_{11} & \sqrt{\lambda_1} \\ s_{12} & \sqrt{\lambda_2} \end{bmatrix} \tag{22}$$

which is the projection of the major semi-axis a_1 onto the orthogonal coordinate axes. The length of the major axis is

$$a_1 = \sqrt{\lambda_1} \tag{23}$$

and its direction cosines are given by the eigenvector

$$s = \begin{bmatrix} s_{11} \\ s_{12} \end{bmatrix} \tag{24}$$

All these relations are known from standard textbooks of adjustment calculus: at least for single point determinations the PCA is completely in correspondence with common practice!

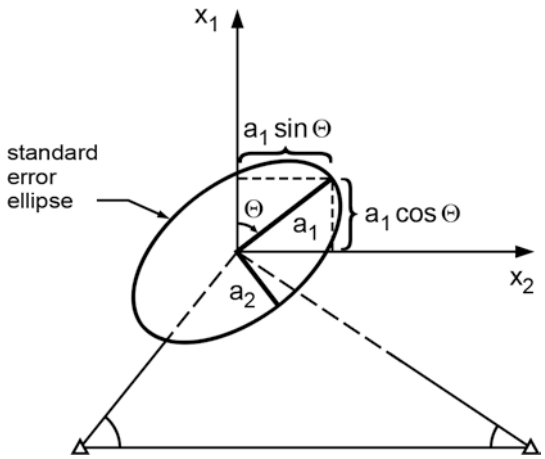


Fig. 3: Error ellipses and principal components for a point intersection.

As a second example, graphically given in Figure 4, a small levelling network of only 3 points is selected (after van Mierlo 1981). Point H_1 in the figure is held fixed.

In deviation from the common rules, as representation of the covariance matrix of the two parameters, belonging to different points, in panel II of Figure 4 a two-dimensional error ellipse is constructed, which shows very clearly the correlation of $\rho = 0.5$ between the two variates. In addition to one-dimensional error or confidence regions, which would be applied usually to depict the precision in this net, the elements of the 1st PC in panel III of this figure represent both the variance and the correlation between both variates!

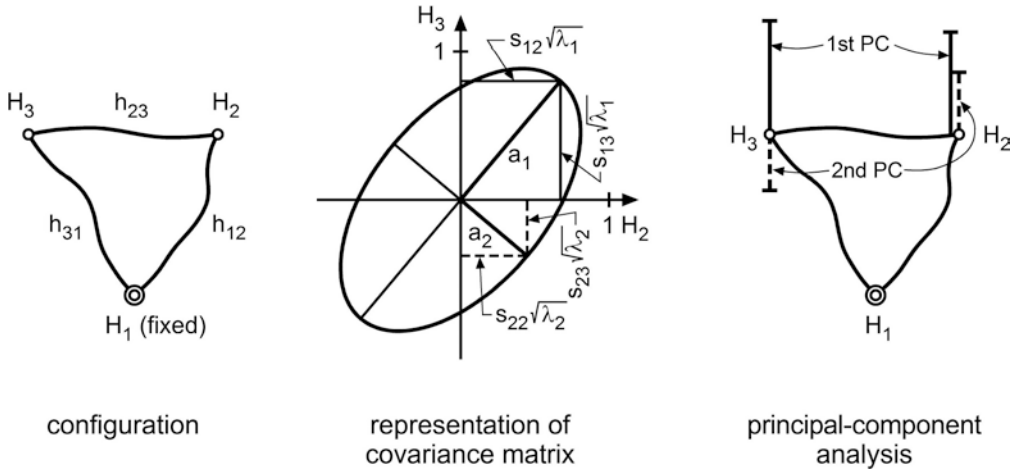


Fig. 4: Principal components for a simple levelling net.

3.2 Horizontal Networks

In Figure 5 and Figure 6 the point confidence ellipses and the components of the 1st PC are depicted for two horizontal networks with direction and distance observations. To both networks and inner constraint adjustment is applied.

Starting with the regular chain of diagonals of Figure 5, it is at first surprising that the areas of the confidence ellipses are not a minimum for the central points, as is often expected to be a result of an inner constraint adjustment. The smallest ellipse areas are found for points 20-25% apart from both ends, i.e. near the so-called Bessel points, if compared with the bending of a beam.

As λ_1 accounts for 57% of the trace of K_{xx}, the PCA is applicable here. The 1st PC, given by the vectors in Figure 5, depict the strong correlation between neighboring points; information that cannot be obtained from point confidence ellipses and is essentially eliminated in relative ellipses. In addition, the prevailing transversal direction of this PC can be interpreted as a weakness of the configuration in this direction. In relation to section 2.3 this means that the parameter estimates may be falsified preponderant in this direction. The increasing sizes of the confidence ellipses for the central and border points fit well to this 1st PC. Only from a PCA can this effect be adequately explained.

A second example is given in Figure 6 where the confidence ellipses and the direction of the 1st PC are depicted for a real triangulateration network from Venezuela. In contrast to the results shown in Figure 5, where the point components of the 1st PC are dominantly in transversal direction, here these components lie mostly in longitudinal direction. To discuss this effect, it has to be mentioned that compared with the regular chain of Figure 5 here the precision of the distance observations relative to the directions is rather poor.

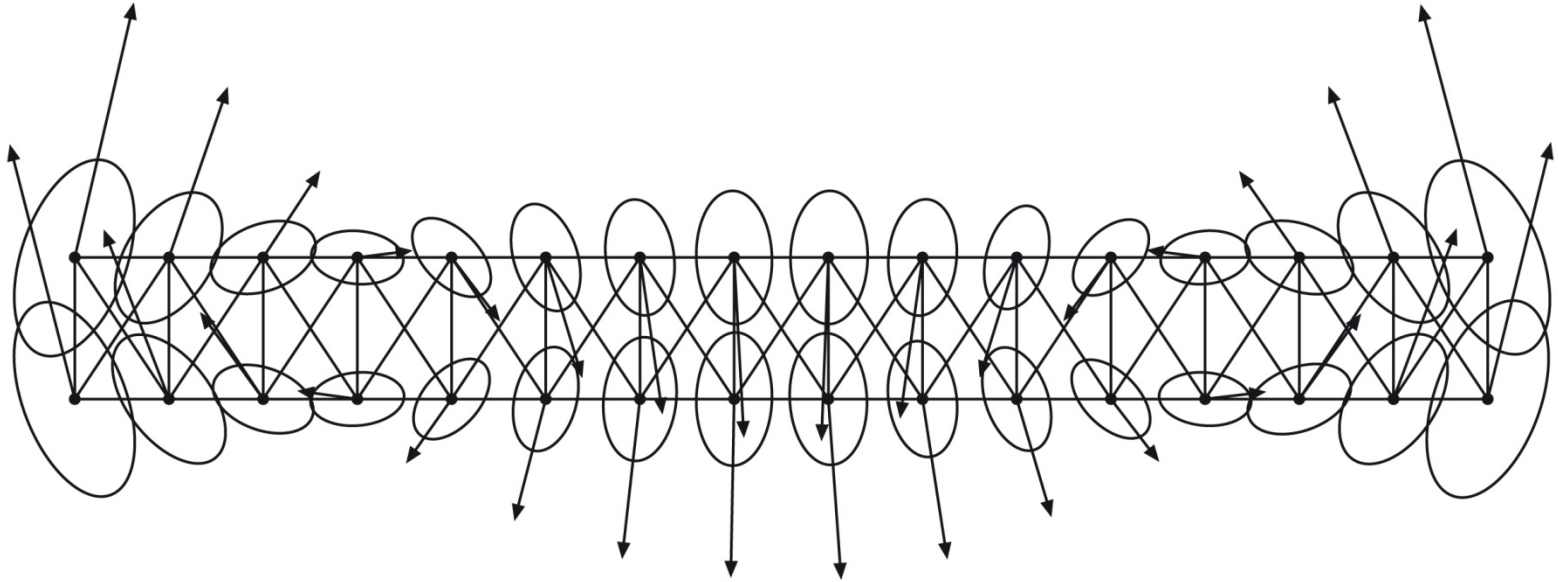


Fig. 5: Confidence ellipses and 1st principal component of a regular horizontal network (Gutberlet 1982).

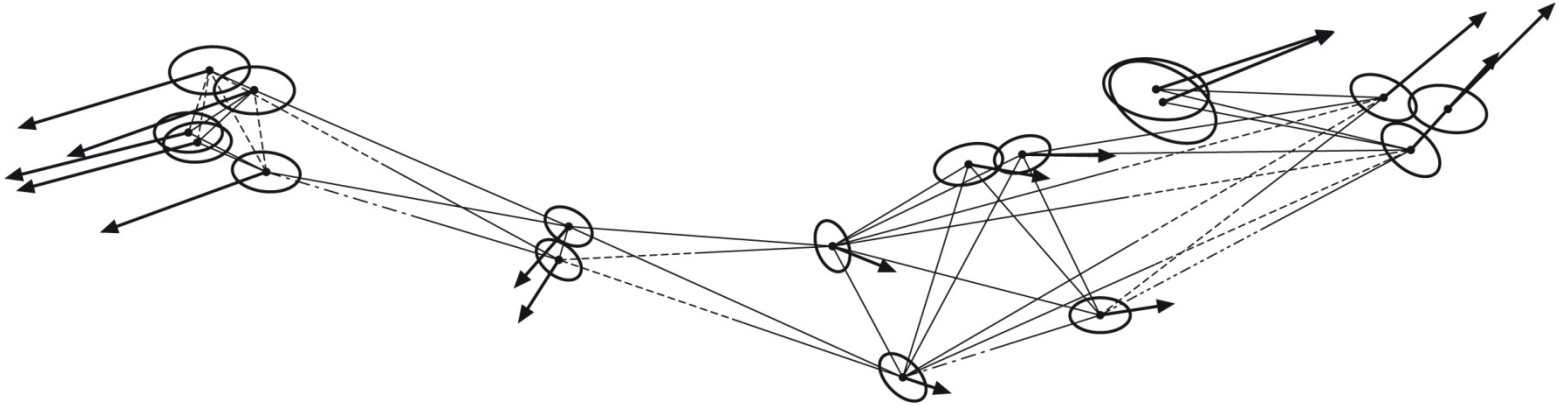


Fig. 6: Confidence ellipses and 1st principal axis of a horizontal network in Venezuela.

One interpretation for this phenomenon may be that the geometry of a network is more or less responsible for the existence of critical (weak) directions, but the character of the weakness is determined by the relative precision within and between different kinds of observations. A more detailed discussion of this interesting aspect is not within the scope of this paper.

3.3 Simulation Study

To ensure the behaviour of networks, with regard to aspects of PCA, simulation studies have been carried out previously (Dupraz & Niemeier 1979, Schmidt 1981). In spite of the critical datum fixation, the example of Dupraz & Niemeier (1979) is presented here again, see Figure 7.

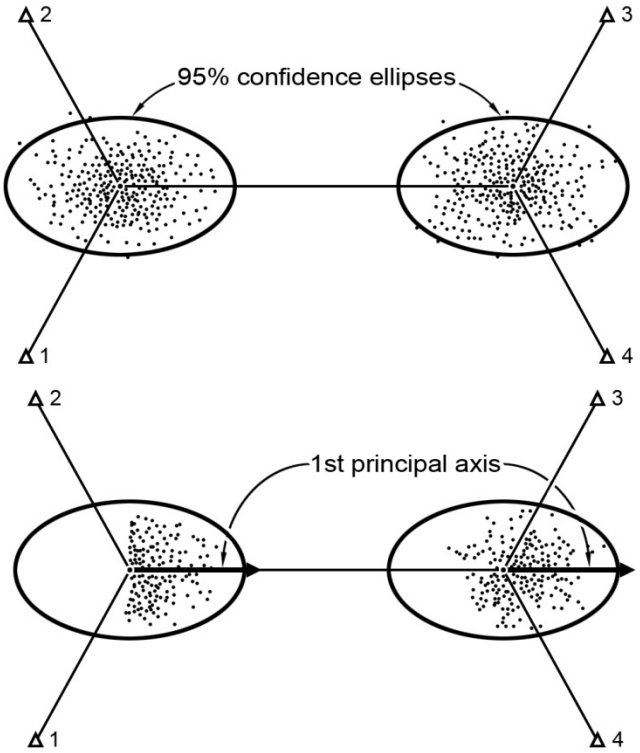


Fig. 7: Simulation study with 1000 adjustments and its relation to the 1st principal component.

In this simulation study of a horizontal trilateration net the true distances were at first derived from the given (true) coordinates. The observations for 1000 adjustments were then created using a random number generator, which gives normal distributed pseudo random numbers with preset standard deviation. The confidence ellipses correspond to a 95% confidence level, and only a few adjustment results exceeded this confidence region. The correlation of the y-values is rather high ($\rho = 0.67$), and to show this numerically, a constraint simulation was carried out: Only the adjustment results which had positive Δy-values for the left point are shown in panel II of Figure 7. It can be derived from the figure that about 70% of the adjustment results for the right point have also positive Δy-values, which is of course due to the high correlation. The 1st PC, which is also

depicted in panel II (without scale) again indicates very clearly the similar behaviour of both points, i.e. the correlation effect.

4. Principal Component and Sensitivity Analysis

4.1 General Remarks

As derived in section 2.3, the highest achievable variance in one dimension is in the direction of s_1. As shown in Figures 2 and 7 this means, the coordinate estimates of different epochs will vary most in the direction of the 1st PC. Now, if a network has a really weak zone, i.e. if λ_1 counts for an essential part of $tr(K_{xx})$, coordinate differences may reach relatively large values and may show a regular pattern (due to their high correlation), even if no real movements took place.

In practice this effect may be very critical, if one tries to develop deformation models out of the results of repeated network determinations. This so-called operational approach is rather frequently applied in geodynamics: One assumes that coordinates, that show a certain regular pattern, are related to real – perhaps not yet significant – crustal movements. For the North-Sea-Coast-Levelling-Net (Augath et al 1979) and the Trans-Canada-Levelling-Line (Lachapelle 1979, Leonhard & Niemeier 1980) an almost complete correspondency between coordinate differences and the 1st PC was found.

4.2 Review of Sensitivity Analysis

To establish a measure for the quality of a net as a monitoring network, the sensitivity analysis (Pelzer 1971, 1976) was established.

Assuming an identical configuration for two monitoring epochs from separate adjustment, parameter estimates $\hat{x}_1$ and $\hat{x}_2$ and their covariance matrices $K_{xx}^{(1)} = \sigma_0^2\, Q_{xx}^{(1)}$, $K_{xx}^{(2)} = \sigma_0^2\, Q_{xx}^{(2)}$ are obtained. The global congruency test of the null hypothesis

$$H_0 = E\{\hat{x}_1\} = E\{\hat{x}_2\} \tag{25}$$

is given by the test statistics (Pelzer 1971)

$$F = \frac{d^T Q_d^+ d}{R(Q_d)\,\sigma_0^2} \tag{26}$$

where

$$d = \hat{x}_2 - \hat{x}_1 \tag{27}$$

$$Q_d = Q_{xx}^{(1)} + Q_{xx}^{(2)} \tag{28}$$

If H_o holds, $\bar{F}$ follows the central F-distribution. For an alternative hypothesis

$$H_A: E\{d\} = d_A \neq 0 \tag{29}$$

the statistics $\bar{F}$ is distributed according to the noncentral F-distribution with the noncentrality parameter

$$\omega = \frac{d_A^T Q_d^+ d_A}{\sigma_o^2} \tag{30}$$

Now the power of a test is defined as the probability $\gamma = 1-\beta$ that d_A will lead to a rejection of H_o at a level of significance α. For specified α and β a critical value for the noncentrality parameter ω_o can either be computed as a function of α, β and $R(Q_d)$, or taken from nomograms (Baarda 1968). The line of thought (corresponding to the reliability concept of Baarda) now is, that a specific d_A is named detectable, if its ω exceeds the critical value ω_o:

$$\omega > \omega_o \qquad \Rightarrow d_A \text{ detectable} \tag{31}$$

The importance of this inequality is that rather frequently the relative magnitude of movements of interest, i.e. a form vector g, is known and thus an assessment can be made in order to determine the just-detectable deformations according to this model. In this case H_A is given by

$$H_A: d_A = c \cdot g \tag{32}$$

where c is a scalar to be determined. The deformation vector d_A now is detectable if

$$\omega = \frac{(c\,g)^T Q_d^+ (c\,g)}{\sigma_o^2} > \omega_o \tag{33}$$

which leads to the inequality for the scalar factor c:

$$c \geq \sigma_o \sqrt{\frac{\omega_o}{g^T Q_d^+ g}} \tag{34}$$

4.3 Sensitivity in Direction of the 1st PC

As was outlined in the previous sections, multiple parameter estimates tend to lie in the direction of the 1st principal axis, i.e. the eigenvector s_1. This is also the worst determined direction in the entire network.
It might now be of some interest, to compute the minimum detectable

movements in direction of s_1. To do this in the concept of sensitivity analysis s_1 is used as a form vector in eq. (32):

$$H_A: d_A = c_1 \cdot s_1 \tag{35}$$

Assuming $Q_d = 2\,Q_{xx}$, for this special case the quadratic form in eq. (34) is

$$g^T Q_d^+ g = \frac{1}{2} s_1^T S D^{-1} S^T s_1 \tag{36}$$

if S is the matrix of eigenvectors s_i and D the matrix of eigenvalues λ_i of Q_{xx}. As all products

$$s_1^T s_j = 0 \qquad \forall\, j \neq 1 \tag{37}$$

eq. (36) takes on the very simple form

$$g^T Q_d^+ g = \frac{1}{2\lambda_1} \tag{38}$$

which gives, substituted in eq. (36), the final inequality

$$c_1 \geq \sigma_o \sqrt{2\,\omega_o\,\lambda_1} \tag{39}$$

which has to be fulfilled by the scalar factor, to lead to detectable movements in direction of s_1.

To give an example for this combination of PCA and sensitivity analysis, in Figure 8 some results from Leonhard & Niemeier (1980) for the analysis of the Trans-Canada-Levelling-Line (TCL) are depicted. As reported by Lachapelle (1979), between the old (1920-1930) and the new (about 1960) observations of this line a discrepancy of 1.96 m was found. Also shown in the figure are the elements of the minimum detectable movements according to eq. (39), computed with preset error probabilities of $\alpha = 0.05$ and $\beta = 0.10$.

Both computations are based on a datum fixation in Halifax, and, as a first result, continuously increasing values for the discrepancies and the critical amount of movements can be found. This close correspondence ensures the tendency of repeated measurements to lie in the direction of s_1. In this case computations for the minimum detectable movements were made with an assumed $\sigma_0^{(1)} = 2\,\text{mm}/\sqrt{\text{km}}$, for the Canadian Adjustment 1928 and $\sigma_0^{(2)} = 4\,\text{mm}/\sqrt{\text{km}}$, for the observations in the new Canadian Vertical Framework (Lachapelle 1979). For points at the Pacific Ocean, the minimum detectable height differences according to $\sigma_0^{(1)}$ are 1.3 mm and according to $\sigma_0^{(2)}$ they rise to 2.6 mm. The conclusion can be drawn that without any systematic effect most of this very large discrepancy may be just an effect of the extreme network geometry and the levelling precision.

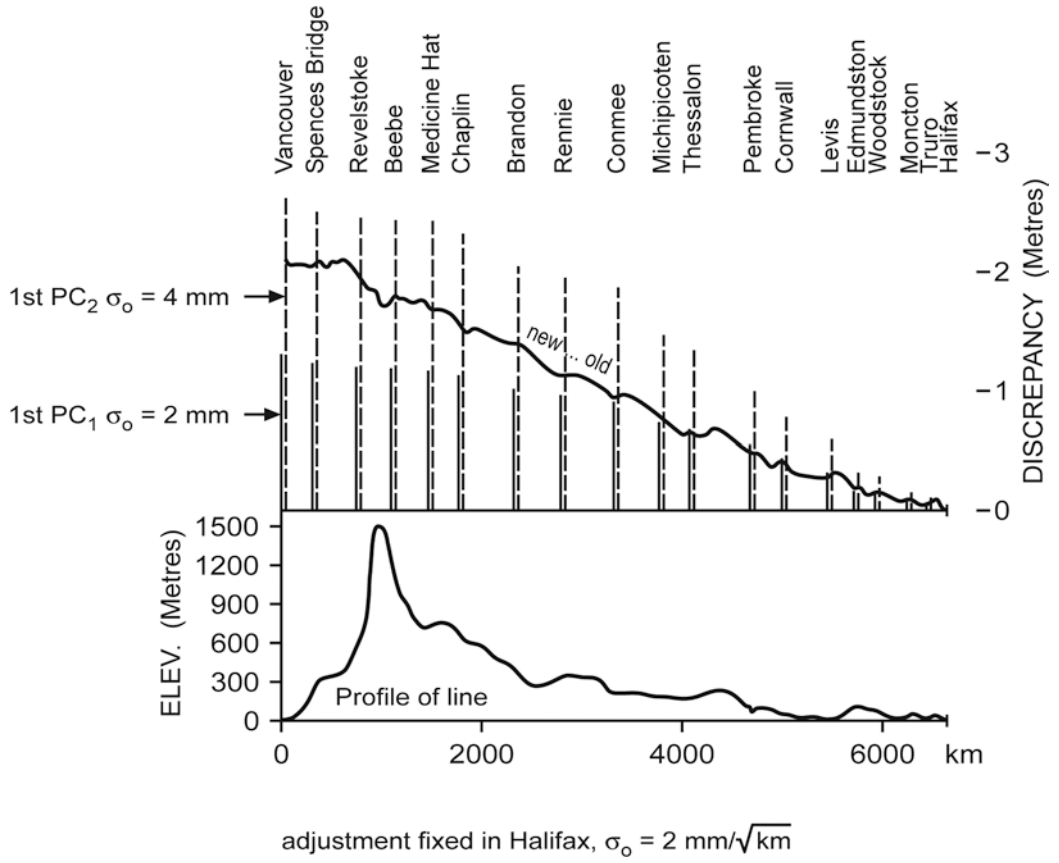


Fig. 8: Computed discrepancy for the Trans-Canada-Levelling Line and its relation to the 1st principal component.

5. Conclusion

Principal Component Analysis can be usefully applied in practice to geodetic networks. The geometrical and statistical properties of PCA allow the detection of possible weak zones of a network and present a geometrically clear picture of correlations between network parameters. This information cannot be provided by other criteria commonly used in network theory. PCA may therefore be a useful additional tool to understand the properties and the behaviour of geodetic networks.

Not included in this paper is a discussion of the datum dependency of the covariance matrix and the resultant PCA (van Mierlo 1981). Preliminary results show that a PCA applied to the so-called inner covariance matrix shows the point correlations in a most correct manner, and therefore the inner constraint adjustment may have some optimality properties with respect to PCA.

6. Acknowledgement

For their useful discussions and suggestions the author wishes to thank Dr. C.S. Fraser and Prof. W.F. Teskey. The financial support provided by a grant from the National Science and Engineering Research Council of Canada is gratefully acknowledged.

REFERENCES

AUGATH, W., NIEMEIER, W., PELZER, H., 1979: On the Significance of Measured Height Differences in Northern Germany. AVN, 82 (1979), pp. 405-412.

BAARDA, W., 1968: A Testing Procedure for Use in Geodetic Networks. Neth. Geod. Comm., Publ. on Geodesy, New Series, Vol. 2, No. 5, Delft.

BAARDA, W., 1977: Measures for the Accuracy of Geodetic Networks. IAG-Symp. on Optimization of Design and Computation of Control Networks. Sopron/Hungary.

DUPRAZ, H., NIEMEIER, W., 1979: Un critère pour l'analyse des reseaux géodésique de contrôle. Mensuration, Photogrammétrie, Genierural, 77, (1979), pp. 70-76.

GIRI, N.C., 1977: Multivariate Statistical Inference. Academic Press, New York, 1977.

GNANADESIKAN, R., 1977: Methods for the Statistical Data Analysis of Multivariate Observations. John Wiley & Son, New York 1977.

GREEN, P.E., CARROLL, J.D., 1976: Mathematical Tools for Applied Multivariate Analysis. Academic Press, New York 1976.

GUTBERLET, M., 1982: Untersuchung geodätischer Netze durch Spektralanalyse der Kovarianzmatrix des Koordinatenvektors. Diplomarbeit, Universität Hannover, 1982, unpublished.

KENDALL, M., 1975: Multivariate Analysis. Charles Griffin, London 1975.

LACHAPELLE, G., 1979: Redefinition of National Vertical Geodetic Networks. The Canadian Surveyor, Vol. 33 (1979), pp. 273-282.

LEONHARD, T., NIEMEIER, W., 1980: Diagnosis of Trans-Canada-Levelling-Networks by Principal Component Analysis. Proc. 2nd Int. Symp. on Problems Related to the Redefinition of North American Vertical Networks, Ottawa, 1980, pp. 485-503.

van MIERLO, J., 1981: Einige Bemerkungen zur Sensitivitätsanalyse, personal communication, February 10, 1981.

NIEMEIER, W., 1982: Design, Diagnosis and Optimization of Monitoring Networks in Engineering Surveying. Invited Paper, Centennial Convention of CIS, Ottawa, 1982.

NIEMEIER, W., ROHDE, G., 1981: On the Optimization of Levelling Networks with Respect to the Determination of Crustal Movements. IAG-Symposium on Geodetic Networks and Computations, Munich, 1981.

PELZER, H., 1971: Zur Analyse geodätischer Deformationsmessungen. Deutsche Geod. Komm., Series C, No. 164, Munich 1971.

PELZER, H., 1976: Über die statistischen Eigenschaften der Ergebnisse von Deformationsmessungen. Proc. VII. Int. Kurs für Ingenieurvermessungen hoher Präzision, Darmstadt, 1976.

PELZER, H., 1979: Some Criteria for the Accuracy and the Reliability of Networks. IUGG-General Assembly, Canberra, 1979.

SCHMIDT, W., 1981: Simulationsstudien zum Kriterium des wesentlichen Eigenvektors für Nivellementslinien. Diplomarbeit, Universität Hannover, 1981, unpublished.

STRANG, G., 1976: Linear Algebra and its Applications. Academic Press, New York 1976.

A NEW METHOD OF LAND SURVEYING NETWORKS OPTIMIZATION

Tosa NINKOV, M.Sc.
Beograd, Yugoslavia

ABSTRACT

The work illustrates a new method of determination of optimum second order project. This method is based on defining optimum weights of land surveying network elements - modifying them, in relation to initial solution, in proportion to their impact onto relevant aim- of $\lambda_{max} \rightarrow$ min forms. Finally, the method is illustrated through optimization of elementary levelling network.

0. Foreword

In spite of the fact that in course of the last few years there has been a remarkable development in many methods of determination of second order optimum project /to both strategies of solving the problem Grafarend /1// - problem has not been fully solved. This because the techniques applied had an imperfection or two which made practical usage not feasible.

The work is indicative of a new approach to determining optimum project of second order to first strategy, i.e. effected is minimization of the defined aim function $\lambda_{max} \rightarrow$ min with meeting specific conditions on unknown weights.

1. Defining the Problem

As defined by G. Schmitt /2/, in the process of determining optimum project of second order configuration matrix $\underline{A}$ and correlation $\underline{Q}_x$ are firm, i.e. unchangeable, while matrix $\underline{P}$ has changeable parameters. Should problem of optimization be solved to first strategy, aim function and limiting condition under general condition can be defined in the following mode:

$$F(x) \rightarrow \text{extremum} \qquad /1/$$

and

$$G(X) \lessgtr bi \qquad /2/$$

where "bi" are coefficients defined.

Analysing accuracy criterion as well as reliability's, as defined in Pelzer /3/ and processed in Ninkov's article /4/, it can be found that as aim function of accuracy criterion as well as reliability's the following form can be applied:

$$(\lambda_{max})\,\underline{Q}_x \rightarrow \min \qquad /3/$$

The correlation matrix $\underline{Q}_x$, whose maximum own value λ_{max} - is obtained to well-known formula:

$$\underline{Q}_x = \left(\underline{A^T P A}\right)^{-1} \qquad /4/$$

where "$\underline{P}$" is a defined diagonal matrix of land surveying network elements weight. Equation /3/ can be also presented in the following form:

$$\max\left\{\det\left(\underline{Q}_x - \lambda\underline{E}\right) = 0\right\} \rightarrow \min \qquad /5/$$

where: λ = scaler at will
E = unit matrix

In its general form, the equation /5/ can be expressed in the following manner:

$$\lambda_{max}\ (\sigma_1,\ \sigma_2,\ \ldots,\ \sigma_n) \rightarrow \min \qquad /6/$$

where "σ_i" is the "ℓ_i" network elements' standard.

This form of the equation /6/ can be utilized as an aim function in determining optimum second order to first strategy.

Limitation conditions for changeable values in the equation /6/ would be of the following form:

$$\sigma_{min} \leq \sigma_i \leq \sigma_{max} \qquad /7/$$

i.e. standards of planned observations have to be within specific limits which depend on planned ones and possible methods of operation, available instruments, etc. int the network realization.

2. Mathematical Basis of Solution to the New Method

This problem's solution starts from the assumption that the first order project /i.e. the network configuration/ is adopted and unchangeable.

In this same manner, possible is to assume values of the standard σ_i of observation ℓ_i - to be located within the limits σ_{min} and σ_{max} defined in advance.

Let a "$\underline{M}$" be used to mark one vector whose elements to an increased order are "σ_i" observation standards, or:

$$\underline{M} = [\sigma_{min} = \sigma_1, \sigma_2, \ldots, \sigma_j, \ldots, \sigma_r = \sigma_{max}] \; / \; j = 1, 2, \ldots, \gamma \; / \qquad /8/$$

If the initial matrix of the $\underline{P}_o$ weights is formed with first part of the vector $\underline{M}$, i.e.

$$\underline{P}_o = \begin{bmatrix} 1/\sigma^2_{min} & & \\ & 1/\sigma^2_{min} & \\ & & 1/\sigma^2_{min} \end{bmatrix} \qquad /9/$$

the following correlation matrix can be formed:

$$o\underline{Q}_x = \left(A^T P_o A\right)^{-1} \qquad /10/$$

to give the least possible maximum own value λ^o_{max}.

This statement derives from the analysis of criterion of accuracy and reliability as expressed by means of won values of correlation matrix $\underline{Q}_x$ (Pelzer /3/).

Now, a demand can be made to determine the impact of modification of standards of individual measured element $\sigma_i = \sigma_{min}$ for $\Delta\sigma$ on value of λ^o_{max}. In other words, necessary is to determine increase in maximum own value λ^o_{max} which is caused by the change mentioned. Mathematically, it can be expressed in the following manner:

$$\frac{\partial\lambda_{max}}{\partial\sigma_i} = \frac{\partial\lambda_{max}/\sigma_1, \sigma_1, \ldots, \sigma_1/}{\partial\sigma_i} \qquad /11/$$

i.e.

$$\frac{\partial\lambda_{max}}{\partial\sigma_i} = \lim_{\Delta\sigma \to o} \frac{max/\sigma_1, \sigma_1, \ldots, \sigma_1 + \Delta\sigma, \ldots, \sigma_1/ - \lambda_{max}/\sigma_1, \sigma_1, \ldots, \sigma_1/}{\Delta\sigma} \qquad /12/$$

This way, possible is to determine increases of aim function /6/ as caused by change in standard of every individual observation for $\Delta\sigma$. Increase in aim function are obtained in the following form:

$$\Delta\lambda^i_{max} = \lambda^i_{max} - \lambda^o_{max} \qquad /13/$$

The measured element ℓ_i to whom corresponds minimum value $\Delta\lambda^i_{max} = (\Delta\lambda_{max})_{min}$ has the least impact on the change in the standard σ_i - in regards to the amount of the aim function. The coefficients of the altered proportionality of the impact in the σ_i standard modification $\Delta\sigma$ - in value of the aim function λ_{max} - are computed to the formula of:

$$K_i = \sqrt{\frac{(\Delta\lambda_{max})_{min}}{\lambda^i_{max}}} \qquad /14/$$

This coefficient of reverse proportionality shall - in continuation of the work - be used for determination of optimum project of second order to the new numerical process. The starting point, in this case, is the fact that in solving any land surveying task always is possible to define - a priori - a desired accuracy of the land surveying network parameters. To Pelzer /3/ and Ninkov /4/ this desired accuracy can be expressed through maximum own amount $\lambda_{max} = \lambda_k$ of the correlation matrix $\underline{Q}_x$.

Almost at all times, in solving problems of designing land surveying amount λ_k is considerably higher than λ°_{max} which is obtained from the $\underline{Q}_x$. That is why possible is, commencing from the P_o, to carry out an increase in standard of individual observations σ_i in proportion to the K_i coefficients, for the following elements of vectors of possible standards of observation $\underline{M}$. Accordingly, in every iteration, standards of observation ℓ_i would be computed to the following formula /the j-iteration/:

$$\sigma_i^{(j)} = \sigma_1 + K_i\,\sigma_j \qquad /15/$$

That way, in every iteration possible is to materialize $\lambda^{(j)}_{max}$ to meet the condition of:

$$\lambda^{(j)}_{max} > \lambda^{(j-1)}_{max} \qquad /16/$$

The iterative process is finalized upon meeting the following condition:

$$\lambda^{(i)}_{max} \approx \lambda_k \qquad /17/$$

In this manner determined is such an arrangement of measurement accuracy which is to provide desired accuracy of the network parameters as defined by λ_k. This way is materialized a plan of observation which considerably depends on impact of σ_i standard of the measured element on the amount of aim function. Realization of the observation plan makes possible remarkable savings in the field of time, work or means - in relation to realization of the plan of observation as obtained through classical previous estimate of accuracy in which all the element of the same kind are measured with same accuracy.

3. Example

For the purpose of illustration of the newly developed method utilized shall be optimization of accuracy of measurement in one single elementary levelling network /Figure 1/ which consists of three given and two demanded Bench Marks. A condition is made that standards of the demanded repers amount to ±2 mm, i.e. a priori defined accuracy in the network as defined by $\lambda_k = 4$.

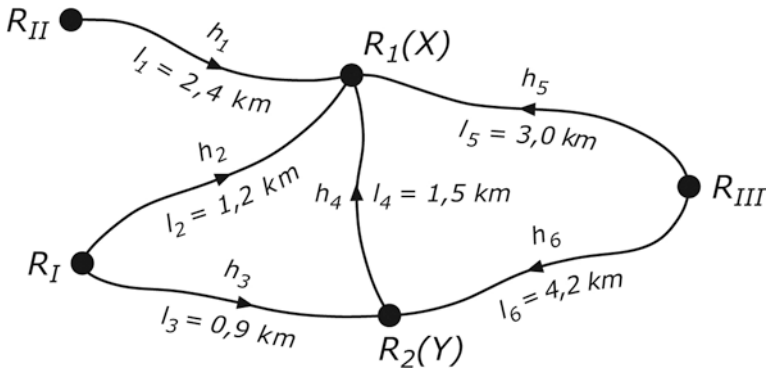


Figure 1.

Vector of possible observation standards shall be adopted in the following form:

$$\underline{M} = [1,2,3,4,5,6,7,8,16,32]$$

The maximum own amount of the correlation matrix $o\underline{Q}_x = /\underline{A}^T \underline{P}\ \underline{A}/^{-1}$ amounts to $\lambda°_{max} = 0.6867$, where configuration matrix A is of the following form:

$$\underline{A}^T = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}$$

The initial matrix of weights $\underline{P}_o$ is computed with diagonal elements $p_i = \frac{1}{\sigma^2_{min}} = 1$.

Now, possible is to perform determination of increase in the $\lambda°_{max}$ when standards of observations σ_1 are changed for $\Delta\sigma = 0.2$. In the observed instance, increases λ^i_{max} and coefficients K_i are illustrated in the Table No. 1.

T A B L E No. 1

Measurements:	ℓ_1	ℓ_2	ℓ_3	ℓ_4	ℓ_5	ℓ_6
max	0.0257	0.0463	0.1217	0.0021	0.0214	0.0212
K_i	0.3861	0.2131	0.1315	1.0000	0.3133	0.3147

By means of changing initial amounts of the standards of observation $\sigma_1 = \sigma_{min} = 1$, in proportion to the coefficients K_i for the following elements of the vector M - finally is obtained the following arrangement of the measurement accuracy:

p_1=0.056474 p_2=0.160629 p_3=0.34919 p_4=0.013600 p_5=0.040196 p_6=0.028554

which provide the desired accuracy of the demanded Bench Marks. That is how determined is an optimum project of the second order by realization of which possible is to gain considerable savings in comparison to realization of classical determined plan of observation.

4. Conclusions

The newly developed method of determining optimum project of the second order has certain advantages in relation to thus far published methods in literatures which deal with the problems.

The advantage is primarily reflected in utilization of aim function of the form $\lambda_{max} \rightarrow min$ which contains criterions of accuracy and reliability. Thereby, there is eliminated a possibility of phenomena of the optimum project of the second order with only necessary number of measurements in the network.

Similarly, by means of simultaneous measurement of the observation standard - in proportion to the K_i coefficients - achieved is an optimum project of the second order with a number of iterations contrary to some methods of the first strategy /GRAFAREND /1// which offered solution after a large number of iterations.

Naturally, this newly developed method, too, has its imperfections which are, first of all, reflected in difficulties of numerical nature in finding its own values of the matrix $\underline{Q}_x$.

However, this problem is daily reduced thanking to new computers with an increased number of numerical possibilities.

REFERENCES

/1/ GRAFAREND, E., 1979: Optimierung Geodätischer Meßoperationen. H.W. Verlag Karlsruhe, 1979.

/2/ SCHMITT, G., 1981: Optimal design of geodetic networks. Symposium on geodetic networks and computations, Munich, 1981.

/3/ PELZER, H., 1976: Genauigkeit und Zuverlässigkeit Geodätischer Netze. Tagung Mathematische Probleme der Geodäsie, Oberwolfach, 1976.

/4/ NINKOV, T., 1982: Mathematička optimizacija projektovanja geodetskih mreža. Doktorska disertacija, Gradjevinski fakultet Beograd, 1982.g.

DENSIFICATION OF A NETWORK AS A SPECIAL CASE OF DEFORMATION ANALYSIS

Haim B. PAPO and Avram PERELMUTER
Haifa and Tel-Aviv, Israel

ABSTRACT

The points in a primary and secondary networks are regarded as being in a state of continuous motion. Accordingly the datum which represents the primary network is also time variable. The adjustment of measurements made for densification of the primary network should model those time variations. This is accomplished by defining the datum through free net constraints imposed on the standard epoch positions and the velocities of the subset of primary points which are involved in the densification measurements. Datum distortions caused by freak primary net points are treated by S-transformation techniques.

1. Time-like Variations in Geodetic Networks

In the adjustment of geodetic networks we can no longer ignore the impact of time-like variations in the position of its points. Modern measurement techniques have reached a degree of precision which justifies the inclusion of velocity parameters in the mathematical models of the adjustment.

The scheme of primary and secondary geodetic network measurements and adjustment is today as it used to be centuries ago as follows:

(1) The primary network is measured over a relatively short period of time depending, of course, on the size of the country. The measurements are adjusted resulting in primary net point coordinates including their variance-covariance matrix.

(2) At various subsequent epochs and as a function of the rate and scope of development of the country the need arises for regional densification of the primary control. This is accomplished by measurements performed in the respective region between new-secondary net points and a number of existing-primary net points which provide control.

The adjustment of the densification measurements results in coordinates and covariance matrix of the secondary net points (See ADLER et al. (1979) and PAPO and PERELMUTER (1981a)). The primary net as a whole has changed during the time interval between the measurements described in (1) and those in (2) above. Reasons for that may be crustal movements, faulty reconstruction of

points which were destroyed etc. Notwithstanding the inevitable variations in the primary net we would like to have the secondary net points defined in the same datum as the one defined in the original adjustment of the primary net. This is of course impossible.

What we can do, instead, is to insure that the datum of the secondary net is consistent with those primary net points which were involved in the densification measurements. This can be accomplished only if the mathematical model for the adjustment of the densification measurements does include parameters which describe the time-like variations of the respective primary net points.

2. Mathematical Model of Measurements in Time-Variable Environment

Let us consider a two dimensional geodetic network defined for simplicity on a plane reference surface. We assume that its points move with respect to a certain well defined reference system (See PAPO and PERELMUTER (1981b)). The Cartesian coordinates of point p_i at epoch t_m would be

$$\begin{bmatrix} x \\ y \end{bmatrix}_i^m = \begin{bmatrix} x \\ y \end{bmatrix}_i^o + \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix}_i \cdot \Delta t_m \qquad (2.1)$$

Here $[x\ y]_i^o$ are the zero (standard) epoch coordinates of point i and $[\dot{x}\ \dot{y}]_i$ are the two components of its velocity. Thus a geodetic measurement which has been performed at epoch t_m between points i and j or between i, j and k would be modeled by the zero epoch coordinates and the velocity of the above points.

Apart of the datum (estimability) problem which is discussed later we need measurements made at least at two different epochs in order to be able to estimate all of the above parameters and in particular the velocities. There is an interesting alternative suggested by MARKUSE (1981) where a formal solution of the velocities is possible from measurements made at a single epoch (t_o). Measurements made at subsequent epochs and processed sequentially result in real solutions for the velocities.

The minimum condition of "at least two epochs" is met for only those primary net points which provide control for the adjustment of densification measurements as they have been measured at epochs t_o and t_m. The rest of the primary net points (measured at t_o only) are of no interest. The secondary net points have been measured only once, at t_m and so we can solve for their positional parameters only. The observation equations of the densification measurements

will be as follows: (See PAPO and PERELMUTER (1981b))

$$V = A_1 \cdot X_1 + [A_2\ A_3] \cdot \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} + \Delta t_m \cdot [A_2\ A_3] \cdot \begin{bmatrix} \dot{X}_2 \\ \dot{X}_3 \end{bmatrix} - L \qquad (2.2)$$

where

X_1 are corrections to secondary net point approximate coordinates,

X_2, X_3 are corrections to the primary net point (t_o epoch) coordinates,

$\dot{X}_2, \dot{X}_3$ are velocities of the primary net points with respect to a particular reference system defined in the next chapter,

$\Delta t_m = t_m - t_o$ is the average time interval between primary and densification measurements.

The partitioning of the primary net points into X_2 and X_3 ($\dot{X}_2, \dot{X}_3$) is arbitrary in general and depends on the size and the nature of the datum defect of the observation system (See PAPO and PERELMUTER (1981a)). In our case, assuming that distances have been measured at t_m, X_3 is of size three and consists of corrections to the x,y (t_o epoch) coordinates of one point and x or y another point in the primary network. $\dot{X}_3$ is of the same size and consists of the respective velocities of the above points.

3. Datum Definition of the Secondary Network

The observation equations system (2.2) is rank deficient due to the lack of datum. Compared to conventional (static) systems in which velocities are not being modeled the size of the defect is exactly doubled (See PAPO and PERELMUTER (1981b)). Our interpretation is that we need datum definition for the zero epoch coordinates $(X_1, X_2, X_3)^o$ as well as for the velocities ($\dot{X}_2, \dot{X}_3$) of the primary net points which were involved in the densification measurements.

A minimal constraints solution of system (2.2) can be obtained by setting to zero the X_3 and $\dot{X}_3$ parameters (a total of two by three quantities). The remaining parameters are then estimated $(X_1^*, X_2^*, \dot{X}_2^*)$ by the solution of the following full rank normal equations system:

$$\begin{bmatrix} N_{11} & N_{12} & \Delta t_m \cdot N_{12} \\ N_{21} & N_{22} & \Delta t_m \cdot N_{22} \\ \Delta t_m \cdot N_{21} & \Delta t_m \cdot N_{22} & \Delta t_m^2 \cdot N_{22} \end{bmatrix} \cdot \begin{bmatrix} X_1^* \\ X_2^* \\ \dot{X}_2^* \end{bmatrix} = \begin{bmatrix} U_1 \\ U_2 \\ \Delta t_m \cdot U_2 \end{bmatrix} \qquad (3.1)$$

where

$$N_{11} = A_1^T \cdot P \cdot A_1 \quad ; \quad N_{12} = A_1^T \cdot P \cdot A_2 \quad ; \quad N_{21} = A_2^T \cdot P \cdot A_1 \quad ; \quad N_{22} = A_2^T \cdot P \cdot A_2$$

$$U_1 = A_1^T \cdot P \cdot L \quad ; \quad U_2 = A_2^T \cdot P \cdot L$$

The datum of the above minimal constraints solution is of limited value due to the entirely arbitrary partitioning of the primary net points into X_2 and X_3. What is more, the primary net points in X_2 are not represented at all in the datum definition. A more meaningful solution, one which represents according to their quality all of the primary net points involved in the t_m epoch measurements, is obtained by applying weighted free net constraints to the (2.2) system. The following two minimum conditions are to be satisfied:

$$[X_2^T X_3^T] \cdot P_X \cdot \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} = \min \quad \text{and} \quad [\dot{X}_2^T \dot{X}_3^T] \cdot P_X \cdot \begin{bmatrix} \dot{X}_2 \\ \dot{X}_3 \end{bmatrix} = \min \qquad (3.2)$$

where P_X the a-priori weight matrix of the X_2, X_3 parameters is based in general on the initial (t_o epoch) solution of the complete primary network. By means of the above two conditions (3.2) a unique reference frame is defined for the velocities of the primary net points as follows:

* the weighted sum of the velocities is zero
* the weighted sum of mixed products of position and velocity $(x_i \dot{y}_i - y_i \dot{x}_i)$ is zero (it is equal to the norm of the vector product of position and velocity vectors of a point, see also PAPO and PERELMUTER (1982b)).

We should note that the datum defined by the above constraints depends on the particular P_X (a-priori weight matrix).

The solution of the free net X and $\dot{X}$ parameters is obtained by a transformation of the X^* and $\dot{X}^*$ parameters (See PAPO and PERELMUTER (1981a)). For completeness we bring here without derivation the sequence of formulae with slight modifications.

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = R \cdot \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}^* \quad ; \quad Q = R \cdot Q^* \cdot R^T \qquad (3.3)$$

$$\begin{bmatrix} 0 \\ \dot{X}_2 \\ \dot{X}_3 \end{bmatrix} = R \cdot \begin{bmatrix} 0 \\ \dot{X}_2 \end{bmatrix}^* \quad ; \quad \dot{Q} = R \cdot \dot{Q}^* \cdot R^T \qquad (3.4)$$

where

$$R = \bar{G}^T \cdot \bar{S}^{-1} \quad ; \quad \bar{S} = G \cdot \bar{G}^T \quad ; \quad G = [\, I \;\; G_{12} \,] \quad ; \quad \bar{G} = [\, I \;\; \bar{G}_{12} \,]$$

$$\bar{G}_{12}^T = \left[P_{33} - G_{12}^T \cdot \begin{bmatrix} 0 \\ P_{23} \end{bmatrix} \right]^{-1} \cdot \left[G_{12}^T \cdot \begin{bmatrix} 0 & 0 \\ 0 & P_{22} \end{bmatrix} - [\, 0 \;\; P_{32} \,] \right] ; \quad P_X = \begin{bmatrix} P_{22} & P_{23} \\ P_{32} & P_{33} \end{bmatrix}$$

$$G_{12}^T = [\, A_3^T \cdot P \cdot A_1 \;\; A_3^T \cdot P \cdot A_2 \,] \cdot \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \qquad (3.5)$$

There is an alternative way for evaluating the G_{12}^T matrix which will be discussed in the next chapter.

The above procedure insures for the secondary net points a datum which differs slightly from the original t_o epoch datum of the primary net. The difference in datum is due to the inclusion of the densification measurements as well as to the additional parameters of the adjustment model (velocities of the primary net points). Those velocities are in our case nuisance parameters. Their principal function is to filter out the effects of relative motion of the primary points on the densification measurements. There is however one additional purpose which can be served by those velocities and their covariance matrix.

We are usually interested to prelude distortions of the datum which may be caused by freak "misbehaved" primary net points. A convenient statistical test of the behaviour of the primary net points is the F-test of hypothesis applied to their relative velocities (See PAPO and PERELMUTER (1981b) and (1982b)). The F statistic which is being tested is defined by the following expression: (See also PELZER (1971) and BRUNNER (1979))

$$F = \frac{\left(\dot{X}_2^T \, \dot{Q}_2^{-1} \, \dot{X}_2 \right) \cdot f}{(V^T \, P \, V) \cdot f_2} \qquad (3.6)$$

where $\dot{Q}_2$ is a submatrix of the variance-covariance matrix of the velocity estimates.

f is the degree of freedom of densification measurements,

f_2 is the number of $\dot{X}_2$ parameters.

If the F-test zero hypothesis is rejected at a given α level the $\dot{X}_2, \dot{X}_3$ velocities are inspected and the suspected "freak" point is suspended from the datum definition club. This is accomplished by setting to zero the respective rows and columns of the P_X matrix. Now $\dot{X}_2$ and $\dot{Q}_2$ are reestimated

in the new reference frame (defined without the suspended point) and the F-test r is repeated. Although the trial and error testing procedure described above can be improved by certain computational techniques, snooping for the freak point or points is highly subjective and depends on the qualifications of the human analyst.

4. Change of Minimal Constraints by S-transformation

The look-out for a freak primary net point can result in one of the following two mutually exclusive cases:

Case 1: none of the parameters of the suspected point is a part of X_3.

Case 2: X_3 contains one or more of the suspected point's parameters.

The method for the suspension of a freak point from the primary net in case 1 is straight forward and was described at the end of last chapter. If the freak point belongs to case 2 we first transform the situation into case 1 and then proceed as above.

Conversion of case 2 into case 1 is performed by the change of minimal constraints, known in the literature as an "S-transformation" (See BAARDA (1973), MIERLO (1980) and MOLENAAR (1981)). In essence it boils down to a repartitioning of the primary net point parameters into a new set of X_2, X_3 ($\dot{X}_2, \dot{X}_3$). A brute force approach would require the formation of a new system of normal equations and their subsequent solution for X_1^*, X_2^* and $\dot{X}_2^*$.

In what follows we describe a simple alternative method for performing the change in minimal constraints which (as in other S-transformation methods) avoid the need for repeated formation and solution of the normal system. Equation (3.3) premultiplied by G results in the trivial inverse transformation

$$\begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix} = G \cdot \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \qquad (4.1)$$

We note again that the X_1, X_2, X_3 vector is unique and depends on the P_X matrix. It doesn't depend on the particular partitioning of the X_2, X_3 parameters. If we reorder X_2, X_3 so that the freak point's parameters in X_3 are exchanged for appropriate parameters in X_2 we can write equation (4.1) for the new vector X_1, X_2, X_3 resulting in the new minimal constraints solution X_1^*, X_2^*. We have, of course, to reevaluate the G matrix to correspond to the new partitioning of X_2 and X_3. Instead of evaluating G_{12}^T ($G = [\, I \; G_{12} \,]$) by

equation (3.5) we can use an easier alternative. We have shown in PAPO and PERELMUTER (1982a) that the G_{12}^{T} matrix can be evaluated from Helmert's transformation matrix C. The computations are simple and can be performed easily by a pocket calculator.

The above S-transformation method is a special case of a more general situation where the repartitioning involves the X_1 vector too. It is easy to see that the proposed method for change of minimal constraints holds also for the velocities according to equation (3.4).

5. Conclusions

The method for the adjustment of horizontal control densification measurements presented in this paper, being a follow-up of our previous work, brings the time element into our considerations resulting in a new mathematical model. The time-like variations in the coordinates of the primary control network points are filtered out from the measurements resulting in a relatively undisturbed datum for the secondary net points.

A similar approach can be taken to the analysis of deformations the difference being in our interpretation of the results as shown in PAPO and PERELMUTER (1982b). In deformation analysis we are interested in the relative velocities of the network points while in our present case we seek the standard epoch positions of the same.

The S-transformation technique presented at the end is one of a number of mathematical tools which are being developed by us for a more efficient and objective analysis of the primary network points involved in the densification measurements.

REFERENCES

ADLER, R., PAPO, H.B., PERELMUTER, A., 1979: The Establishment of Standards for Lower Order Horizontal Control. Survey Review, Vol. XXV, No. 194.

BAARDA, W., 1973: S-transformation and Criterion Matrices. Netherlands Geodetic Commission, Vol. 5, No. 1.

BRUNNER, F.K., 1979: On the Analysis of Geodetic Networks for the Determination of the Incremental Strain Tensor. Survey Review, Vol. XXV, No. 192.

MARKUSE, U.I., 1981: Analysis and Adjustment of Geodetic Networks without Forming Normal Equations. Izvestia V.U.Z. Geodesia i Aerofotosiemka, No. 3.

MIERLO, J. van, 1980: Free Network Adjustment and S-transformation. D.G.K., Reihe B, Heft Nr. 252, München.

MOLENAAR, M., 1981: S-transformation and Artificial Covariance Matrices in Photogrammetry. I.T.C. Journal, No. 1.

PAPO, H. and PERELMUTER, A., 1981a: Datum Definition by Free Net Adjustment. Bulletin Géodésique, No. 55.

PAPO, H. and PERELMUTER, A., 1981b: Kinematic Analysis of Deformations. FIG XVI. International Congress in Montreux, Switzerland.

PAPO, H. and PERELMUTER, A., 1982a: Free Net Analysis of Close Range Photogrammetry. Photogrammetric Engineering, Vol. XLVIII, No. 4.

PAPO, H. and PERELMUTER, A., 1982b: Deformations as Reflected in the Kinematics of a Network of Points. III. International Symposium in Budapest, Hungary.

PELZER, H., 1971: Zur Analyse geodätischer Deformationsmessungen. D.G.K., Reihe C, Heft Nr. 164, München.

INFLUENCE OF SYSTEMATIC EFFECTS IN STOCHASTIC AND FUNCTIONAL MODELS

Hans PELZER
Hannover, Federal Republic of Germany

1. Insufficient adjustment models

In the adjustment of geodetic networks as well as in other adjustment problems we describe the relationships between observed quantities and unknown parameters by a functional model, e. g. in the linearized form

$$\underline{\tilde{l}} = \underline{A}\ \underline{\tilde{x}}\ , \tag{1.1}$$

$\underline{\tilde{l}}$: true values of observed quantities (n x 1 vector),

$\underline{\tilde{x}}$: true values of unknown parameters (u x 1 vector),

$\underline{A}$: n x u matrix of coefficients.

Replacing the true observation vector $\underline{\tilde{l}}$ in (1.1) by its observed value $\underline{l}$ we get the observation equation

$$\underline{l} + \underline{v} = \underline{A}\ \underline{\hat{x}} \tag{1.2}$$

$\underline{v}$: vector of residuals,

$\underline{\hat{x}}$: parameter vector to be estimated.

The stochastic properties of $\underline{l}$ are described by

$$E(\underline{l}) = \underline{\mu}_l = \underline{\tilde{l}} \tag{1.3}$$

$$\underline{\varepsilon} = \underline{l} - \underline{\mu}_l\ , \tag{1.4}$$

$$\underline{\Sigma}_{ll} = E(\underline{\varepsilon}\underline{\varepsilon}^T) = \sigma_0^2\, \underline{Q}_{ll} = \sigma_0^2\, \underline{P}^{-1} \tag{1.5}$$

$\underline{\mu}_l$: expected value of $\underline{l}$,

$\underline{\Sigma}_{ll}$: covariance matrix of $\underline{l}$,

$\underline{Q}_{ll}$: cofactor matrix,

$\underline{P}$: weight matrix,

σ_0^2 : variance of unit weight.

The covariance matrix $\underline{\Sigma}_{ll}$ together with equation (1.3) is called the stochastic model, and in combination with the functional model (1.1) the least square solution of (1.2) is given by the well-known formulae

$$\underline{N}_{xx}\,\underline{\hat{x}} - \underline{n}_x = \underline{0} \;, \tag{1.6}$$

$$\underline{N}_{xx} = \underline{A}^T \underline{P}\,\underline{A} \;;\; \underline{n}_x = \underline{A}^T \underline{P}\,\underline{l} \;, \tag{1.7}$$

$$\underline{Q}_{xx} = \underline{N}_{xx}^{-1} \;, \tag{1.8}$$

$$\underline{\hat{x}} = \underline{Q}_{xx}\,\underline{n}_x \;, \tag{1.9}$$

$$\underline{\Sigma}_{xx} = \sigma_0^2\,\underline{Q}_{xx} \;, \tag{1.10}$$

$\underline{\Sigma}_{xx}$: covariance matrix of $\underline{\hat{x}}$.

Of course, the adjustment model cannot be more than an approximation to the physical reality, which may be sufficient or not. If the model is not sufficient, the residual vector $\underline{v}$ does not agree with its covariance matrix

$$\underline{\Sigma}_{vv} = \sigma_0^2\left(\underline{Q}_{ll} - \underline{A}^T \underline{Q}_{xx}\,\underline{A}\right) \;, \tag{1.11}$$

detectable by statistical tests.

In the case of insufficient adjustment model the functional model (1.1) has to be extended. Generally this can be done by additional parameters, forming a vector $\underline{\xi}$,

$$\underline{\tilde{l}} = \underline{A}\;\underline{x} + \underline{B}\;\underline{\tilde{\xi}} = [\underline{A} \vdots \underline{B}] \left|\begin{matrix} \underline{\tilde{x}} \\ \cdots \\ \underline{\tilde{\xi}} \end{matrix}\right| \tag{1.12}$$

$\underline{\tilde{\xi}}$: true value of additional parameters (m x 1 vector),

$\underline{B}$: n x m matrix of coefficients.

If the true value $\tilde{\underline{\xi}}$ is known, its influence to the observation vector $\underline{l}$ may be considered as a systematic model error $\underline{\Delta}$,

$$\tilde{\underline{l}} - \tilde{\underline{\Delta}} = \underline{A}\ \tilde{\underline{x}}\ , \qquad (1.13)$$

$$\tilde{\underline{\Delta}} = \underline{B}\ \tilde{\underline{\xi}}\ , \qquad (1.14)$$

$\tilde{\underline{\Delta}}$: systematic model error in (n x 1 vector).

In this case in (1.2) and (1.7) the observation vector $\underline{l}$ has to be replaced by

$$\underline{l} = \underline{l} - \tilde{\underline{\Delta}} = \underline{l} - \underline{B}\ \tilde{\underline{\xi}} \qquad (1.15)$$

and the ordinary solution (1.6) – (1.10) can used now as before.

Normally, however, the true value $\tilde{\underline{\xi}}$ is not known, i.e. the model (1.12) leads to another solution.

2. Typical examples

2.1 Calibration parameters of EDM-instruments

Normally the calibration of EDM-instruments, i.e. in simple cases the determination of the scale factors and additive constants, will be carried out in special calibration networks or by laboratory tests. But, as an alternative procedure, these calibration parameters or some of them can be determined directly from the network observations, this method is often called "on-the-job-calibration". In these cases the calibration parameters play the role of the additional parameters $\underline{\xi}$ in equation (1.12).

On the other hand, even if a special calibration measurement has been carried out, the resulting parameters are of limited accuracy and may be improved by the network adjustment.

2.2 Centering errors

In the functional model (1.1) all observations from or to a certain network station are considered as centered to one and the same point. Of course, this model is disturbed if one or more centering errors occur, such centering errors may be considered as additional parameters $\underline{\xi}$ in the extended model (1.12).

In practice we have to distinguish two cases. In the first one we have only a limited number of possibly gross centering errors in the network; these errors can be estimated in the extended model (1.12). In the second case, however, all observations contain small centering errors which can be described by their statistical distribution, this case is considered in Ch. 4 and 5.

3. Estimation of additional parameters without stochastic information

In this model the additional parameters $\underline{\xi}$ in (1.12) are considered as nonstochastical quantities like the main parameters $\underline{x}$; i.e. they have to be estimated in the same manner. The model may be used, for example, in the case of on-the-job-calibration (Ch. 2.1), or for the purpose of determination of gross centering errors (Ch. 2.2).

The extended model (1.12) leads to the extended observation equation

$$\underline{l} + \underline{v} = [\underline{A} \vdots \underline{B}] \begin{vmatrix} \hat{\underline{x}} \\ \cdots \\ \hat{\underline{\xi}} \end{vmatrix} \qquad (3.1)$$

$\hat{\underline{\xi}}$: estimated value of $\tilde{\underline{\xi}}$.

and together with the unchanged stochastic model (1.5) we get the normal equation

$$\left|\begin{array}{c|c} \underline{A}^T\underline{P}\ \underline{A} & \underline{A}^T\underline{P}\ \underline{B} \\ \hline \underline{B}^T\underline{P}\ \underline{A} & \underline{B}^T\underline{P}\ \underline{B} \end{array}\right| \left|\begin{array}{c} \hat{\underline{x}} \\ \hline \hat{\underline{\xi}} \end{array}\right| - \left|\begin{array}{c} \underline{A}^T\underline{P}\ \underline{l} \\ \hline \underline{B}^T\underline{P}\ \underline{l} \end{array}\right| = \underline{0} \qquad (3.2)$$

or, with other symbols for abbreviation,

$$\left|\begin{array}{c|c} \underline{N}_{xx} & \underline{N}_{x\xi} \\ \hline \underline{N}_{\xi x} & \underline{N}_{\xi\xi} \end{array}\right| \left|\begin{array}{c} \hat{\underline{x}} \\ \hline \hat{\underline{\xi}} \end{array}\right| - \left|\begin{array}{c} \underline{n}_x \\ \hline \underline{n}_\xi \end{array}\right| = \underline{0}\ . \qquad (3.3)$$

With

$$\left|\begin{array}{c|c} \underline{Q}_{xx} & \underline{Q}_{x\xi} \\ \hline \underline{Q}_{\xi x} & \underline{Q}_{\xi\xi} \end{array}\right| = \left|\begin{array}{c|c} \underline{N}_{xx} & \underline{N}_{x\xi} \\ \hline \underline{N}_{\xi x} & \underline{N}_{\xi\xi} \end{array}\right|^{-1}\ , \qquad (3.4)$$

the solution of (3.2) resp. (3.3) is

$$\left|\begin{array}{c} \hat{\underline{x}} \\ \hline \hat{\underline{\xi}} \end{array}\right| = \left|\begin{array}{c|c} \underline{Q}_{xx} & \underline{Q}_{x\xi} \\ \hline \underline{Q}_{\xi x} & \underline{Q}_{\xi\xi} \end{array}\right| \left|\begin{array}{c} \underline{n}_x \\ \hline \underline{n}_\xi \end{array}\right|\ . \qquad (3.5)$$

This solution is advantageous if one is interested in the numerical value of $\hat{\underline{\xi}}$, e.g. if the statistical significance of $\hat{\underline{\xi}}$ has to be tested.

If, however, the numerical value of $\hat{\underline{\xi}}$ is not of special interest $\hat{\underline{\xi}}$ may be eliminated beforehand by

$$\bar{\underline{N}}_{xx} = \underline{N}_{xx} - \underline{N}_{x\xi}\,\underline{N}_{\xi\xi}^{-1}\,\underline{N}_{\xi x}\ , \qquad (3.6)$$

$$\bar{\underline{n}}_x = \underline{n}_x - \underline{N}_{x\xi}\,\underline{N}_{\xi\xi}^{-1}\,\underline{n}_\xi\ , \qquad (3.7)$$

$$\bar{\underline{N}}_{xx}\,\hat{\underline{x}} - \bar{\underline{n}}_x = \underline{0}\ . \qquad (3.8)$$

The solution of (3.8) is simply

$$\underline{\hat{x}} = \underline{\bar{N}}_{xx}^{-1}\,\underline{\bar{n}}_x = \underline{Q}_{xx}\,\underline{\bar{n}}_x \quad , \tag{3.9}$$

$$\underline{Q}_{xx} = \underline{\bar{N}}_{xx}^{-1} \quad , \tag{3.10}$$

$\underline{\hat{x}}$ normally distributed with expected value

$$E(\underline{\hat{x}}) = \underline{\tilde{x}} \tag{3.11}$$

and covariance matrix

$$\underline{\Sigma}_{xx} = \sigma_0^2\,\underline{Q}_{xx} \quad . \tag{3.12}$$

From (3.6) follows that the matrix $\underline{\bar{N}}_{xx}$ may be singular though $\underline{N}_{xx}$ is not, i.e. the additional parameters $\underline{\xi}$ may be not estimable from the observation vector $\underline{l}$. But even in the case where $\underline{\bar{N}}_{xx}$ is regular and $\underline{Q}_{xx} = \underline{\bar{N}}_{xx}^{-1}$ exists, the inequality.

$$\mathrm{tr}\left(\underline{Q}_{xx}\right) \geq \mathrm{tr}\left(\underline{N}_{xx}^{-1}\right)$$

holds, i.e. the determination of $\underline{\xi}$ in the extended model is connected with a loss in precision of the main parameters $\underline{x}$. That is the price we have to pay for an unbiased estimation of $\underline{x}$.

In order to proof (3.12) we take from (3.4)

$$\underline{Q}_{xx} = \underline{N}_{xx}^{-1} + \underline{N}_{xx}^{-1}\,\underline{N}_{x\xi}\,\underline{Q}_{\xi\xi}\,\underline{N}_{\xi x}\,\underline{N}_{xx}^{-1} = \underline{N}_{xx}^{-1} + \underline{F}\;\underline{F}^T \quad , \tag{3.14}$$

with an auxiliary matrix

$$\underline{F} = \underline{N}_{xx}^{-1}\,\underline{N}_{x\xi}\,\underline{Q}_{\xi\xi}^{\frac{1}{2}} \quad . \tag{3.15}$$

Therefore

$$\mathrm{tr}\left(\underline{Q}_{xx}\right) = \mathrm{tr}\left(\underline{N}_{xx}^{-1}\right) + \mathrm{tr}\left(\underline{F}\;\underline{F}^T\right) \tag{3.16}$$

and, for any real matrix $\underline{F}$,

$$\mathrm{tr}\left(\underline{F}\;\underline{F}^T\right) \geq 0 \quad . \tag{3.17}$$

4. Estimation of additional parameters as pseudo observations

In many cases it is impossible to follow the solution of Ch. 3 because the matrix $\underline{\bar{N}}_{xx}$ in (3.10) is singular and no simple inverse $\underline{Q}_{xx}$ exists. In such cases, however, we may often regard the additional parameters $\underline{\xi}$ as stochastic quantities with expected value

$$E\left(\underline{\xi}\right) = \underline{0} \tag{4.1}$$

and given matrix

$$\underline{\Sigma}_{\xi\xi} = \sigma_o^2 \underline{Q}_{\xi\xi} \quad . \tag{4.2}$$

But, independent of the question whether or not the matrix $\underline{\bar{N}}_{xx}$ is singular, the parameter vector $\underline{\xi}$ often is of stochastic nature and should be treated in the way described below.

For example in the case of determination of calibration parameters (Ch. 2.1) preliminary values of these parameters may be known from calibration procedures, together with the corresponding covariance matrix $\underline{\Sigma}_{\xi\xi}$. Then we may interpretate the remaining errors in the parameters as stochastic values with properties (4.1) and (4.2) and the network adjustment only as a possibility to improve these calibration parameters.

As another example we may consider the centering errors (Ch. 2.2) as random errors to be described by (4.1) and (4.2), where the covariance matrix $\underline{\Sigma}_{\xi\xi}$ has to be estimated based on practical experience.

On the basis of (4.1) and (4.2) we may extend the observation equation (3.1) by addition of pseudo observations

$$\left|\begin{matrix}\underline{l}\\ --\\ \underline{0}\end{matrix}\right| + \left|\begin{matrix}\underline{v}\\ --\\ \hat{\underline{\xi}}\end{matrix}\right| = \left|\begin{matrix}\underline{A} & | & \underline{B}\\ ----&+&----\\ \underline{0} & | & \underline{I}\end{matrix}\right| \left|\begin{matrix}\hat{\underline{x}}\\ --\\ \hat{\underline{\xi}}\end{matrix}\right| , \tag{4.3}$$

corresponding to the stochastical model

$$\underline{\Sigma} = \left|\begin{matrix}\underline{\Sigma}_{ll} & | & \underline{0}\\ ---&+&---\\ \underline{0} & | & \underline{\Sigma}_{\xi\xi}\end{matrix}\right| = \sigma_o^2 \left|\begin{matrix}\underline{Q}_{ll} & | & \underline{0}\\ ---&+&---\\ \underline{0} & | & \underline{Q}_{\xi\xi}\end{matrix}\right| = \sigma_o^2 \left|\begin{matrix}\underline{P} & | & \underline{0}\\ ---&+&---\\ \underline{0} & | & \underline{P}_{\xi\xi}\end{matrix}\right|^{-1} \quad . \tag{4.4}$$

The resulting normal equation is, with symbols defined in (3.3)

$$\left|\begin{array}{c|c} \underline{N}_{xx} & \underline{N}_{x\xi} \\ \hline \underline{N}_{\xi x} & \underline{N}_{\xi\xi}+\underline{P}_{\xi\xi} \end{array}\right| \left|\begin{array}{c} \bar{\underline{x}} \\ \hline \bar{\underline{\xi}} \end{array}\right| - \left|\begin{array}{c} \underline{n}_x \\ \hline \underline{n}_\xi \end{array}\right| = \underline{O} \qquad (4.5)$$

with the solution

$$\left|\begin{array}{c} \underline{x} \\ \hline \underline{\xi} \end{array}\right| = \left|\begin{array}{c|c} \bar{\underline{Q}}_{xx} & \bar{\underline{Q}}_{x\xi} \\ \hline \bar{\underline{Q}}_{\xi x} & \bar{\underline{Q}}_{\xi\xi} \end{array}\right| \left|\begin{array}{c} \underline{n}_x \\ \hline \underline{n}_\xi \end{array}\right| , \qquad (4.6)$$

where the cofactor matrix $\bar{\underline{Q}}$,

$$\bar{\underline{Q}} = \left|\begin{array}{c|c} \bar{\underline{Q}}_{xx} & \bar{\underline{Q}}_{x\xi} \\ \hline \bar{\underline{Q}}_{\xi x} & \bar{\underline{Q}}_{\xi\xi} \end{array}\right| = \left|\begin{array}{c|c} \underline{N}_{xx} & \underline{N}_{x\xi} \\ \hline \underline{N}_{\xi x} & \underline{N}_{xx}+\underline{P}_{\xi\xi} \end{array}\right|^{-1} , \qquad (4.7)$$

is not identical with the corresponding matrix in (3.4). Similar to Ch. 3, the vector $\bar{\underline{\xi}}$ can be eliminated beforehand by

$$\bar{\underline{N}}_{xx} = \underline{N}_{xx} - \underline{N}_{x\xi}\left(\underline{N}_{\xi\xi}+\underline{P}_{\xi\xi}\right)^{-1}\underline{N}_{\xi x} , \qquad (4.8)$$

$$\bar{\underline{n}}_x = \underline{n}_x - \underline{N}_{x\xi}\left(\underline{N}_{\xi\xi}+\underline{P}_{\xi\xi}\right)^{-1}\underline{n}_\xi , \qquad (4.9)$$

leading to the reduced normal equation

$$\bar{\underline{N}}_{xx}\,\underline{x} - \bar{\underline{n}}_x = \underline{O} \qquad (4.10)$$

with the solution

$$\bar{\underline{Q}}_{xx} = \bar{\underline{N}}_{xx}^{-1} , \qquad (4.11)$$

$$\bar{\underline{x}} = \bar{\underline{Q}}_{xx}\,\bar{\underline{n}}_x , \qquad (4.12)$$

where, of course, the parameter vector $\underline{\bar{x}}$ is different from the corresponding vector $\underline{\hat{x}}$ in (3.9).

Evidently, the model considered in this chapter is a general one and contains both, the model of Ch. 1 without additional parameters as well as that of Ch. 3. The first model is defined by

$$\underline{\Sigma}_{\xi\xi} = \underline{O} \;,\; \underline{P}_{\xi\xi} \to \infty\, \underline{I} \;,\; \left(\underline{N}_{\xi\xi}+\underline{P}_{\xi\xi}\right)^{-1} \to \underline{O} \;, \tag{4.13}$$

and the second one follows simply from

$$\underline{\Sigma}_{\xi\xi} \to \infty\, \underline{I} \;,\; \underline{P}_{\xi\xi} = \underline{O} \;, \tag{4.14}$$

see (4.8) and (4.9).

5. Correlation model

On the same conditions as in Ch. 4, i.e. if the additional parameters $\underline{\xi}$ can be considered as random variates with properties (4.1) and (4.2), we may introduce a quite different adjustment model. The model equation (1.12) we can write in the form

$$\underline{\tilde{l}} - \underline{B}\ \underline{\tilde{\xi}} = \underline{\tilde{l}} - \underline{\tilde{\Delta}} = \underline{A}\ \underline{\bar{x}} \;, \tag{5.1}$$

cf. (1.13), leading to the observation equation

$$\underline{l} - \underline{\Delta} = \underline{A}\ \underline{\bar{x}} \;, \tag{5.2}$$

where $\underline{\Delta}$ is a random vector of systematic effects with expected value (cf. (4.1) and (1.14))

$$E(\underline{\Delta}) = E\left(\underline{B}\,\underline{\xi}\right) = \underline{B}\,E\left(\underline{\xi}\right) = \underline{O} \tag{5.3}$$

and variance matrix $\underline{\Sigma}_{\Delta\Delta}$,

$$\underline{\Sigma}_{\Delta\Delta} = \sigma_o^2 \underline{Q}_{\Delta\Delta} \quad , \tag{5.4}$$

where $\underline{Q}_{\Delta\Delta}$ follows with (1.14) from the law of error propagation,

$$\underline{Q}_{\Delta\Delta} = \underline{B}\,\underline{Q}_{\xi\xi}\underline{B}^T \quad . \tag{5.5}$$

In accordance with (4.3) the observed value $\underline{\Delta}$ of systematic model errors in $\underline{l}$ may be zero,

$$\underline{\Delta} = \underline{B}\,\underline{\xi} = \underline{0} \quad . \tag{5.6}$$

In this case the functional model (1.1) resp. the observation equation (1.2) remains unchanged

$$\underline{l} - \underline{\Delta} + \underline{v} = \underline{l} + \underline{v} = \underline{A}\,\overline{\underline{x}} \quad , \tag{5.7}$$

but the stochastic model, i.e. the covariance matrix $\underline{\Sigma}_{ll}$ of $\underline{l}$ changes to

$$\overline{\underline{\Sigma}}_{ll} = \sigma_o^2 \overline{\underline{Q}}_{ll} = \sigma_o^2 \left(\underline{Q}_{ll} + \underline{Q}_{\Delta\Delta}\right) = \sigma_o^2 \left(\underline{Q}_{ll} + \underline{B}\,\underline{Q}_{\xi\xi}\underline{B}^T\right) \quad . \tag{5.8}$$

The resulting normal equation is

$$\underline{A}^T \left(\underline{Q}_{ll} + \underline{B}\,\underline{Q}_{\xi\xi}\underline{B}^T\right)^{-1} \underline{A}\,\overline{\underline{x}} - \underline{A}^T \left(\underline{Q}_{ll} + \underline{B}\,\underline{Q}_{\xi\xi}\underline{B}^T\right)^{-1} \underline{l} = \underline{0} \tag{5.9}$$

its solution is identical with the solution of (4.10), because in both cases the same functional and stochastical information is used; for a numerical verification see EBNER (1973). For the same reason $\overline{\underline{Q}}_{xx}$ from (4.11) is equal to

$$\overline{\underline{Q}}_{xx} = \left(\underline{A}^T \left(\underline{Q}_{ll} + \underline{B}\,\underline{Q}_{\xi\xi}\underline{B}^T\right)^{-1} \underline{A}\right)^{-1} \quad . \tag{5.10}$$

6. Conclusion

Systematic errors in the observations may be interpreted as errors in the mathematical adjustment model, which can be eliminated by insertion of additional parameters $\underline{\xi}$ in the model. A general solution of this problem can be found if these additional parameters are considered as random variables. In this solution, all information about these parameters is concentrated in the covariance matrix $\underline{\Sigma}_{\xi\xi}$ of $\underline{\xi}$.

With any matrix norm $\|\cdot\cdot\|$ we get

$$\|\underline{0}\| \leq \left\|\underline{\Sigma}_{\xi\xi}\right\| < \|\infty\,\underline{I}\| \quad . \tag{6.1}$$

In the first (left hand) extreme case ($\|\underline{0}\| = 0$) we have no further information about the systematic effect and have to detect them from the network adjustment alone (s. Ch. 3). In the right hand extreme case, however, we know exactly the true values $\underline{\xi}$ of the systematic errors and can reduce our observations $\underline{l}$ before carrying out the network adjustment.

In most practical cases the covariance matrix $\underline{\Sigma}_{\xi\xi}$ is neither the zero nor an infinite matrix and therefore one of the solutions given in Ch. 4 and 5 have to be used. Because the results of both solutions are identical, we can use one of these solutions with respect to the practical computation. From this point of view, normally the solution given in Ch. 4 may be more suitable.

REFERENCES

EBNER, H., 1973: Zusätzliche Parameter in Ausgleichungen. Zeitschrift für Vermessungswesen 98, p. 385-391.

MIKHAIL, E.M., 1976: Observations and least squares. New York.

PELZER, H., 1981: Leistungsfähigkeit und Grenzen des Funktionalmodells. Meeting of the Working Group III/1 of the International Society of Photogrammetry and Remote Sensing, Stuttgart, November 26-27, 1981.

SCHMID, H.H. and SCHMID, E., 1965: A Generalized Least Squares Solution for Hybrid Measuring Systems. The Canadian Surveyor XIX, p. 27-41.

SOME ESTIMABLE FUNCTIONS IN GEODETIC NETWORKS

Gligorije PEROVIĆ
Beograd, Yugoslavia

ABSTRACT

Using Drozdov's definition of Geodetic Networks and Gauss-Markov's Theoreme it is proved that angles and distances are estimable functions.

Introduction

Definition (Drozdov, 1972): Set of known (for example of fixed points), unknown $(X_1, X_2, \ldots, X_m)$ and observed values $l_1, l_2, \ldots, l_n \quad (n > m)$ are called Network, if:

first, if among observed values $l_i \quad (i = 1, 2, \ldots, n)$ can be found m such as, knowledge of exact values, together with known values, is necessary and sufficient to determine unknown values, i.e.

$$X_i = X_i\,(l_1, l_2, \ldots, l_m) \qquad i = 1, 2, \ldots, m \qquad (1')$$

and

$$l_i = l_i\,(X_1, X_2, \ldots, X_m) \qquad i = 1, 2, \ldots, m \qquad (1'')$$

(those m values are called necessary values);

second, any of the other observed values $l_{m+1}, \ldots, l_n$ (there are $n - m$ and are called unnecessary) of the given set can be presented as function necessary values $l_1, l_2, \ldots, l_m$, i.e.

$$l_{m+k} = f_k\,(l_1, l_2, \ldots, l_m) \qquad k = 1, 2, \ldots, n - m \qquad (2)$$

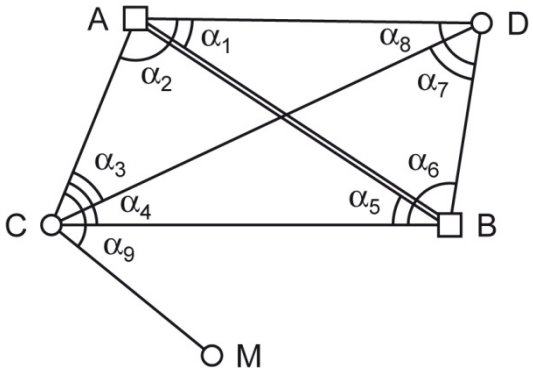


Fig. 1.

So, for example, set of observed angles $\alpha_1, \alpha_2, \ldots, \alpha_9$ together with known points A and B and unknown coordinates of the points C and D (Fig. 1.) do not form a Network, although, for example, angles $\alpha_1, \alpha_2, \alpha_5$ and α_6 are necessary for the solution of that problem, but the angle α_9 can not be expressed by them. However the set measured angles $\alpha_1, \alpha_2, \ldots, \alpha_8$ together with the known and unknown values form a Network (without the point M).

Equations *(1")* and *(2)*, after substitution *(1")* in *(2)*, present n expressions in the form of

$$l_i = l_i(X_1, X_2, \ldots, X_m) \qquad i = 1, 2, \ldots, n \ . \qquad (3)$$

If X_{i0} is an approximate value of the unknown X_i, then $X_i = X_{i0} + x_i$. After linearisation of the equation *(3)* we get

$$l_i = l_{i0} + a_{i1}x_1 + \cdots + a_{im}x_m \ , \qquad i = 1, 2, \ldots, n \ , \qquad (4)$$

or, in matrix form

$$l = l_0 + A\,x \qquad (5)$$

where

$$A = \{a_{ij}\}, \qquad l = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{bmatrix}, \qquad l_0 = \begin{bmatrix} l_{10} \\ l_{20} \\ \vdots \\ l_{n0} \end{bmatrix}, \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad (6)$$

$$l_{i0} = l_i(X_{10}, X_{20}, \ldots, X_{m0}) \ . \qquad (6')$$

Corollary. From such a definition of the Network it results in, that any element of the Network (for example angle or distance) can be expressed by the function of necessary values.

Estimable functions

Theoreme 1. For a linear function (see, for example, SCHEFFE, 1959)

$$y = p^T x \qquad (7)$$

we say that it is estimable, then and only then, when vector p is a linear combination of rows of A matrix, i.e. if exists a vector q, such that

$$p^T = q^T A^T \ . \qquad (8)$$

Theoreme 2 - Theoreme Gauss-Markov. With these suppositions: $E(y) = p^T x$, $K_y = \sigma^2 I$, every function $y = p^T x$ has a linear unbiased estimate $\hat{y}$, with minimum variance, and that estimate is unique in the class of linear unbiased estimate.

Corollary T.2. If $y_1, y_2, \ldots, y_k$ are estimable functions, then any linear combination $y = \sum_i^k g_i y_i$ is an estimable function.
From *(1')*, *(1")* and *(2)* we get that any function $h = h(l_1, l_2, \ldots, l_n)$ can be reduced to the function $h = h'(X_1, X_2, \ldots, X_m)$. So, it is necessary and sufficient to express y like a function of observed values l_1, $l_2, \ldots, l_n$, i.e. that $y = y(l_1, \ldots, l_n)$.

Estimable functions in Geodetic Networks

Using previously exposed, we will give a proof that the angles and distances in Geodetic Networks are estimable functions. In that aim, we shall consider two-dimensional Network - Fig. 2. In such a Network different values can be measured. We shall consider three characteristic cases:

1. Networks with observed directions (angles),
2. Networks with measured distances,
3. Networks with measured distances and angles.

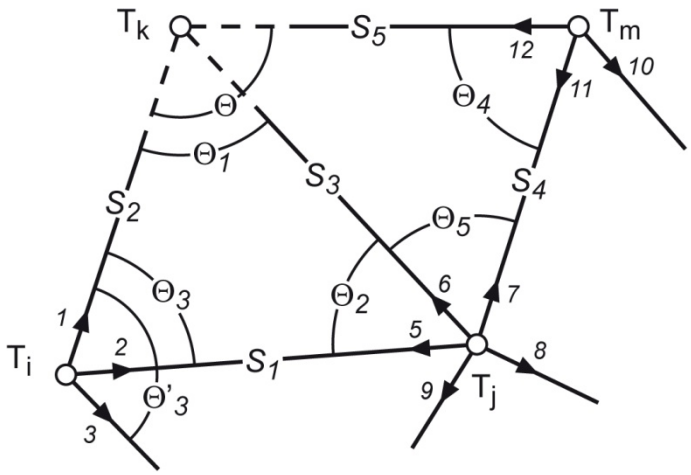


Fig. 2.

As necessary parameters we shall adopt coordinates of the points i, j, k, m, ..., of first (initial) m observed values (of course, after renumeration).
First case. Networks with observed directions (angles). The results of the observations, on Fig. 2, are marked by the numbers *1*, *2*, ...
The residual equations at the observations are

$$
\begin{aligned}
v_1 &= v_{ik} = \Delta n_{ik} + z_i - l_{ik} + l_{ik0} \\
v_2 &= v_{ij} = \Delta n_{ij} + z_i - l_{ij} + l_{ij0} \\
&\text{----------} \\
v_5 &= v_{ji} = \Delta n_{ji} + z_j - l_{ji} + l_{ji0} \\
v_6 &= v_{jk} = \Delta n_{jk} + z_j - l_{jk} + l_{jk0} \\
&\text{----------}
\end{aligned}
\qquad (9)
$$

where

$$
\begin{aligned}
\Delta n_{ik} &= a_{ik}x_i + b_{ik}y_i - a_{ik}x_k - b_{ik}y_k \\
\Delta n_{ij} &= a_{ij}x_i + b_{ij}y_i - a_{ij}x_j - b_{ij}y_j \\
&\text{----------} \\
&\text{etc.}
\end{aligned}
\qquad (10)
$$

For the angle, which is between observed directions, for example angle

$$\Theta_3 = l_{ij} - l_{ik} \qquad (11)$$

we get

$$v_{\Theta_3} = v_{ij} - v_{ik} = v_2 - v_1 \quad , \qquad (12)$$

which is obviously linear combination of A matrix row (difference between the second and first A matrix row).

For the angle which is not directly between the observed directions, for example angle Θ_1, we can establish a connection

$$
\begin{aligned}
\Theta_1 &= 180° - (\Theta_2 + \Theta_3) \\
&= 180° - \left(l_{jk} - l_{ji} + l_{ij} - l_{ik}\right) \\
&= 180° - (l_6 - l_5 + l_2 - l_1) \quad ,
\end{aligned}
\qquad (13)
$$

so we get that

$$v_{\Theta_1} = v_1 - v_2 \quad . \qquad (14)$$

For any angle in the Network, for example angle Θ, we can establish a connection

$$\Theta = \Theta\,(l_1, l_2, \dots, l_m, \dots, l_n) \qquad (15)$$

(in this particular case $\Theta = 360° - (l_2 - l_1 + l_7 - l_5 + l_{12} - l_{11})$).

If we can not establish the connection *(15)*, than, by Drozdov's definition, it is not a Network.

So (after linearisation) we get that

$$v_\Theta = g_1 v_1 + g_2 v_2 + \cdots + g_n v_n \quad , \qquad (16)$$

(for the angle Θ: $v_\Theta = v_1 - v_2 + v_5 - v_7 + v_{11} - v_{12}$).

Now, we take into consideration distances. If we don't have fixed points, we shall give the distance between two points any value. It could be, for example, the distance $S_1 = C =$ constant value. Applying sinus Theoreme on the triangle $i-j-k$, for S_3 we get

$$S_3 = C \ \sin\Theta_3 \ / \ \sin(\Theta_2 + \Theta_3) \quad , \qquad (17)$$

which, after linearisation becomes

$$v_{S_3} = a_1 v_1 + a_2 v_2 + a_5 v_5 + a_6 v_6 \quad , \qquad (18)$$

where a_i are some coefficients.

Applying gradually the sinus Theoreme on the adjacent triangles, any distance can be expressed as a function of observed directions l_1, l_2,..., l_n. If such connection can not be expressed then, by Drozdov, it is not a Network.

Second case. Network with measured distances. Now we have measured distances S_1, S_2, ..., (Fig. 2). Applying cosine Theoreme to triangle $i-j-k$ we get

$$\Theta_1 = \arccos((S_2^2 + S_3^2 - S_1^2)/2\, S_2 S_3) \quad , \qquad (19)$$

from where, after linearisation we have

$$v_{\Theta_1} = b_1 v_{S_1} + b_2 v_{S_2} + b_3 v_{S_3} \quad , \qquad (20)$$

where b_1, b_2, ... are constant coefficients.

Gradually applying cosine Theoreme to the adjacent triangles we conclude that any angle can be expressed by a function of measured distances. On the contrary (by Drozdov) it would not be a Network. Further, we shall consider the distance $i-m=S$. First, through the cosine Theoreme the angles Θ_2 and Θ_5 can be expressed by S_1, S_2, S_3, S_4 and S_5, and then, again apply the cosine Theoreme to the triangle $i-j-m$, we can express the distances with measured elements S_1, S_2, S_3, S_4 and S_5:

$$S = S\,(S_1, \ldots, S_5) \qquad (21)$$

where

$$v_S = c_1 v_{S_1} + c_2 v_{S_2} + \cdots + c_5 v_{S_5} \quad . \qquad (22)$$

Gradually applying this procedure we get to the conclusion that the distance between any two points can be expressed by measured distances, on the contrary it would not be a Network.

So, we get to the conclusion that the angles and the distances are estimable functions in Geodetic Networks.

Third case. Networks with measured angles and distances. This case includes the two previous ones, so the conclusions from them are valid.

In leveling Networks the height differences would be estimable functions (they could match the distances in two-dimensional Networks) which could be established like in the previous cases.

Conclusion

From the mentioned cases we get to the conclusion that the angles and the distances in Geodetic Networks are estimable functions.

REFERENCES

ДРОЗДОВ, Н. Д., 1972: Линейная алгебра в теории уравнивания измерений. Недра, Москва, стр. 17-19.

SCHEFFE, H., 1959: The analysis of variance. New York.

DATA PROCESSING AND ADJUSTMENT

Knud PODER
Charlottenlund, Denmark

ABSTRACT

The paper deals with data processing in connection with control networks, supplemented by a few special details of importance for such programmes. A possible role of the control networks as scaling elements of fundamental networks and possible methods for a reasonable treatment of control networks in connection with fundamental networks of poor scale ends the paper.

Contents.

1. Introduction.
2. EDP programmes for Control Networks.
3. Selected Problems.
4. Control Networks and Fundamental Networks.
5. Concluding Remarks.

1. Introduction.

Control networks before the electronic age of geodesy relied heavily upon the fundamental (first order) networks for location, orientation, and scale. Computation was frequently carried out by inclusion of one or a few stations at a time with simplified algorithms in order to make the task feasible and because the obtainable accuracy was much lower – also in the fundamental network.

The obtainable accuracy now is increased by nearly one order of magnitude not least due the use of EDM, and scale is frequently observed directly in the control network. It has even been suggested to dispense with a classic fundamental network and build a network from only a relatively short-sided trilateration network, Weigel, 1932 (not based upon EDM), Töpfer, 1956, Gerke and Pelzer, 1970, Schädlich, 1970. Location and thus indirectly orientation may with present and coming satellite techniques be observed directly for a control network.

The numerous and frequently very good scale information observed in control networks may actually give a problem of inconsistency if such a control network should be adjusted to a classical fundamental network with too old and inaccurate scale, but the availability of EDP gives a remedy for this problem.

When EDP was introduced in geodesy, most geodesists probably saw it as the tool for handling normal equations and maybe also doing the more tedious computing work in accordance with the usual computation forms established in the institution. Later came the idea that the results could be filed and retrieved by means of EDP and at last we realized, that data should be manipulated by EDP as soon as the observations were available.

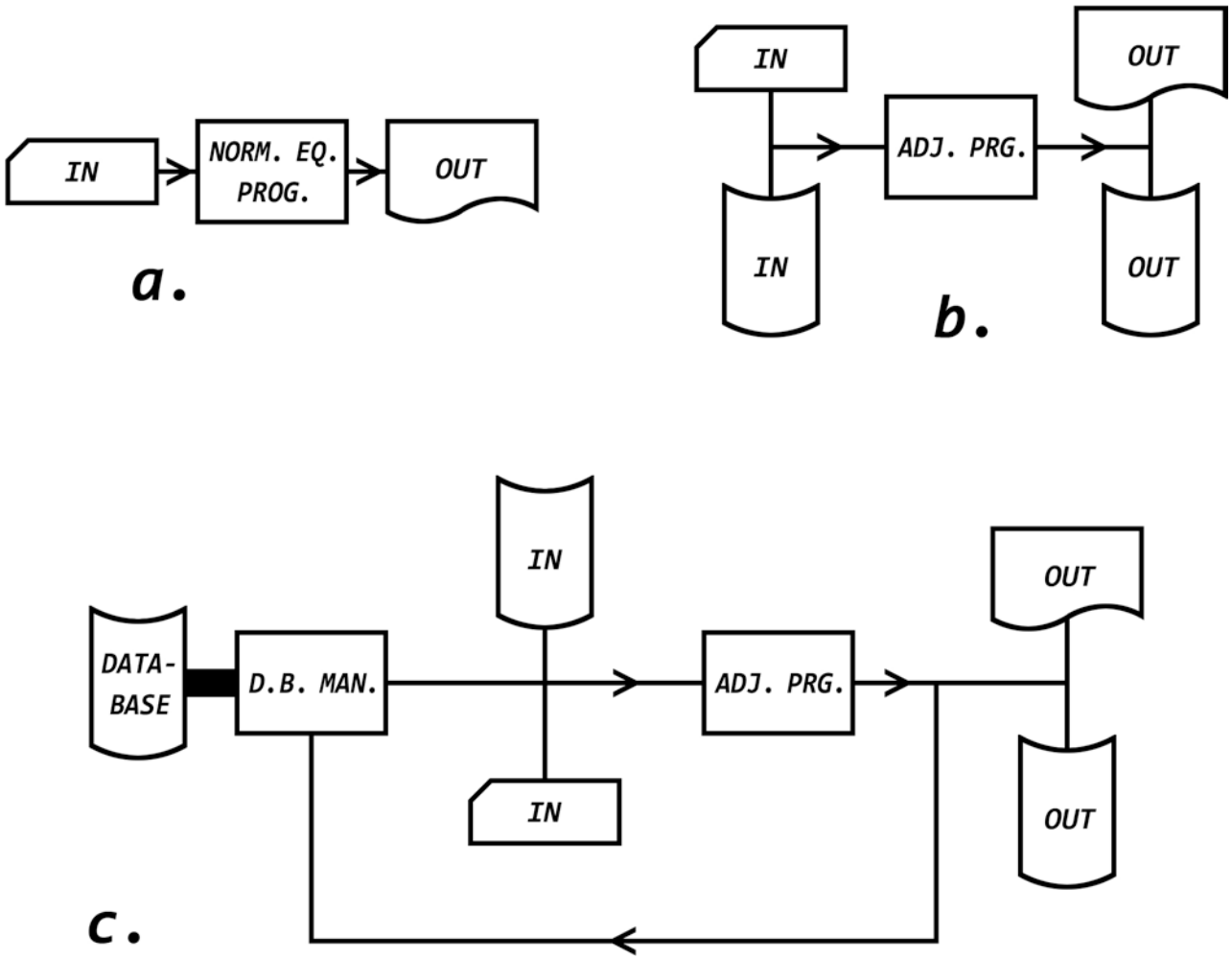


Fig. 1. EDP in the evolution of network adjustment.

Fig. 1 shows the evolution in three steps. Fig 1 a is the simple system were punched cards or paper tape with normal equations are fed in for computing and results are printed out on a line printer or a typewriter. Fig. 1 b shows the next step where data can be fed in for network adjustment from an external reader or from a disc file, adjusted and output on an external printer or on a disc file. Fig. 1 c shows a network adjustment programme system, where the majority of data come from or go to a data base.

We shall in the following concentrate us on the common area of data processing and network adjustment, but end with suggesting how by virtue of data processing one can build a more homogenous network than was feasible before the electronic age.

2. EDP Programmes for Control Networks.

If builders built buildings the way programmers wrote programs, then the first woodpecker that came along would destroy civilization.

(Weinberg’s second law.)

The first attempts of programming geodetic work were possibly not made after the principles and requirements given here, but one has to realize that the problems look different from the points of view of the user and the programmer. The user asks for processing facilities by specifying what he wants done. The programmer supplies programmes to meet these specifications and maintains the programmes.

Table 1 illustrates the points of importance in this communication (which in practice may be much more diffuse and with many details for the actual problem).

User	Programmer
1. Reliability	1. Clear problem definition
2. Versatility	2. Modularity
3. Data base access	3. Well-structured
4. Information	4. Maintainability

Table 1. User’s and programmer’s views.

2.1 The User’s Demands.

Let us for a moment disregard Weinberg’s second law and look on the concept of programme reliability. Reliability means that the programme should ensure that if correct data was available then the results would be correct, and if errors detectable by virtue of the redundancy were present then clear signals of the errors should be given. E.g. coordinates in a system differing from the one selected for the adjustment should be rejected or translated to the proper system, undefined coordinates through missing observations should be flagged clearly, erroneous observations should be listed, etc.

Versatile processing means that the user has a free choice of coordinates and datums, that any reasonable kind of observations can be used, and that the compilation of the data needed is

automatic, so that the user just has to specify the names of the files containing the data and does not have to worry about details. The processing must be able to continue even with large observation blunders and/or missing observations, so that as many errors and omissions as possible can be realized in a single run.

The access to a data base is essential for supporting a safe retrieval of needed data and filing of results. This is probably the most recent experience with data processing in geodesy.

The processing must apart from standardized results also provide the user with a more detailed information if it is desired to clarify problems. This will mean plotting of selectable parts of a network, printing (some) inverse matrix elements, listing of observations and their corrections, etc.

2.2 The Programmer's Work.

A clear definition of the overall task is not so difficult for a network adjustment (especially if one has made some mistakes the first times and avoids persisting in making errors). As an example one could specify a programme giving a least squares adjustment of networks in 1, 2, or 3 dimensions using corresponding observation data. This would give a versatile programme. Likewise, one would need a specification of the capacity of the programme counted as the usual size of the networks to be handled. However it is evident, that one should not specify more than actually needed, but it may clarify the programming if the structure is general enough to permit extension to more comprehensive tasks. The actual size of the programming work may range from a modest programming of a desk-top minicomputer to 10 man-years for a large comprehensive system. In practice most systems to-day will probably not start from scratch, but rather be based on existing systems.

Modularity and structuring are very essential for an orderly and economic development of the programmes. Modularity means that the programmes are composed of rather independent modules, which of course again may consist of modules recursively. It is a great help that such modules regarded as system components can be tested rather independently and they may frequently be used in the other programmes of the entire lot of software at the installation. Structuring of the programmes means that the programmes carefully branch out in parallel paths for the different conditions due to coordinate system and observation kinds and follow the same path when the same treatment is required. Structured programming is described in Dahl, Dijkstra, and Hoare, 1972 in general and an example of a structured programme can be found in Poder and Madsen, 1978.

The maintenance may be regarded as an euphemism for correcting errors found in the system, but it may also mean a further development of the system, expansion of capacity, etc. Modularity means that the insertion of better components of software as

replacement for weak or erroneous ones can be made safely. A clear structuring will at the same time support the verification of the proper effect of the replacing module. The maintenance work on a medium size adjustment system may easily attain the size of about one man-year per year.

Let us finally turn Weinberg's second law on again. It is surprising how long time an error may live in a programme. It is not unusual that the special conditions – not expected by the programmer – first will turn up after years of satisfactory running.

2.3 Scope and Size of Control Network Adjustment Systems.

The actual size and capacity of such a system, which jointly with the adjustment programme(s) includes a certain number of auxiliary programmes for preparation, post-treatment and general maintenance service of the data will of course depend upon the scope of the institution using the system. If one never adjusts more than 10 stations at one time, a micro-computer will probably meet all demands, but a national survey institution may need the capacity of several thousands of stations computed simultaneously, see e.g. Schwarz, 1978, Poder and Madsen, 1978, Knight and Mepham, 1978, Harada, 1980. It is in the latter case quite reasonable to take advantage of the overall capacity to use a programme with a high numerical quality, not least because the programme then also would be applicable for computations of the fundamental network with high precision. In short: the extra time for precise computation is immaterial if a reasonable large and fast machinery is available. The available capacity can also be used for photogrammetry tasks, and such a system will be very useful for the probably increasing joint adjustments of photogrammetric and ordinary geometrical geodetic observation data.

2.4 Staff.

It is a general experience that the communication between users and programmers may be difficult. This may be due to the infancy stage of EDP, and later generations may wonder why it was so. One of the possible remedies for the present situation is to use persons which have some background in both the actual application (geodesy, surveying) and in programming jointly with the ordinary users and programmers.

3. Selected Problems.

Before turning to the changing situation of control networks, some problems in the adjustments systems should be mentioned. The

survey is by no means complete, but the problems get a good deal of attention in the present stage of art.

3.1 Input Data Compilation.

A control network of 2-3000 stations may have 10-25000 direction observations and 500-2000 distance observations. It would be very difficult to collect such an amount of data manually without making errors or omissions, so obviously an automatic compilation would be preferable. The compilation of the given coordinates and approximate values of the new coordinates is also a major task in this case.

An apparently simple method would be to start the processing with a list of station identifiers (numbers or names) of the stations involved in the process. It is then possible to compile the observation data from the files of the data base, guided by the defining list of stations. Coordinates can likewise be collected, possibly after a transformation if they exist in some coordinate system in the files. Missing given stations can be flagged and reported, and it is possible to find approximate coordinates for the new stations, e.g. if the approximate orientation of directions and/or distances are known. Stations failing to meet this can then be flagged and must then be supplied with approximate coordinates in a manual way.

However, there are tasks where this method fails, e.g. in photogrammetry and in levelling, where the identification and presence of auxiliary points are not clearly visible to the user. One could in principle then scan the observation files to get a list of station identifiers (the multiple occurrences can easily be removed), but the problem is that this will mostly suggest all stations in the observation files as candidates for the adjustment. It seems therefore, that a better approach would be to let the user start the definition and then a dialog with this list and the observation may give suggestions for which stations should be included in the task. Special parameters as camera constants, refraction coefficients etc. may also require a more active dialog when the input data are compiled by the adjustment programme.

3.2 Blunder Detection.

It cannot be avoided that the observation data in an adjustment contains errors. Such errors may concern us in three ways. They may be so large, that the convergence of the adjustment is threatened or they may be so much smaller, that they can make the coordinates less correct. Finally the errors may be just so much larger than the normally distributed observations, that one has to use some error snooping method to detect them. Statistical methods has of course no interest for the really large errors of, say, degrees or minutes of arc, due to simple human

mistakes. Such error will mostly be detected in a validating on-line input system, whereas the second type of, say, minutes or some tens of seconds may be permitted to pass the validation. (One of the safest ways to miss the truth is to reject all information in disagreement with established values). A purely pragmatic method as follows will handle the two first cases very well. If one lets the coordinates found by the first run through the adjustment be preliminary coordinates for further runs, observations with large deviations may be flagged to the user and used with a reduced weight in the adjustment. If the deviation was due to bad preliminary coordinates, the signals will cease in the subsequent passes when the coordinates become good enough, and if the deviation was due to a blunder, the error may come out more and more clearly, when its effect on the determination of coordinates is reduced due to the reduced weight applied to it. It must be emphasized, that the purpose of this method is not an improvement of the results as such, but only a method to keep the adjustment alive instead if diverging and leaving the finding of the blunder to the user.

The statistically based search for blunders has been pioneered by Prof. W. Baarda at Delft, Baarda, 1967, 1968. Later contributions come from Dr. Allen Pope, NGS, USA, Pope, 1976. Quite recently some attempts have been made at AUC to use a statistically based method on the data during the active adjustment. As the adjustments include increasingly larger networks, such statistical methods become really important, because human inspection of the observations becomes more and more impossible and ineconomical.

3.3 Normal Equations.

The normal equation matrices of geodetic networks are mostly very sparse. It is therefore very reasonable to take advantage of this by suitable methods. The method of indexing each element of the matrix is not suited to array processors, which may operate on whole columns of a matrix with one instruction, and the fill-in occurring will be a nuisance when a paging method is applied for normal equations of a size requiring such a method. The so-called datamatic blocking, Andersen and Krarup, 1962, Poder and Tscherning, 1973, Poder, 1978, where the matrix is mapped columnwise on blocks, which are paged to and from the backing storage when required seems to have a capacity of 5-10000 unknowns when the matrix is reasonably sparse. The method presupposes for an efficient storage, that the equations are ordered so that the elements are placed reasonably near the diagonal of the matrix. This may be achieved by suitable algorithms using the observation structure to predict a reasonably good ordering, see e.g. Mark and Poder, 1981, for a review, or Snay, 1976, but all applicable methods cannot find the optimum. Such an ordering saves both space (good methods can reduce the space requirement to less than 10 percent of what is required for a full upper triangular matrix) and computing time (mostly to between 1 and 10 percent of the time required for a full matrix). The coefficients are packed

columnwise near the diagonal (some zeroes are unavoidably imbedded), so that the addressing for an array processor is very easy.

The limitation of the method comes from the same reason that gives it its capacity, the blocking on the backing storage. Each block must be formed one at a time (giving very nice restart possibilities), but this presupposes a complete scanning of the likewise blocked observation equations.

The way out of this problem, which appears when the number of unknowns exceeds maybe 10000 is to use Helmert-blocking, Wolf, 1978, Meissl, 1980, Poder, 1981. It is most likely that this method will only rarely be applied, unless one wants a joint adjustment of an entire national control network.

The Cholesky method for reducing the symmetric normal equations matrix into a triangular matrix seems to have almost advantages over other methods, because the matrix is symmetric and positive definite, Wilkinson, 1965. Lacking observations may make the matrix singular. It is however easy to realize this situation in the algorithm, clearly indicated by the loss of significant bits when the diagonal terms of the reduced matrix are formed. The algorithm can then neglect the corresponding rows and columns in a very simple way (if the inverse diagonal element is stored, a zero will do the trick). At the same time the defect can be flagged to the user, so that the situation is harmless for the continuation of the processing and informative for the user.

The numerically very promising methods based upon orthogonal transformations of the observation equation matrix, have up to 8 times as many operations as Cholesky and much more paging operation if the matrix is large.

4. Control Networks and Fundamental Networks.

4.1 Joint Networks.

As mentioned in the beginning control networks are no longer solely dependent upon a fundamental network. The do actually have distance observation which with proper precautions will produce a better scale than obtainable by invar bases (when extended and propagated) or by direct observation of the fundamental network sides with EDM. We may note following facts about EDM over short ranges of 2-10 km.

1. It is easier to obtain a good model of the atmosphere from terminal observations, when the line is short.

2. Observations in a control network are mostly spanning a much bigger time interval than the one used for EDM observations in the surrounding fundamental network.

3. Systematic errors in the EDM-instrumentation being harmless for the single distance, will add up heavily and may give a large error in the scale of the fundamental network, if the scale is propagated to the latter.

We may sum this up in the following conclusions. The distances by EDM for a real improvement of the scale of the fundamental network should short enough for a good modelling of the refractive index and not so short, that the zero error bias adds up to many times. The new experimental fundamental network in DK uses sidelengths of about 10 km, but it is not known if a better choice could be made. The exploitation of the scale information is made very simply by adjusting the entire network form the fundamental one down to the network where the desired distances are observed. This means in the studied case about 900 stations, being well within the capacity of the programmes used for the recomputation of the entire control network blocks.

The suggested exclusive use of shorter distances as a rational replacement of the classical fundamental network poses the same requirement of a high capacity of the adjustment system, but the method is certainly feasible from a point of view of computation.

However, there are very good operational reasons for maintaining direction observations jointly with distances. The choice of netform elements is much more flexible if both directions and distances can be observed.

4.2 Adjustment to Fundamental Networks.

It is reasonable to ask for coordinates, which are as good a transformation of the observation data as possible, but for many purposes changes in coordinates with small amounts may be just troublesome and useless. Conversely the coordinates of the fundamental network may be so much in error, that the better scale from the local control network cannot be utilized, if one has to adjust to the fundamental one. In the case of best possible conservation of the old coordinates, a scale factor on all distances may be found jointly with the coordinates of the control network. This actually reduces the distance observations to distance ratio observations. The situation of best possible control network coordinates may be met by assuming that the fundamental network coordinates are observations with a finite weight matrix (possibly with nonzero off-diagonal elements). It is then even possible to locate the poorest fundamental coordinates by blunder detection methods.

5. Concluding Remarks.

EDP has been used in geodesy and land surveying almost as soon as it became available. The status is that the computing capacity available permits computations in a scale one could never attain by manual methods. There has been collected a fair amount of experience in the applications and EDP seems to be one of the most valuable tools ever given to the geodetic community.

References.

Anderson, C. and T. Krarup: Linear Equations. Selected Numerical Methods, ed. Chr. Gram, Copenhagen, 1962.

Baarda, W.: Statistical concepts in Geodesy. Netherlands Geodetic Commission, New Series, Vol. 2, No. 4, 1967.

Baarda, W.: A Testing Procedure for Use in Geodetic Networks. Netherlands Geodetic Commission, New Series, Vol. 2, No. 5, 1968.

Dahl, C.-J., E.W. Dijkstra, and C.A.R. Hoare: Structured Programming. Academic Press, 1972.

Gerke, K. and H. Pelzer: Netzverdichtung oder Netzaufbau aus Funktionen gemessener Größen? Allg. Verm. Nachr., No. 8, 1970.

Harada, Takehisa: Universal Program for Adjustment of any Geodetic Network. Journ. Geod. Soc. of Japan, Vol. 26, No. 3, 1980.

Jewell, J.G.: Ordering systems of equations. Journ. Soc. of Industrial and Applied Mathematics, Vol. 9, 1961.

Knight, W. and M.P. Mepham: Report on Computer Programs for Solving Large Systems of Normal Equations. Proc. Second Int. Symp. on the North American Geodetic Networks, Washington, 1978.

Mark, A. and K. Poder: Ordering and Dissection of Geodetic Least Squares Equations. Proceedings Symposium on Geodetic Networks and Computations, Munich, Aug. 31 – Sept. 5, 1981.

Meissl, P.: A Priori Prediction of Roundoff Error Accumulation in the Solution of a Super-large Geodetic Normal Equation System. NOAA Prof. Paper 12, NGS, NOAA, Rockville, Md., 1980.

Poder, Knud: Processing of Datamatically Blocked Normal Equations. Proc. Second Int. Symp. on the North American Geodetic Networks, Washington, 1978.

Poder, K.: Top-down Strategy for Helmert Blocking. Proceedings Symposium “Management of Geodetic Data”, København, August 24-26, 1981, pp. 326-340, 1981.

Poder, Knut and F. Madsen: Status of the new Adjustment in Greenland and the Adjustment System of the Danish Geodetic Institute. Proc. Second Int. Symp. on the North American Geodetic Networks, Washington, 1978.

Poder, K. and C.C. Tscherning: Cholesky’s Method on a Computer. The Danish Geodetic Institute Internal Report No. 8, 1973.

Pope, Allen J.: The Statistics of Residuals and the Detection of Outliers. NOAA Tech. Rep., NOS65, NGS1, 1976.

Schwarz, C.S.: TRAV 10 horizontal network adjustment program. NOAA Techn. Mem., NOS, NGS-12, Rockville, Md., 1978.

Schädlich, M.; Das Verfahren der kurzen Schritte – ein Beitrag zur Modellierung geodätischer Übertragungsmessungen. Vermessungstech. 18, No. 3, 1970.

Snay, R.E.: Reducing the profile of sparse symmetric matrices. Bull. Geod., Vol. 50, No. 94, 1976.

Töpfer, F.: Ableitung von Winkeln übergeordneter Dreiecke aus Dreiecksnetzen. Veröff. des Geod. Inst. der TH Dresden, Heft 1, Dresden, 1956.

Weigel, K.: Triangulierung ohne Winkelmessung. Verhandlungen der Baltischen Geodätischen Kommission, Sechste Tagung, Warschau 1932, Helsinki, 1933.

Wilkinson, J.H.: The algebraic Eigenvalue Problem. pp. 229-249, Oxford, 1965.

Wolf, Helmut: The Helmert Block Method – its Origin and Development. Proc. Second Int. Symp. on the North American Geodetic Networks, Washington, 1978.

ON THE STOCHASTICITY OF ALIGNMENT-FUNCTIONS FOR AUTOMATED TRACK MAINTENANCE MACHINES

Herman QUEE

Utrecht, the Netherlands

ABSTRACT

The maintenance of railway tracks is periodically carried out by hydraulic tamping machines (main tracks in the Netherlands about 1-2 times per year). Modern tamping machines are since 1975 equipped with a servo-guided laser-scanner for use in straight tracks. The Netherlands Railways have introduced a system by which the method of laser-guided tamping can also be applied in curves (circles and clothoides).
The surveying department established in each curve a local co-ordinate system in which the co-ordinates of constraint points in the track and reference points for the laser-beam are determined. Then the ideal alignment for the tracks is designed, consisting of straight lines, clothoides and circles. From these functions setting-out-distances for the laser-beam and guiding data for the tamping machine are computed and via cassette tapes transferred to a microprocessor on board the machine.

All quantities mentioned are derived from stochastic co-ordinates in a local system; this applies also to be alignment functions! The internal system of the tamping machine requires a certain relative precision, whereas the constraints in the tracks (fixed points on bridges, road-crossings, switches etc.) require a certain reliability of these derived quantities.

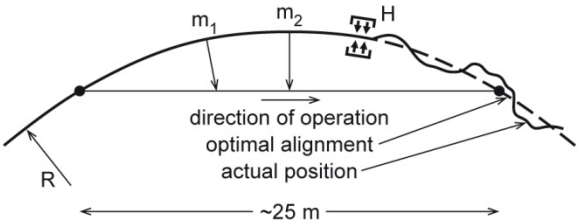
It will be proved that the stochasticity of a point on the alignment elements (circles etc.), formulated according to BAARDA's theory of S-transformations, is identical to the stochasticity of a network point at the same place. This means that precision and reliability of short distances and angles are invariant (independent from the choice of "base points" for the network-adjustment).
Angles and short distances are the variables on which the above-mentioned requirements for precision and reliability are based, which means that also the requirements are invariant.

1. Introduction : track maintenance machines.

Track geometry is subject to large deformation stresses, not only on their horizontal and vertical positioning, but also on the canting, caused by the effects of railway traffic. The rate at which track deformities occur, is dependent on the condition of the ballast bed and the subsoil on the one side, and the train traffic (tonnage, speed, number of axles) on the other.
Track deformities are corrected periodically by various types of hydraulic track maintenance machines, see fig. 1, which ensure that the track is

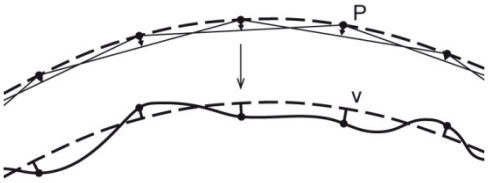
returned to its optimal position both horizontally and vertically, and which can also renew the cant (in track curves). The method by which these "tamping machines" carry out their correction work is illustrated in fig. 2.

A steel wire is stretched between front and rear of the machine, and is used as a reference line for the process control unit.



The leading point is guided over the desired (optimal) alignment; the trailing end plus the measuring point m_1 are to be found over the already corrected track section. From the measured distance m_1 and the imputed curve radius R, the desired alignment at m_2 is computed by a microprocessor. Subsequently the displacement of the track and the improvement of the cant are carried out by hydraulically operated machinery. This process is repeated from sleeper to sleeper, reaching a speed of about 0.5 km/hour.

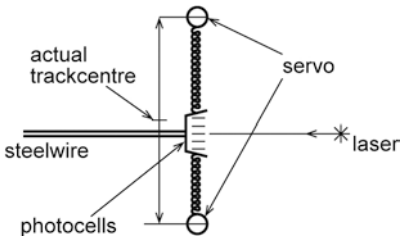
The optimal position, over which the leading point is guided, was determined up to now by one of the classical methods of measurement of "versines", p, from which the displacements v between actual and optimal position are derived in each measurement point (see fig. 3).



The use of this method of track re-alignment has many practical and theoretical disadvantages, i.e.:

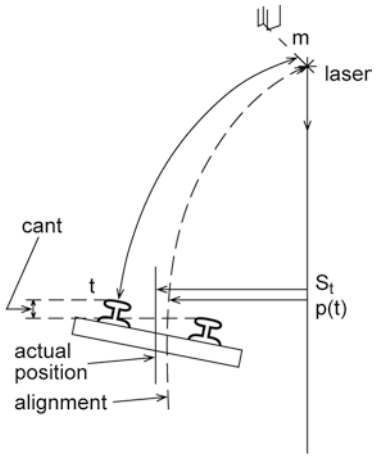
1. The labour intensive versine measurements must be repeated for every maintenance period, because the displacements v are rendered useless after the track correction.
2. The optimal alignment is badly defined : a radius can have an inaccuracy of 100 to 250 m, and the position and length of transistionary curves can have an inaccuracy of 10 to 20 m.
3. During the operations much manual work is needed to impute the data as mentioned above in the microprocessor.

An important technical development in mechanical track maintenance, about 1975, was the application of the laser-beam as adapted for the guidance of the leading point of the tamping machine. The laser is erected in the track centre, and the emanating beam is then intercepted by a row of photocells, which are positioned on the front end of the machine. The cells follow the laser-beam through a servomechanism, thus holding the leading point of the wire continuously in the laser-beam (= the optimal track alignment : see fig. 4.

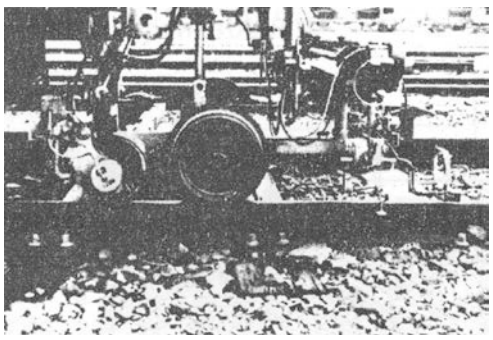


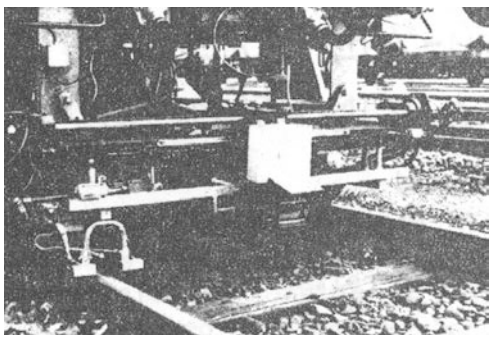
This system is only suitable, of course, in straight track sections. However, the Netherlands Railways has, in the period 1978-1982, worked closely with the manufacturer of the tamping machines to develop a system that also allows the use of laser guided machines in curves. This system is called DRIVER (Driving by external reference). The working principle is illustrated in fig. 5 : The laser-beam is set out in relation to fixed points next to the tracks, generally these are catenary masts, in a chord of the optimal curve (the alignment). The tamping machine is now equipped with the following apparatus : (fig. 6).

1: A tachometer; this measures the curve length t from the reference point.
2: A row of photocells, which guides itself in the laser-beam by a servo-mechanism and thus measures the distance S from the existing track position to the laser-beam.
3: A microprocessor, which reads the guidance data, previously determined with surveying techniques, from a cassette tape and computes continuously the distance p (t) between the curve alignment and the chord.



The difference $p(t)-S_t$ is passed to the automatic guidance mechanism, which shifts the track over this distance.





2. The geodetical activities for laser-guided track maintenance.

The surveying department has three tasks with mechanical track maintenance:

1. The production of a field of co-ordinate points in every railway curve. This implies the measurement and adjustment of a network, consisting of reference points (catenary masts) and points of the actual track position; computation of co-ordinates in a local co-ordinate system.
2. The design of the optimal track position, the “alignment”, in this co-ordinate system and the computation of deduced variates for the setting out of the laser-beam and the guidance of the machine.
3. The storage of all information in a database and the transmission to the DRIVER-system on board the tamping machine.

From the method of mechanical track maintenance, as explained in chapter 1, criteria are deduced for the accuracy of the data for the DRIVER-system, and subsequently for the accuracy of the co-ordinates; the complete process is given in next scheme, in which the terminology and notation as introduced in BAARDA (1967) and (1968) have been used.

process	accuracy		criteria
	precision	reliability	
measurement	σ_r ; σ_s	alternative hypothesis	
	↓	↓	
network adjustment co-ordinates	$(\underline{g})$	$\widetilde{\nabla Y}^{\alpha}$ →	the layout of the network
	↓	↓	↑
alignment functions (parameter p^j and deduced quantities m^i)	σ_{p^j} ; σ_{m^i}	$\widetilde{\nabla p}^j$; $\widetilde{\nabla m}^i$	1 constraint 2 transfer 3 reference 4 machine-control
setting out of laser: M^i and guidance of machine: $R(p^j)$	σ_{R^j} ; σ_{M^i}	$\widetilde{\nabla R}^j$; $\widetilde{\nabla M}^i$	

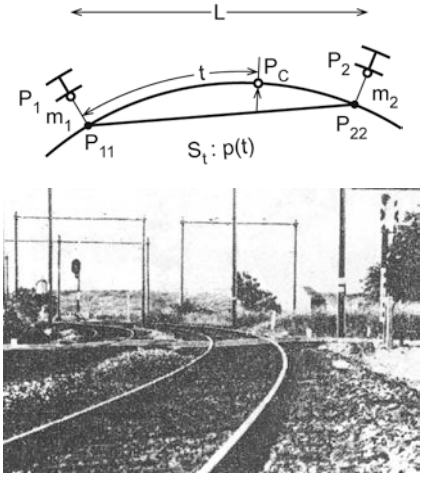
The four criteria mentioned in this scheme are : (stochastic quantities are underlined)

1. The constraint criterion (fig. 7).

The criterion variate c_c reads: (approximate formula)

$$\underline{c}_c \approx \left(1-{}^t/_L\right)\underline{m}_1 + {}^t/_L\,\underline{m}_2 + \underline{p}(t) + -\left(1-{}^t/_L\right)\underline{M}_1 - {}^t/_L\,\underline{M}_2 - \underline{S}_t \ .$$

$p(t)$ = computed
S_t = measured by laser scanner



Requirement for reliability :

$$\left|\widetilde{\nabla c}_c\right| < 8 \text{ mm} \ .$$

Points : P_1, P_c, P_2 : network points
P_{11}, P_{22} : projections on alignment

2. The transfer criterion (fig. 8).

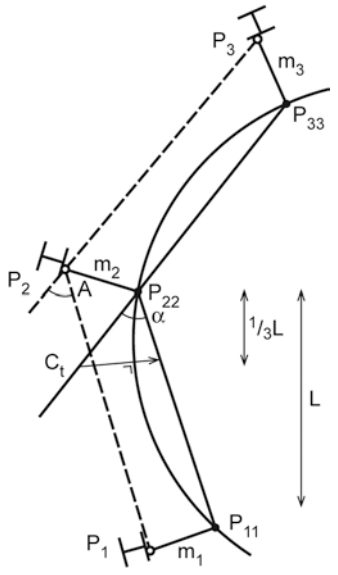
When the operation has been carried out over 2/3 of the distance between two reference points, the laser is transferred to the next position. P_{33}; the scanner is transferred over the distance c_t, which is the criterion variate :

$$\underline{c}_t \approx {}^1/_3\,L\left(\underline{\alpha}-\underline{A}\right) = {}^1/_3\,L\,\frac{\underline{m}_1 - \underline{M}_1 - 2\left(\underline{m}_2 - \underline{M}_2\right) + \underline{m}_3 - \underline{M}_3}{L}$$

Requirement for reliability :

$$\left|\widetilde{\nabla c}_t\right| < 10 \text{ mm} \ .$$

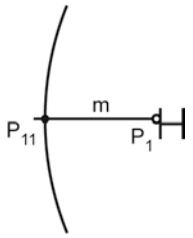
Points : P_1, P_2, P_3 : network points.
P_{11}, P_{22}, P_{33} : projections on alignment



3. The reference criterion (fig. 9).

The criterion variate c_r reads :

$$\underline{c}_r = \sqrt{\left(\underline{X}_1 - \underline{X}_{11}\right)^2 + \left(\underline{Y}_1 - \underline{Y}_{11}\right)^2} - \underline{M}_1$$



Requirement for reliabiliaty :

$$\left|\widetilde{\nabla c}_r\right| < 5 \text{ mm}$$

Points : P_1 en P_{11}.

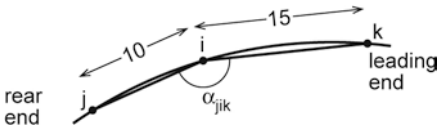
4. The criterion for the machine control (fig. 10).

The control unit of the machine requires a certain level of precision; the criterion variate reads:

$$\underline{C}_m = 15\,\alpha_{jik}$$

Requirement for precision :

$$\sigma_{\underline{C}_m} < 2 \text{ mm}$$



Conclusion :

The points on the alignment play, by all four criteria, an important role. Therefore a functional model has to be made for the alignment function, with which the stochasticity of the points on the alignment can be derived from the stochasticity of the measured points (from which the alignment is computed).
Considering that these last points have to be defined as S-co-ordinates, it will be obvious that the theory of S-transformations, BAARDA (1973), is indispensable.

3. The computation of the alignment.

An alignment is built up from connecting straight lines, clothoides and circles. In railways there is always a clothoide between a straight line and a circle, and often between two circles of considerably differing radius. In the field of application as viewed here, curves of more than 100 years old are idealized. This means that in many cases the circle must be split up into smaller units with slightly differing radii, for example 908, 897, 904 m. etc.. The alignment to be computed has to comply with a number of requirements :

- constraint points,
- start and finishing points must connect to the existing track axis,
- the distance between alignment (optimal situation) and existing track may not exceed 8 cm, otherwise also the catenaries must be altered.

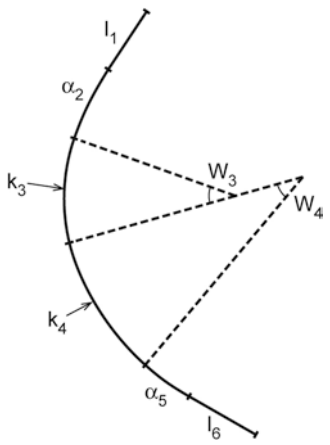
Different methods have been used to find a computation algorithm for the solution of alignment parameters, i.e. : radii and lengths of alignment elements (i.e. : straight lines, clothoides and circles).
As far as the functional aspect is concerned, spline functions are used (RIJKSWATERSTAAT, 1968), and for the optimalisation aspect one made use of certain methods of operational research (DEUTSCHE BUNDESBAHN, 1971).
The Netherlands Railways found a solution to this problem in 1974. The alignment parameters are resolved iteratively from a large number of linearised condition equations, with misclosures q^i. The parameters p^j are curvatures (= 1/radius) and lengths of the elements, see fig. 11.

$p^j = 1$: length of straight line

" $= w = 1/R$: centre angle of circle

" $= k = 1/R$: curvature of circle

" $= \alpha = \frac{dk}{dl}$: constant of clothoide



Remark : the length of a clothoide is not an independent parameter because it follows from α and the curvature k of the connecting circle:

$$l_{cloth} = \frac{k}{\alpha} = \frac{1}{\alpha R}$$

The misclosures q^i of the condition equations are functions of the parameters and of co-ordinates and bearings ; see fig. 12.

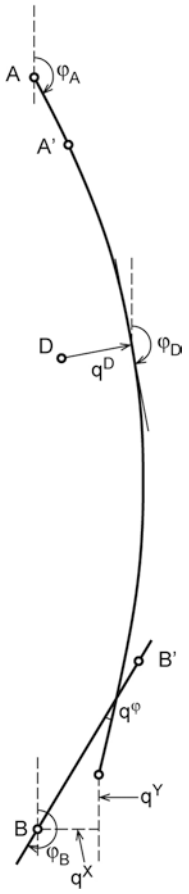
$$q^1 \equiv q^\varphi = \varphi_{AA'} - \varphi_{B'B} + I\varphi_1 + I\varphi_2 + I\varphi_3 + \ldots + I\varphi_m =$$

$$\downarrow \quad \downarrow \quad \downarrow \qquad \downarrow$$

$$=0 \quad =\frac{k_3}{2\alpha_2} \quad =W_3 \qquad =0$$

$$= \varphi_{AA'} - \varphi_{B'B} + I\varphi\left(\ldots p_o^{j(1)}\right) + I\varphi\left(\ldots p_o^{j(2)}\right) + \ldots + I\varphi\left(\ldots p_o^{j(m)}\right) \qquad (1)$$

$$q^2 \equiv q^X = X_A - X_B + IX\left(\ldots p_o^{j(1)}, \varphi_{AA'}\right) +$$
$$+ IX\left(\ldots p_o^{j(2)}, \varphi_{AA'} + I\varphi_1\right) +$$
$$+ IX\left(\ldots p_o^{j(3)}, \varphi_{AA'} + I\varphi_1 + I\varphi_2\right) +$$
$$\vdots$$
$$+ IX\left(\ldots p_o^{j(m)}, \varphi_{AA'} + I\varphi_1 + I\varphi_2 + \ldots + I\varphi_{m-1}\right) \qquad (2)$$

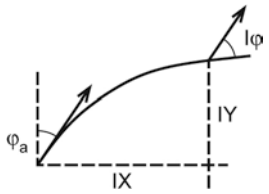


$$q^3 \equiv q^Y = Y_A - Y_B + IY\left(\ldots p_o^{j(1)}, \varphi_{AA'}\right) + \ldots$$
$$\ldots + IY\left(\ldots p_o^{j(m)}, \varphi_{AA} + I\varphi_1 + I\varphi_2 + \ldots + I\varphi_{m-1}\right) \qquad (3)$$

$$q^4 \equiv q^{D_1} = \cos\varphi_{D_1}\left[X_A + IX_{A\to P_1}\left(\ldots p_o^j, \varphi_{AA'} + \sum I\varphi_i\right) - X_{D_1}\right] +$$
$$\vdots \qquad - \sin\varphi_{D_1}\left[Y_A + IY_{A\to P_1}\left(\ldots p_o^j, \varphi_{AA'} + \sum I\varphi_i\right) - Y_{D_1}\right] \qquad (4)$$

$q^{3+n} \equiv q^{D_n} = \ldots$ similar to ↴

m = number of elements
n = number of constraint points
$p^{j(1)}$: parameters in element 1.
p_o^j : approximate values.
IX = interval X
IY = interval Y
Iφ = interval φ
} of elements; (see fig. 13)



Example of interval X :

$$IX_{i(circle)} = \frac{1}{k_i}\left[\cos\varphi_a - \cos(\varphi_a + w_i)\right]. \qquad (5)$$

Differentiating of these condition equations to the parameters p now gives:

$$(\Delta q^i) = \left(\frac{\partial q^i}{\partial p^j}\right)(\Delta p^j) = (Q^{ij})(\Delta p^j). \qquad (6)$$

$$i = 1 \ldots n+3 \qquad j = 1 \ldots n+3$$

Consequently the corrections for the parameters are calculated from :

$$(p_1^j - p_o^j) = -(Q^{ij})^{-1}\left(q^i(\ldots p_o^j)\right). \qquad (7)$$

hence :

$$(p_1^j) = (p_o^j) - (Q_o^{ij})^{-1}\left(q^i(\ldots p_o^j)\right). \qquad (8)$$

This is computed iteratively, every time with the improved p^j :

$$(p_2^j) = (p_1^j) - (Q_1^{ij})^{-1}\left(q^i(\ldots p_o^j)\right) \qquad (8')$$

until all improvements are smaller than 1 mm.

The number of condition equations must be equal to the number of parameters that have to be solved. If an alignment consisting of three circle elements has to be computed, then one can choose, for example, the following conditions and parameters :

parameters p^j : l_1 k_3 w_3 k_4 w_4 k_5 w_5 l_7

conditions q^i : q^φ q^x q^y q^{D1} q^{D2} q^{D3} q^{D4} q^{D5}

Another possibility is to choose given values for the w-parameters :

parameters p^j : l_1 k_3 k_4 k_5 l_7

conditions q^i : q^x q^y q^φ q^{D2} q^{D4}

A different approach is found by the introduction of other types of conditions, for example : equality of two w parameters or of two k parameters ;

parameters p^j : l_1 k_3 w_3 k_4 w_4 k_5 w_5 l_7

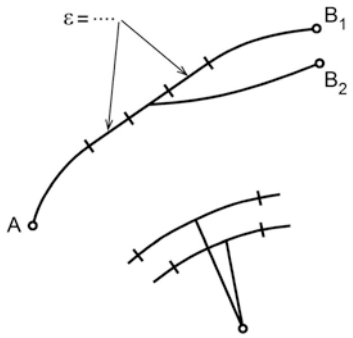
conditions q^i : q^φ q^x q^y q^{D1} q^{D2} q^{D3} q^{D4} q^8

with : $q^8 = w_3 - w_5 = 0$, or : $k_3 - k_5 = 0$

So much for the production of alignment parameters. This algorithm has been used since 1974, and during that time about 1000 alignments have been computed, for reconstructions and expansion of railways. Also other types of conditions are used, for example :

- concentricity of two circle elements,
- the sum total of the lengths of two straight lines,
- constraint zones,
- S-curves ($k_i = k_j$).

(see fig. 15).



Experience has learnt that the iterative process generally converges so quickly that the end result is reached with less than 5 iterations.

4. The stochasticity of alignment parameters.

We now consider (8) :

$$\left(p^j - p_o^j\right) = -(Q^{ij})^{-1}(q^i)$$

in which q^i, the misclosures of the condition equations, are functions of the approximate values p_o^j and of co-ordinates X_i, Y_i and bearings φ ; for example :

$$q^x = X_A - X_B + IX\left(\ldots p_o^{j(1)}, \varphi_{AA'}\right) + IX\left(\ldots p_o^{j(2)}, \varphi_{AA'} + I\varphi_1\right) + \ldots$$
$$\ldots + IX\left(\ldots p_o^{j(m)}, \varphi_{AA'} + I\varphi_1 + \ldots + I\varphi_{m-1}\right).$$

For reasons that will shortly be made clear in this article we are now going to use, instead of p^j, dimensionless parameters p^j, by dividing or multiplying parameters with length dimension with a distance of two co-ordinate points; choosing $l_{AA'}$ we get :

- length of straight line, l_i : $\rightarrow l_i / l_{AA'} = L_i$
- curvature of circle, k_i : $\rightarrow k_i / l_{AA'} = K_i$
- clothoide constant, α_i : $\rightarrow \alpha_i l_{AA'}^2 = A_i$

$\Big\}$ p^j

- the w-parameter is dimensionless already and remains unaltered.

The misclosures with length dimension, q^x, q^y, and q^D, subsequently pass into :

$$q^x = X_A - X_B + l_{AA'}\left[IX\left(\ldots p_o^{j(1)}, \varphi_{AA'}\right) + IX\left(\ldots p_o^{j(2)}, \varphi_{AA'}+I\varphi_1\right) + \ldots \right.$$
$$\left. \ldots + IX\left(\ldots p_o^{j(m)}, \varphi_{AA'}+I\varphi_1+\ldots+I\varphi_{m-1}\right)\right]$$

$$q^y = Y_A - Y_B + l_{AA'}\left[IY\left(\ldots p_o^{j(1)}, \varphi_{AA'}\right) + \ldots + IY\left(\ldots p_o^{j(m)}, \varphi_{AA'}+I\varphi_1+\ldots+I\varphi_{m-1}\right)\right]$$

$$q^D = \cos\varphi_D\left[X_A - X_D + l_{AA'} IX_{A\to P}\left(\ldots p_o^j, \varphi_{AA'}\right)\right] +$$
$$- \sin\varphi_D\left[Y_A - Y_D + l_{AA'} IY_{A\to P}\left(\ldots p_o^j, \varphi_{AA'}\right)\right]$$

(q^j remains unaltered)

Generally:

$$q^i = q^i\left(\ldots p_o^j, X_A, Y_A, X_{A'}, Y_{A'}, X_B, Y_B, X_{B'}, Y_{B'}, X_{D_1}, \ldots, Y_{D_n}\right) =$$
$$= q^i\left(\ldots p_o^j, \ldots Y^\alpha\right) \qquad (9)$$

The parameters p^j occurring in this formula are dimensionless quantities; the co-ordinates Y^α are to be considered as operationally defined, stochastic co-ordinates, i.e. S-co-ordinates in the terminology of BAARDA (1973); they are defined with respect to a S-base r,s (the non stochastic computation base used for the network adjustment).

We now differentiate (9), at first to the parameters :

$$(\Delta q^i) = \left(\frac{\partial q^i}{\partial p^j}\right)(\Delta p^j) =$$
$$= (R^{ij})\ (\Delta p^j) \qquad (10)$$

From this we can solve the dimensionless parameters, similar to (8) :

$$\left(p^j - p_o^j\right) = -(R^{ij})^{-1}(q^i) \qquad (11)$$

Further we differentiate (9) to the stochastic co-ordinates; underlining all stochastic variates :

$$\left(\underline{\Delta q^i}\right) = \left(\frac{\partial q^i}{\partial Y^\alpha}\right)\left(\underline{\Delta Y^\alpha}\right) =$$

$$\begin{pmatrix} \underline{\Delta q^\varphi} \\ \underline{\Delta q^X} \\ \underline{\Delta q^Y} \\ \underline{\Delta q^D} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & Y_{AB} & X_{AB}/l_{AA'} & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -X_{AB} & Y_{AB}/l_{AA'} & 0 & -1 & 0 & 0 & 0 \\ \cos\varphi_D & -\sin\varphi_D & L^o & L^{oo}/l_{AA'} & 0 & 0 & 0 & -\cos\varphi_D & \sin\varphi_D \end{pmatrix} \begin{pmatrix} \underline{\Delta X_A} \\ \underline{\Delta Y_A} \\ \underline{\Delta\varphi_{AA'}} \\ \underline{\Delta l_{AA'}} \\ \underline{\Delta X_B} \\ \underline{\Delta Y_B} \\ \underline{\Delta\varphi_{B'B}} \\ \underline{\Delta X_D} \\ \underline{\Delta Y_D} \end{pmatrix}$$

or:

$$\left(\underline{\Delta q^i}\right) = (S^{i\alpha})\left(\underline{\Delta Y^\alpha}\right) \qquad (12)$$

The stochasticity of the coordinates will be transmitted via the

misclosures q^i into the parameters p^j; using the difference equation of (11) :

$$(\triangle p^j) = -(R^{ij})^{-1}\,(\triangle q^i)$$

we obtain :

$$\boxed{\left(\underline{\triangle p}^j\right) = -(R^{ij})^{-1}\,(S^{i\alpha})\,\left(\underline{\triangle Y}^{\alpha}\right)} \qquad (13)$$

From this difference equation weight co-efficients for the p^j can be computed, together with those of the co-ordinates Y^{α} of constraint points and Y^r of reference points :

Weight co-efficients of :
- parameters p^j
- co-ordinates Y^{α}
- co-ordinates Y^r

$$\begin{pmatrix} g^{jj} & g^{\alpha j} & g^{rj} \\ g^{j\alpha} & g^{\alpha\alpha} & g^{r\alpha} \\ g^{jr} & g^{\alpha r} & g^{rr} \end{pmatrix} = \begin{pmatrix} -(R^{ij})^{-1}(S^{i\alpha}) & (0) \\ \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} & \end{pmatrix} (\overline{g}) \begin{pmatrix} -(R^{ij})^{-1}(S^{i\alpha}) & (0) \\ \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} & \end{pmatrix}^{*} \qquad (14)$$

5. The stochasticity of the alignment.

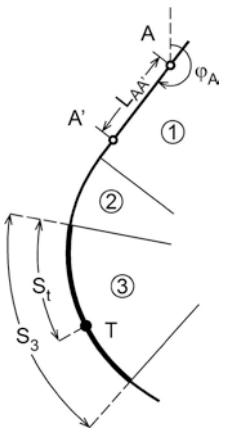
A point T on the alignment has the following co-ordinates : (suppose T is situated in the third element)

$$\begin{aligned} \underline{X}_T &= \underline{X}_A + \underline{l}_{AA'} \Big[IX\left(\ldots \underline{p}^{j(1)}, \underline{\varphi}_{AA'}\right) + IX\left(\ldots \underline{p}^{j(2)}, \varphi_{AA'} + \underline{I\varphi}_1\right) + \\ &\qquad + IX\left(f_T, \ldots \underline{p}^{j(3)}, \underline{\varphi}_{AA'} + \underline{I\varphi}_1 + \underline{I\varphi}_2\right) \Big] \\ \underline{Y}_T &= \underline{Y}_A + \underline{l}_{AA'} \Big[IY\left(\ldots \underline{p}^{j(1)}, \underline{\varphi}_{AA'}\right) + IY\left(\ldots \underline{p}^{j(2)}, \varphi_{AA'} + \underline{I\varphi}_1\right) + \\ &\qquad + IY\left(f_T, \ldots \underline{p}^{j(3)}, \underline{\varphi}_{AA'} + \underline{I\varphi}_1 + \underline{I\varphi}_2\right) \Big] \end{aligned} \qquad (15)$$

with :

$$\underline{I\varphi}_i = I\varphi\left(\ldots \underline{p}^j\right)$$

$$f_T = \frac{S_t}{S_3} \quad ; \quad 0 < f_T < 1$$

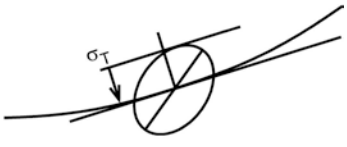


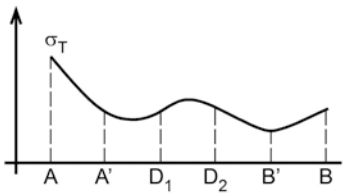
We now differentiate (15) :

$$\underline{\Delta X_T} = \underline{\Delta X_A} + (Y_T - Y_A)\,\underline{\Delta\varphi_{AA'}} + \frac{X_T - X_A}{l_{AA'}}\,\underline{\Delta l_{AA'}} + \sum_{j=1}^{n+3} \frac{\partial IX_{AT}}{\partial p^j}\,\underline{\Delta p^j}$$

$$\underline{\Delta Y_T} = \underline{\Delta Y_A} + (X_T - X_A)\,\underline{\Delta\varphi_{AA'}} + \frac{Y_T - Y_A}{l_{AA'}}\,\underline{\Delta l_{AA'}} + \sum_{j=1}^{n+3} \frac{\partial IY_{AT}}{\partial p^j}\,\underline{\Delta p^j} \qquad (16)$$

Using these difference equations and the matrix of weight co-efficients (14) we can calculate weight co-efficients for X_T and Y_T. These can be expressed for every point T in the form of a standard ellipse; a more meaningful expression is obtained, if the component perpendicular to the alignment, σ_T, is expressed in a graphics for the whole alignment. We call this the "orthogonal precision" of the alignment. It can be computed from the difference equation, derived from (16) :





$$\cos\varphi_T\,\underline{\Delta X_T} - \sin\varphi_T\,\underline{\Delta Y_T} = N_T\left(\underline{\Delta X_A}, \underline{\Delta Y_A}, \underline{\Delta\varphi_{AA'}}, \underline{\Delta l_{AA'}}, \ldots \underline{\Delta p^j}\right) \qquad (17)$$

It now becomes clear, that through the use of dimensionless parameters p^j the stochasticity of the alignment falls into two parts, whereby the first part (with the terms $\underline{\Delta X_A}$, $\underline{\Delta Y_A}$, $\underline{\Delta\varphi_{AA'}}$ and $\underline{\Delta l_{AA'}}$) is dependent on the deliberate choice of an S-base, but the second part (with the terms $\underline{\Delta p^j}$) is not.

We now wrote (16) in complex numbers, in the notation of BAARDA (1969).

$$\underline{\Delta z_T} = \underline{\Delta z_A} + \underline{\Delta z_{AT}}\,\underline{\Delta\Lambda_{AA}} + z_{AT}\left(\ldots \underline{\Delta p^j}\right) \qquad (18)$$

$$\longrightarrow \sum \frac{\partial IY}{\partial p^j}\,\underline{\Delta p^j} + i\sum \frac{\partial IX}{\partial p^j}\,\underline{\Delta p^j}$$

with:

$$\underline{\Delta Z} = \underline{\Delta Y} + i\,\underline{\Delta X}$$

$$\underline{\Delta\Lambda} = \underline{\Delta \ln l} + i\,\underline{\Delta\varphi}$$

$$\underline{\Lambda_{AA'}} = \underline{\ln z_{AA'}} \rightarrow \underline{\Delta\Lambda_{AA'}} = \underline{\Delta \ln z_{AA'}} = \frac{\underline{\Delta z_{AA'}}}{z_{AA'}}$$

$$\frac{z_{AT}}{z_{AA}} = e^{\Pi A'AT}$$

Therefore, suppose the network is adjusted in the S-system (a) :

$$\underline{\Delta z_T^{(a)}} = \underline{\Delta z_A^{(a)}} + e^{\Pi A'AT}\,\underline{\Delta z_{AA'}^{(a)}} + z_{AT}\left(\ldots \underline{\Delta p^j}\right)$$

We now apply an S-transformation from the (a) to the (r, s) system :

$$\underline{\Delta z}_T^{(rs)} = \underline{\Delta z}_A^{(rs)} + e^{\Pi_{A'AT}}\, \underline{\Delta z}_{AA'}^{(rs)} + z_{AT}\left(\ldots \underline{\Delta p}^j\right) \qquad (19)$$

these are non variant !

In view of (2.8') from BAARDA (1973) :

$$\underline{\Delta z}_A^{(rs)} = \underline{\Delta z}_A^{(a)} - \frac{z_{sA}}{z_{sr}}\,\underline{\Delta z}_r - \frac{z_{rA}}{z_{rs}}\,\underline{\Delta z}_s$$

$$\underline{\Delta z}_{AA'}^{(rs)} = \underline{\Delta z}_{AA'}^{(a)} - \frac{z_{AA'}}{z_{sr}}\,\underline{\Delta z}_r - \frac{z_{AA'}}{z_{rs}}\,\underline{\Delta z}_s$$

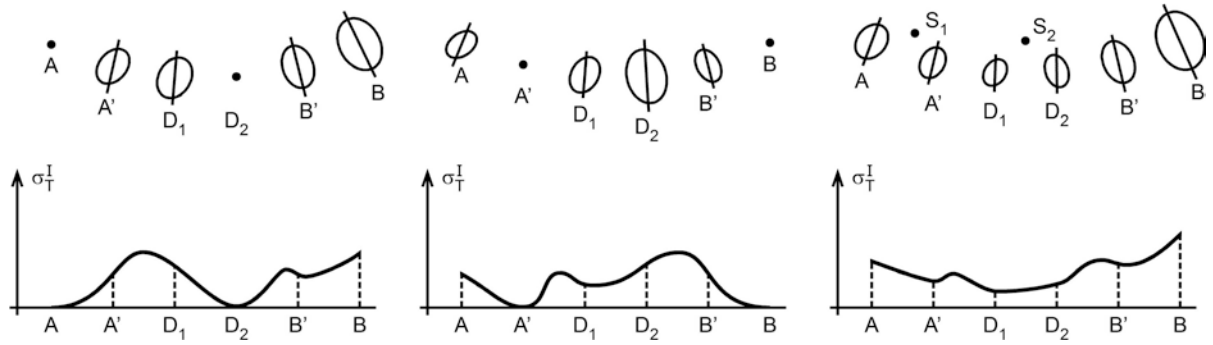
Substitution of these in (19) gives :

$$\underline{\Delta z}_T^{(rs)} = \underline{\Delta z}_T^{(a)} - \frac{z_{sA} + e^{\Pi_{AAT}}\, z_{AA}}{z_{sr}}\,\underline{\Delta z}_r - \frac{z_{ra} + e^{\Pi_{AAT}}\, z_{AA}}{z_{rs}}\,\underline{\Delta z}_s =$$

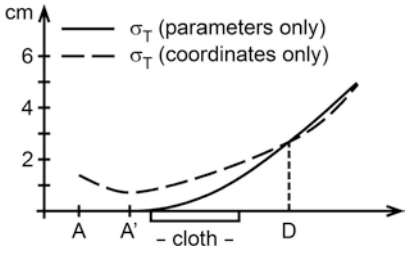
$$= \underline{\Delta z}_T^{(a)} - \frac{z_{sT}}{z_{sr}}\,\underline{\Delta z}_r - \frac{z_{rT}}{z_{rs}}\,\underline{\Delta z}_s \qquad (20)$$

This means, that the stochasticity of a point on an alignment reacts in exactly the same way on a S-transformation as that of a network point P_m, that lies at the same place as T, and is determined directly from the network adjustment.

Figure 18 gives an example of the graphics of σ_T, computed from (16, 17, 18), in three different S-sytems:



The $\underline{\Delta p}$-terms of (16, 17, 18) are <u>dimensionless</u> parameters; their stochasticity is <u>not dependent</u> on the S-system, but certainly on the lay-out of the alignment (distances of constraint-points, the radii, the lengths of elements etc.). Figure 19 gives some examples of graphics of σ_T, if calculated only from the terms $\underline{\Delta p}^j$:



From (16) it appears, that when T lies in A :

$$\underline{\Delta X}_{T=A} = \underline{\Delta X}_{A} \qquad \text{(all other co-efficients = 0)}$$
$$\underline{\Delta Y}_{T=A} = \underline{\Delta Y}_{A} \qquad \text{(all other co-efficients = 0)} \tag{21}$$

The standard ellipse of T is then equal to that of A. This also applies if T lies in B ;
substitute (13) in (16) :

$$\underline{\Delta X}_{T=B} = \underline{\Delta X}_{A} + (Y_B - Y_A)\,\underline{\Delta\varphi}_{AA'} + \frac{X_B - X_A}{l_{AA'}}\,\underline{\Delta l}_{AA'} - \left(\frac{\partial I X_{A\to T=B}}{\partial p^j}\right)(R^{ij})\,(S^{i\alpha})\,(\underline{\Delta Y}^{\alpha})$$
$$\underline{\Delta Y}_{T=B} = \underline{\Delta Y}_{A} + (X_B - X_A)\,\underline{\Delta\varphi}_{AA'} + \frac{Y_B - Y_A}{l_{AA'}}\,\underline{\Delta l}_{AA'} - \left(\frac{\partial I Y_{A\to T=B}}{\partial p^j}\right)(R^{ij})\,(S^{i\alpha})\,(\underline{\Delta Y}^{\alpha}) \tag{22}$$

In view of (10) the co-efficients :

$$\frac{\partial I X_{A\to T=B}}{\partial p^j} \qquad (j = 1 \ldots n+3)$$

are equal to the second row of (R^{ij}), i.e. the row of $\underline{\Delta q}^{X}$;
the same applies for :

$$\frac{\partial I Y_{A\to T=B}}{\partial p^j} \qquad (j = 1 \ldots n+3).$$

and the third row of (R^{ij}), the row of $\underline{\Delta q}^{Y}$;
Therefore :

$$\left(\frac{\partial I X_{A\to T=B}}{\partial p^j}\right)(R^{ij})^{-1} = (0\ 1\ 0\ 0\ \ldots\ldots\ 0)$$

$$\left(\frac{\partial I Y_{A\to T=B}}{\partial p^j}\right)(R^{ij})^{-1} = (0\ 0\ 1\ 0\ \ldots\ldots\ 0)$$

In view of (12) :

$$(S^{i\alpha})\,(\underline{\Delta Y}^{\alpha}) = \begin{pmatrix} \ldots\ldots\ldots\ldots\ldots\ldots \\ 1 \;\; 0 \;\; Y_{AB} \;\; X_{AB}/l_{AA'} \;\; -1 \;\; 0 \ldots \\ 0 \;\; 1 \;\; -X_{AB} \;\; Y_{AB}/l_{AA'} \;\; 0 \;\; -1 \ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{pmatrix} \begin{pmatrix} \underline{\Delta X}_{A} \\ \underline{\Delta Y}_{A} \\ \underline{\Delta\varphi}_{AA'} \\ \underline{\Delta l}_{AA'} \\ \underline{\Delta X}_{B} \\ \underline{\Delta Y}_{B} \end{pmatrix}$$

therefore, indeed :

$$\underline{\Delta X}_{T=B} = \underline{\Delta X}_{B}$$

$$\underline{\Delta Y}_{T=B} = \underline{\Delta Y}_{B}$$

If T lies in one of the constraint points, the co-efficients of the $\underline{\Delta p^j}$ in (17) will become equal to the row of (R^{ij}) belonging to the relevant constraint point, so :

$$\left(\cos\varphi_D \frac{\partial I X_{A\to T=D}}{\partial p^j} - \sin\varphi_D \frac{\partial I Y_{A\to T=D}}{\partial p^j}\right)(R^{ij})^{-1} = (0\ 0\ 0 \ldots 1\ 0 \ldots).$$

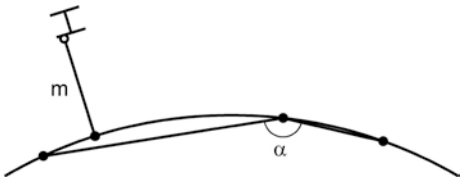
and therefore :

$\sigma_{T=D}$ = the component of D's standard ellipse perpendicular to the alignment.

Conclusion:
The orthogonal precision σ_T is, in all network points from which the alignment parameters are solved, equal to that of the network point (see fig. 18).

6. The application of the theory when planning the network for constraint and reference points.

In the last chapter, we analyzed the stochasticity (precision) of points on the alignment, and there was through the use of dimensionless parameters, demonstrated that the precision of an alignment points reacts in the same way to a S-transformation as a network point that lies at the same place. Hence the property that the precision (and also the reliability) of angles and short distances are non-variant in S-transformations, is also valid for angles between three alignment points, and for short distances between an alignment and a network point.
The criterion variates formulated in chapter 2 are composed of these quantities; therefore the criterion variates are non-variant too !



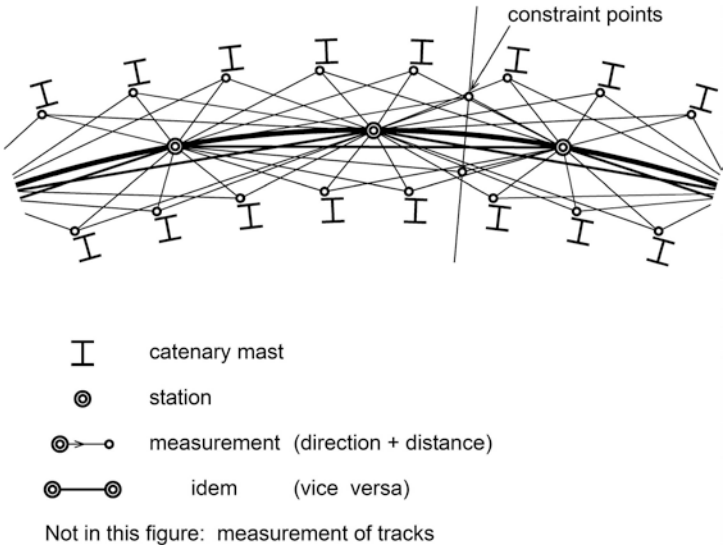
Now one can design the network for the measurement of constraint points and reference points as follows :

1. Choose a lay out for the network, and assume the precision for the instruments to be used.
2. Compute the matrix of weight co-efficients of the co-ordinates, $(\overline{g})$, in a S-system (computation base).
3. Compute standard deviations and boundary values of the criterion variates c_c, c_t, c_r, and c_m and affirm that these comply with the requirements.

The network, designed following this method is illustrated in figure 22.

- instrument locations 125 m from each other (= two catenary distances).
- from every instrument location one or two in front and behind are measured.
- reference points are measured in principle from three different locations.
- constraint points are always measured from three different locations.

The Netherlands Railways will measure approximately 500 of such networks for the DRIVER system, with an average length of $\frac{3}{4}$ km.



REFERENCES:

BAARDA (1967): Vereffeningstheorie (Adjustment theory) – stencilled edition, Delft.

BAARDA (1968): A testing Procedure for Use in Geodetic Networks - Netherlands Geodetic Commission, New Series, Vol. 2, No. 5, Delft.

BAARDA (1969): Puntsbepalingssystemen in het platte vlak (Polygon Theory in the Complex Plane) – stencilled edition, Delft.

BAARDA (1973): S-transformations and Criterion matrices – Netherlands Geodetic Commission, New Series, Vol. 5, No. 1, Delft.

RIJKSWATERSTAAT (1968): Het Snuffelprogramma -
Article in “Geodesia” 1971 No. 10.

DEUTSCHE BUNDESBAHN (1971): Dissertation Dr. W. Hupfeld – Technische Universität Carolo Wilhelmina, Braunschweig.

ESVELD, QUEE and SCHIPPERS (1982): Driver, a New Concept for Automatic Curve Correction – AICFF/UIC-Congres Malaga, Sept. 1982.

MODELLING ERRORS IN GEOMETRIC LEVELLING

Ole REMMER

Copenhagen, Denmark

1. Introduction

In [3] and [4] it is shown that the Kukkamäki corrections as applied to the levelling observations of the second Finnish levelling only are one third their correct size.

This insight is gained through a thorough analysis of the closing errors of this same levelling.

We shall now show that we may carry this analysis of closing errors further, finding one more until yet undiscovered systematic effect in the same closing errors.

We shall show that the character of this effect is the same as the error associated with the Kukkamäki correction namely a wrong factor to an otherwise correct correction.

The theme of this paper is the correction for land uplift or secular change. We shall show that this land uplift correction is too large by a factor of 1.6.

We shall furthermore explain how this erroneous assessment of the Finnish land uplift does not stem from incorrect geodetic data but from an erroneous analysis of these same data.

In addition we get the km-variance from the adjustment to fall to $0.19^{\left(10^{-3}\text{gpu}\right)^2}$/km which is exactly the same estimate as one gets from the forward-backward discrepancies; I believe this to be the first time such a phenomenon has been reported in geodetic literature.

2. The Evidence for Overcompensation

We use as starting point the closing errors completely corrected for refraction such as it has been described in [3] and [4].

We write these below in Table 1 column 3. In column 2 we write the land uplift corrections for the corresponding polygons taken from [2] Table VI. Column 1 is the circumference and column 4 will be explained later.

Circum-ference	Land uplift corrections	Closing Errors without refraction	Residual
F_i	z_i	y_i	$y_i - \gamma z_i$
239.02	0.51	7.96	7.76
428.87	- 1.22	- 1.43	- 0.95
369.79	2.36	4.65	3.71
290.49	- 0.16	2.87	2.93
413.90	0.31	-14.20	-14.32
413.61	- 5.95	- 3.28	- 0.92
451.17	-11.00	-17.07	-12.70
709.25	14.73	4.20	- 1.65
556.29	23.60	4.15	- 5.23
545.30	- 6.60	14.28	16.90
723.36	12.74	10.28	5.22
562.56	- 8.75	-17.86	-14.38
566.68	14.05	3.57	- 2.01
588.17	9.49	-16.15	-19.92
551.15	1.56	0.09	- 0.53
662.12	6.01	2.88	0.49
688.02	- 4.04	-11.10	- 9.49
324.29	2.25	15.96	15.07

Table 1

We now claim that in the mean the closing errors y_i above are nothing but a constant γ times the land uplift reductions above. If this is true we should be able to compute an estimate, r, of the correlation coefficient between the land uplift corrections z_i and the closing errors y_i above with is significantly different from zero.

We find, using the well known standard procedure for computing r:

$$r = \frac{\sum_{18}(z_i - \bar{z})(y_i - \bar{y})}{\sqrt{\sum_{18}(z_i - \bar{z})^2}\sqrt{\sum_{18}(y_i - \bar{y})^2}} = \frac{594.96}{\sqrt{1429.33}\sqrt{1916.50}} =$$

$$= \frac{594.96}{37.81 \cdot 43.78} = \frac{594.96}{1655,32} = 0.3594$$

We may test the null-hypothesis $\rho = 0$ with the aid of:

$$t_{16} = \frac{0.3594}{\sqrt{1-0.3594^2}}\sqrt{16} = \frac{0.3594 \cdot 4}{\sqrt{0.8708}} = \frac{1.438}{0.9332} = 1.54 > t_{0.92}$$

We thus see that we are reasonably sure that the z_i's and y_i's are in fact correlated.

Now of course if things were correct we should certainly not be able to see the land uplift corrections mirrored in the closing errors after they had bee officially removed.

The positive sign of the correlation coefficient tells us obviously that the land uplift has been overcompensated by the land uplift corrections cited in [2] Table VI and above in Table 1.

We shall want to remove this overcompensation by a small adjustment where our observation equations simply express the fact that the closing errors y_i with the refraction effect removed simply reflect the land uplift corrections z_i, i.e.

$$E\{y_i\} = \gamma z_i \qquad (1)$$

which of course by the Method of Least Squares leads to minimizing (cf. our λ-adjustment in [3] and [4]).

$$\boxed{\sum_{18}\frac{1}{F}(y_i - \gamma z_i)^2 = \min} \qquad (2)$$

resulting in the following normal equation for our parameter γ:

$$\hat{\gamma} = \frac{\sum^{18} \frac{y_i z_i}{F_i}}{\sum^{18} \frac{z_i^2}{F_i}} \qquad (3)$$

Let us find this estimate $\hat{\gamma}$ of γ; we get from Table 1:

$$\sum^{18} \frac{y_i z_i}{F_i} = 1.0814$$

$$\sum^{18} \frac{z_i^2}{F_i} = 2.7210 \qquad \text{i.e.}$$

$$\hat{\gamma} = 0.3974$$

For the variance factor σ^2, we find an estimate s^2:

$$s^2 = \frac{1}{17} \sum \frac{1}{F_i} (y_i - \gamma z_i)^2 = \frac{3.67168}{17} = 0.2160$$

For the estimate $s^2\{\hat{\gamma}\}$ of the variance $\sigma^2\{\hat{\gamma}\}$ we have:

$$s^2\{\hat{\gamma}\} = \frac{s^2}{\sum \frac{z_i^2}{F_i}} = \frac{0.2160}{2.7210} = 0.07938$$

This we may use to test the null-hypothesis:

$$H_0 : \gamma = 0$$

against the alternative hypothesis:

$$H_1 : \gamma > 0$$

This test is given by its region of rejection

$$t_{17} = \frac{\hat{\gamma}}{s\{\hat{\gamma}\}} = \frac{0.3944}{\sqrt{0.07938}} = \frac{0.3944}{0.2817} =$$

$$1.40 \sim t_{0.91}$$

which we again shall take as being a statement identical with saying that we are reasonably sure that γ is different from zero; i.e. we are reasonably sure that the land uplift corrections for Finland (from [2]) are overcompensated and that by a factor which is:

$$\frac{1}{1 - 0.3974} = 1.66$$

or to put in another way: On the land uplift corrections displayed in [2] Table VI or above in Table 1 we have to make a uniform reduction of 40%.

If we make a similar λ and γ-analysis on the closing errors of the old network ([2] Table VI bottom), we find $\hat{\lambda} = 2.8$ and $\hat{\gamma} = 0.7$.

We therefore conclude that λ and γ in the two levellings are identical.

We admit that we have made first a λ–adjustment and then a γ-adjustment on the Finnish material.

It should of course have been a unified adjustment where we determine λ and γ at the same time. This we are going to make now including the old network in the common adjustment.

Our new observation will then be

$$E\{y_i\} = -\lambda x_i + \gamma z_i \qquad (4)$$

where y_i is the closing error from [3] column 8, i.e. y_i is the closing error without refraction correction, x_i is the Kukkamäki refraction correction and z_i is the land uplift correction. Since the land uplift corrections in which we are interested is computed only on the basis of that part of the two levelling networks which is common we can only use the 13 first polygons of the second levelling together with all the 11 polygons from the first levelling i.e. we have 24 observation equations of type (4). The coefficients of these 24 observation equations are displayed below:

x_i	z_i	y_i
0.39	0.51	6.79
- 0.68	- 1.22	0.62
- 2.13	2.36	11.07
- 0.17	- 0.16	3.38
0.76	0.31	-16.49
8.34	- 5.95	-28.40
- 4.40	-11.00	- 3.82
- 0.92	14.73	6.97
- 3.72	23.60	15.35
4.35	- 6.60	1.18
- 7.42	12.74	32.63
2.48	- 8.75	-25.33
2.07	14.05	- 2.66
0.75	9.49	-18.41
5.61	1.56	-16.81
- 1.36	6.01	6.98
1.33	- 4.04	-15.11
- 2.31	2.25	22.92
0.51	- 8.45	-26.11
2.86	- 5.55	- 6.63
9.39	- 4.82	-29.05
0.16	- 1.76	- 6.09
- 2.94	-12.89	- 0.54
- 9.56	27.27	1.49
-2.57	-14.18	-54.42
-12.16	- 5.13	61.74
- 4.26	30.41	43.31
- 5.60	- 1.47	7.06
- 8.12	7.75	58.12

Table 2

Furthermore we remark that in our analysis until now (and also in the analysis in [3] and [4]) we have ignored the correlation between the different closing errors which is due to the simple fact that neighbouring polygons use the same measurements but with opposite signs. This incorrectness of method we are going to remedy by using the full non-diagonal weight-matrices.

First we write down, except for a constant, the full variance-covariance matrix for the 13 first polygons of the new network. It is:

P_{new}^{-1} =

239.0200												
-98.2900	428.8700											
-51.7400	-12.2100	369.0000										
0.0000	0.0000	-48.0300	290.4900									
0.0000	0.0000	0.0000	-11.5500	413.9000								
0.0000	-128.5500	0.0000	0.0000	0.0000	413.6100							
0.0000	0.0000	0.0000	0.0000	0.0000	-97.9600	451.1200						
0.0000	-73.3300	-120.9400	0.0000	0.0000	-40.2900	-114.5300	709.2500					
0.0000	0.0000	0.0000	-79.9200	-55.2800	0.0000	0.0000	-185.6100	556.2900				
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-123.6500	0.0000	0.0000	545.3000			
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-79.9900	0.0000	-115.3400	723.3600		
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-85.5200	0.0000	0.0000	-254.9200	562.5600	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-162.4400	0.0000	0.0000	-102.7500	566.6800

which gives the following weight matrix for the new network:

P_{new} =

0.00497												
0.00142	0.00315											
0.00093	0.00056	0.00323										
0.00022	0.00018	0.00066	0.00378									
0.00004	0.00005	0.00008	0.00021	0.00247								
0.00056	0.00117	0.00035	0.00014	0.00004	0.00307							
0.00029	0.00052	0.00036	0.00017	0.00006	0.00097	0.00305						
0.00049	0.00072	0.00087	0.00041	0.00014	0.00064	0.00086	0.00222					
0.00023	0.00032	0.00046	0.00078	0.00036	0.00028	0.00038	0.00095	0.00253				
0.00012	0.00021	0.00017	0.00008	0.00003	0.00036	0.00106	0.00040	0.00019	0.00229			
0.00014	0.00021	0.00024	0.00014	0.00005	0.00022	0.00042	0.00060	0.00034	0.00056	0.00220		
0.00018	0.00026	0.00032	0.00020	0.00008	0.00025	0.00040	0.00078	0.00054	0.00038	0.00121	0.00283	
0.00011	0.00016	0.00021	0.00029	0.00013	0.00015	0.00021	0.00047	0.00089	0.00015	0.00040	0.00083	0.0023

For the old network one finds correspondingly for the variance-covariance matrix:

$$P^1_{old} = 6.5x$$

229.3600										
-94.3000	425.1400									
0.0000	-123.3300	412.6600								
0.0000	0.0000	-34.4400	125.9600							
0.0000	0.0000	-94.9300	-40.3300	633.2800						
0.0000	0.0000	0.0000	0.0000	-70.7200	796.5800					
0.0000	0.0000	0.0000	0.0000	0.0000	-255.3800	808.1600				
0.0000	0.0000	0.0000	0.0000	0.0000	-45.1800	-65.5600	668.1400			
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-202.8800	641.0500		
-29.1500	-13.4500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	335.8300	
0.0000	-74.5100	-41.0500	0.0000	-139.8200	-182.9100	0.0000	0.0000	-183.7400	-119.0000	712.9400

which gives the following weight matrix for the old network:

$$P_{old} = \frac{1}{6.5}x$$

0.00499										
0.00133	0.00309									
0.00049	0.00111	0.00305								
0.00019	0.00042	0.00105	0.00848							
0.00018	0.00036	0.00068	0.00080	0.00191						
0.00012	0.00021	0.00022	0.00018	0.00036	0.00161					
0.00004	0.00007	0.00008	0.00006	0.00012	0.00053	0.00142				
0.00005	0.00008	0.00008	0.00005	0.00010	0.00024	0.00022	0.00172			
0.00012	0.00021	0.00019	0.00012	0.00022	0.00025	0.00013	0.00062	0.00195		
0.00062	0.00047	0.00029	0.00016	0.00026	0.00024	0.00008	0.00010	0.00026	0.00333	
0.00036	0.00064	0.00057	0.00036	0.00064	0.00062	0.00022	0.00027	0.00068	0.00079	0.00208

The factor 6.5 with which the weights in the old network are divided (or the variances multiplied) comes from the two possible separate adjustments of old and new network and their corresponding estimates of the variance factor.

The observation equation coefficients of Table 2 together with our two full weight matrices above give then the following normal equations:

$$\begin{Bmatrix} 0.54330 & 0.71457 \\ 0.71457 & 4.60552 \end{Bmatrix} \begin{Bmatrix} \hat{\lambda} \\ \hat{\gamma} \end{Bmatrix} = \begin{Bmatrix} 1.83166 \\ 3.71507 \end{Bmatrix} \qquad (5)$$

which gives

$$\boxed{\begin{aligned} \underline{\hat{\lambda}} &= 2.90277 = \underline{2.90} \\ \underline{\hat{\gamma}} &= 0.35628 = \underline{0.36} \end{aligned}}$$

For the variance factor σ^2 we find an estimate s^2 from the adjustment:

$$\underline{s^2} = \frac{V^*VP}{n-2} = \frac{4.2688}{22} = 0.19404 = \underline{0.19}$$

In order to compute the variances for $\hat{\lambda}$ and $\hat{\gamma}$ we have to compute $(N^{-1})_{11}$ and $(N^{-1})_{22}$. We find:

$$(N^{-1})_{11} = 2.31251$$

$$(N^{-1})_{22} = 0.27280$$

giving the following variances:

$$s^2\{\hat{\lambda}\} = 2.31251 \cdot 0.19404 = 0.44872$$

$$s^2\{\hat{\gamma}\} = 0.27280 \cdot 0.19404 = 0.05293$$

We may then pay our previous tests a short visit; they were tests of the hypothesis $\lambda = 1$ and $\gamma = 0$ respectively. We find:

$$t_{22} = \frac{2.90277 - 1}{\sqrt{0.44872}} = \frac{1.90277}{0.66987} = 2.84 > t_{0.995}$$

and

$$t_{22} = \frac{0.35628}{\sqrt{0.05293}} = \frac{0.35628}{0.23007} = 1.549 > t_{0.925}$$

We are now thus very sure that the Kukkamäki correction is under correcting the refraction and reasonably sure that the land uplift in Finland has been overcompensated by a factor of

$$\frac{1}{1-\gamma} = \frac{1}{1 - 0.36} = 1.56$$

For the refraction correction we knew already from [3] and [4] that the original Kukkamäki correction should be multiplied with a constant, which we have now determined to be

$$\lambda = 2.9$$

i.e. it is undercompensated by this same amount.

We end finally by remarking that if you determine the kilometer-variance σ^2 of the second Finnish levelling from the discrepancy between forward and backward levelling (the so-called local variance) you find the following estimate

$$s^2_{local} = 0.20\,mm^2/km = \underline{0.19^{(10^{-3}gpu)^2}/km}$$

while as we have seen we get from our adjustment above the so-called global variance estimate:

$$\underline{s^2_{global} = 0.19^{(10^{-3}gpu)^2}/km}$$

i.e. we have for the first time in geodetic literature a levelling adjustment where the adjustment variance (= the global variance) and the variance from the forward-backward discrepancies (= the local variance) is one and the same!

Furthermore this unique situation has been created by removing only two systematic errors: the refraction error and the land uplift error.

The equality of our two variance estimates, the local one and the global one, means that by removing the land uplift overcompensation and the Kukkamäki refraction correction under compensation we have taken out all systematic errors in the Finnish network or to put it another way:

If you make proper corrections for land uplift and refraction in the Finnish levelling of high precision, you have no other systematic errors left!

At the same time your global km-variance will be as low as

$$\sigma^2 = 0.19^{(10^{-3}\text{gpu})^2}/\text{km}$$
$$\sigma = 0.44^{10^{-3}\text{gpu}}/\sqrt{\text{km}}$$

or

which is far below any hitherto known variances, and equivalent to a true technological break-through!

In [4] has already been explained how the erroneous size of the refraction correction is born out of Kukkamäki's procedure.

We shall now in the next section accordingly concentrate on explaining in details how the erroneous overcompensating size of the land uplift correction is born.

To this purpose we shall use only already published Finnish material namely [1] and [2].

3. The Origin of the Overestimation in the Geodetic Data of the Land Uplift Values.

The explanation of the afore-mentioned error in the estimation of the land uplift rate in Finland may be very simply explained. The cause lies in erroneous treatment of correct data.

We shall prove this assertion by first describing the correct way in which to treat the levelling data from the two Finnish levellings; then we shall describe how the computations of the data were actually carried out according to [1]. Finally then we shall prove that this discrepancy between the computations in [1] and the correct computations completely explain the overcompensation of the land uplift in the Finnish computation procedure.

We shall start assuming that we have for all levelling sections 2 height-differences w_{1_i} and w_{2_i} from the first and second levelling respectively.

$$E\left\{\sum\left[w_{2_i}+\frac{w_{2_i}-w_{1_i}}{T_{2_i}-T_{1_i}}\left(T_{2_0}-T_{2_i}\right)\right]\right\}=0 \tag{6}$$

$$E\left\{\sum\left[w_{1_i}+\frac{w_{2_i}-w_{1_i}}{T_{2_i}-T_{1_i}}\left(T_{1_0}-T_{1_i}\right)\right]\right\}=0 \tag{7}$$

Here (6) belongs to the new and (7) to the old network, E means mathematical expectation (i.e. loosely said it means that if we used error-free observations in (6) and (7) then these would be exactly satisfied), T_{1_i} and T_{2_i} are the points of time of the actual observations w_{1_i} and w_{2_i} respec-

tively, while T_{1_0} and T_{2_0} are the common epochs of the first and second levelling respectively.

We underline that (6) and (7) are strictly true and completely hypothesis-free. They are also linear in w_{1_i} and w_{2_i} which we may underline by slightly rewriting them into.

$$E\left\{\sum\left[\left(1+\frac{T_{2_0}-T_{2_i}}{T_{2_i}-T_{1_i}}\right)w_{2_i}-\frac{T_{2_0}-T_{2_i}}{T_{2_i}-T_{1_i}}w_{1_i}\right]\right\}=0 \tag{8}$$

$$E\left\{\sum\left[\frac{T_{i_0}-T_{1_i}}{T_{2_i}-T_{1_i}}w_{2_i}+\left(1-\frac{T_{1_0}-T_{1_i}}{T_{2_i}-T_{1_i}}\right)w_{1_i}\right]\right\}=0 \tag{9}$$

For each polygon in a twice levelled network we shall thus get 2 condition equations (6) and (7) (or (8) and (9)). This will then lead to the well known adjustment by correlates which will then finally furnish us with correct adjusted values of $\bar{w}_{2_i}$ and $\bar{w}_{1_i}$ for all levelling sections (correct in the Least Squares sense).

In [1] however this corrects adjustment procedure has not been followed.

Instead an attempt has been made to break down the adjustment in phases, the first phase being an isolated adjustment of the land uplift values (see [1] p. 45-46).

Again we can with the aid of (6) and (7) write down the correct condition equations for this isolated adjustment. We find by subtracting (7) from (6)

$$E\left\{\sum\left[(w_{2_i}-w_{1_i})\left(1+\frac{T_{2_0}-T_{2_i}}{T_{2_i}-T_{1_i}}-\frac{T_{1_0}-T_{1_i}}{T_{2_i}-T_{1_i}}\right)\right]\right\}=0$$

or

$$E\left\{\sum\left[(w_{2_i}-w_{1_i})\frac{T_{2_i}-T_{1_i}+T_{2_0}-T_{2_i}-T_{1_0}-T_{1_i}}{T_{2_i}-T_{1_i}}\right]\right\}=0 \qquad \text{i.e.}$$

$$E\left\{\sum\left[\frac{w_{2_i} - w_{1_i}}{T_{2_i} - T_{1_i}}\left(T_{2_0} - T_{1_0}\right)\right]\right\} = 0$$

and dividing with $\left(T_{2_0} - T_{1_0}\right)$ we finally find as the condition equation for an isolated determination of the land uplift values:

$$E\left\{\sum\frac{w_{2_i} - w_{1_i}}{T_{2_i} - T_{1_i}}\right\} = 0 \qquad (10)$$

As can be seen from [1] page 46 it is exactly (10) which is being solved! In other words [1] starts with finding the correct land uplift values!

Unfortunately these correct values of land uplift are again destroyed in the next phase of the broken-down adjustment!

How and why does this happen? The answer is that by making the correct adjustment of land uplift values through (10) one thing is missing: the adjusted high values $\bar{w}_{2_i}$ and $\bar{w}_{1_i}$ of epochs T_{1_0} and T_{2_0}, we only have their differences $\bar{w}_{2_i} - \bar{w}_{1_i}$.

As explained above we can only get these correct adjusted values through the adjustment connected to (6) and (7).

However, probably in an effort to solve as few equations as possible at the same time, [1] tries to solve two separate adjustments instead of the correct total adjustment connected to (6) and (7).

These two separate adjustments are connected to the two following erroneous condition equations, which get by computing the land uplift change part in (6) and (7) through the use of the values of $\bar{w}_{2_i} - \bar{w}_{1_i}$ from the adjustment above connected to (10), i.e. to the following erroneous equations.

$$E\left\{\sum\left[w_{2_i} + \frac{\bar{w}_{2_i} - \bar{w}_{1_i}}{T_{2_i} - T_{1_i}}\left(T_{2_0} - T_{2_i}\right)\right]\right\} = 0^* \qquad (11)$$

$$E\left\{\sum\left[w_{1_i} + \frac{\bar{w}_{2_i} - \bar{w}_{1_i}}{T_{2_i} - T_{1_i}}\left(T_{1_0} - T_{1_i}\right)\right]\right\} = 0^* \qquad (12)$$

(The asterisks above indicate that the equations are wrong; it should be the unadjusted values w_{2_i} and w_{1_i} which should appear in (11) and (12) everywhere as it is in (6) and (7)).

It is, however, quite easy to see how the error is committed namely by writing

$$\frac{\bar{w}_{2_i} - \bar{w}_{1_i}}{T_{2_i} - T_{1_i}} = \text{corr}$$

i.e. by treating the last part of (11) and (12) as something not connected to w_{2_i} and w_{1_i}, i.e. by writing (11) and (12) as

$$E\left\{\sum \left[w_{2_i} + \text{corr}\left(T_{2_0} - T_{2_i}\right)\right]\right\} = 0 \qquad (11^*)$$

$$E\left\{\sum \left[w_{1_i} + \text{corr}\left(T_{1_0} - T_{1_i}\right)\right]\right\} = 0 \qquad (12^*)$$

hereby masking the fundamental fact that w_{1_i} appear also in (11*) and w_{2_i} appear also in (12*); one even succeeds in getting incorrect coefficients for w_{2_i} in (11*) and w_{1_i} in (12*).

The most serious error is of course that one completely ignores by using (11*) and (12*) separately the covariance which exists between (11*) and (12*).

This procedure of course gives wrong results which is amply illustrated by the figures in [1] itself.

When making the erroneous adjustment connected to (11*9 or (12*) new wrong values $\bar{w}^*_{2_i}$ and $\bar{w}^*_{1_i}$ are produced through this adjustment.

In [1] one might of course have been so lucky that the errors in $\bar{w}^*_{2_i}$ and $\bar{w}^*_{1_i}$ were negligible due to numerical reasons or due to the cancelling out of different kinds of errors. This is, however, not so because if one by accident (even through using wrong equations and wrong covariances) had stumbled close to the correct values $\bar{w}_{2_i}$ and $\bar{w}_{1_i}$ then of course:

$$w^*_{2_i} - w^*_{1_i} = \bar{w}_{2_i} - \bar{w}_{1_i} \qquad ^*$$

But as can be seen from [1] page 47 this is not the case. I.e. the estimates $\bar{w}^{*}_{2_i}$ and $\bar{w}^{*}_{1_i}$ are in fact wrong since their differences does not agree with the correct ones $\bar{w}_{2_i} - \bar{w}_{1_i}$ computed on the basis of (10).

The correct conclusion is then that $\bar{w}^{*}_{2_i}$ and $\bar{w}^{*}_{1_i}$ are wrong and therefore cannot be used for anything. However, in [1] the exact opposite conclusion is drawn! Namely that since $\bar{w}^{*}_{2_i} - \bar{w}^{*}_{1_i} \neq \bar{w}_{2_i} - \bar{w}_{1_i}$ the latter (correct value) has to be discarded!

The next phase is then to use these new erroneous values to compute the correction in (11*) and (12*) i.e.

$$\text{corr} = \frac{\bar{w}^{*}_{2_i} - \bar{w}^{*}_{1_i}}{T_{2_i} - T_{1_i}}$$

This quantity, corr, is then used in two new separate adjustments based on (11*) and (12*) which produces two new values $\bar{w}^{**}_{2_i}$ and $\bar{w}^{**}_{1_i}$ being even more in disagreement with $\bar{w}_{2_i} - \bar{w}_{1_i}$ than was $\bar{w}^{*}_{2_i}$ and $\bar{w}^{*}_{1_i}$.

This iteration procedure is then in principle continued until the last “adjusted” values do not differ from the last but one.

As we have pointed out above this procedure is erroneous. That this is so can be clearly seen even from the material published in [1]. According to [1] page 49 and page 51 this iterative erroneous adjustment produces a larger and larger variance factor (“mean square error of the unit weight”) for each iteration. This means of course that the observations satisfy less and less the condition equations for each new iteration, i.e. become more and more erroneous. This growth in variance factor goes in the new levelling from

$0.20\ \text{mm}^2/\text{km}$ to $0.244\ \text{mm}^2/\text{km}$
([1] page 51 and page 53)

and in the old levelling from

$1.74\ \text{mm}^2/\text{km}$ to $2.45\ \text{mm}^2/\text{km}$
([1] page 48 and page 50)

We have thus in the very material published in [1] a clear indication of growing errors when moving through the iterations described above.

This is then the whole explanation of the overcompensation of the land uplift values found in section 2.

References:

[1] Kääriäinen, Erkki: On the Recent Uplift of the Earth’s Crust in Finland. Veröff. des Finn. Geod. Inst., No. 42, Helsinki 1953.

[2] Kääriäinen, Erkki: The Second Levelling of Finland in 1935-1955. Veröff. des Finn. Geod. Inst., No. 61, Helsinki 1966.

[3] Remmer, Ole: Role and Evaluation of Refraction for the Processing of Levelling Networks. Proceedings NAD-Symposium, Ottawa 1980, pp. 623-646.

[4] Remmer, Ole: Refraction in Levelling. Geod. Inst. Medd. No. 54, Copenhagen 1980.

OPTIMIZATION OF CONTROL NETWORKS - STATE OF THE ART

Günter SCHMITT
Karlsruhe, FRG

ABSTRACT

A short review is presented about the current state of research in the optimization of geodetic networks. Special emphasis is given to the analytical methods in the second order design (weight problem) with criterion matrices as ideal variance-covariance matrices of the adjustment parameters.

1. Introduction

If in geodesy a new network is planned, the requirements of the users of the network must be considered in such a way, that the design of the net is able to fulfil these requirements. The adoption of the network design to the user's requirements is in general an optimization problem; we speak about the "optimization of geodetic networks" or the "optimal design of geodetic networks". A geodetic control network has three characteristics: the location of the points, the observations which are connecting them, and the datum to which the coordinates are referred. It is usual and widely accepted to classify the general optimization problem into different stages (see for example Grafarend (1974)):

The datum problem as the zero order design problem is concerned with the choice of an optimal reference system for the coordinates of the points and their variance-covariance matrix.

The configuration problem as the first order design problem, where the optimal positioning of the points is chosen under consideration of the possible measurements.

The weight problem as the second order design problem is dealing with the optimal accuracies of the observations in a fixed configuration.

The densification problem as the third order design problem, which has to give the answer to the question, how to improve an existing network by additional points and observations. Of course a lot of problems, which are coming up in geodetic applications, are including subproblems of different orders.

2. Objectives and general strategies

In earlier studies the quality of a network has been judged by its precision and its costs. The objectives of the design have been at first a maximal precision of the elements of the network (points and derived quantities) and secondly minimal costs for the marking of the points and for the performance and the measurements. Nowadays a third criterion for the quality of a network has turned out to be of high importance, that is its reliability. The reliability includes the following three topics: maximal power to detect outliers in the measurements, minimal influence of undetected errors on the results and high sensitivity to detect movements in deformation networks. It is well known that the three requirements - precision, reliability and economy - have a contrary influence on the optimal design of a network. Generally one can say that the realization of reliability is mainly the task for the first order design, whereas the second order design is particularly concerned with the precision. The minimization of the costs can be integrated in both first and second order design. In opposition to that, one can see the cost problem as a secondary transformation problem after a first and/or second order design.

The solution strategies of the optimization problems in the different orders of the design are dependent both on the mathematical form, to which the problem has been brought, and the shape of the objective function, which is representing the aim of the design. Purely analytical solution ways are mainly known in the second order design, where the standard problems can be expressed in terms of linear equations or linear inequalities – we come later back to this topic. Standard algorithms of linear of nonlinear programming are in use, such as the well known simplex method. In the last time, also more or less direct solutions have had success, which are using generalized inverses.

A method which is most popular and suited for first, second and third order design problems is the network simulation. An arbitrary number of design variants can be generated by variation of the free parameters of the problem, and compared to the objectives of the design. The variation of the parameters can be arranged by selection or by random processes. Most effective is the implementation of the simulation within an interactive computer system with graphic terminals, where the designs can be criticized and directly improved in a dialog mode. One great advantage of the simula-

tion method is, that arbitrary decision criteria can be used and compared together, in order to find an optimal design; there is no forcing need to bring these criteria into a strong mathematical form what, is indispensable, if one uses purely analytical solutions with discrete risk functions.

3. Scalar objective functions

The most optimization strategies require the formulation of a risk function. Generally, the objectives of maximal accuracy and minimal costs can be combined to a closed objective function, such as proposed by Wolf (1970):

$$Z = R_T \cdot T + R_S \cdot S.$$

T is standing for the precision requirements, S for the costs, R_T and R_S are standardisation factors, by which the influence of both parts can be weighted adequately. There are a lot of approaches to bring the costs into a mathematical form S. The most approaches try to split the costs for the measurements into constant terms for driving to the stations, setting up of the instruments and signalizing the points. The remaining free parameters are then the repetition numbers of the observations. Different instrumentations can be considered in this concept by the introduction of special efficiency numbers for these instruments or error formulas for a single measurement.

The precision term T of the above general risk function Z is based upon the variance-covariance matrix of the point coordinates or, in special geodetic problems, upon the variance-covariance matrix of derived quantities, which are to be understood as functions of the coordinates. T is expressed in terms of the eigenvalues λ_i of the respective v.c.-matrix. In the second case mainly the trace of the v.c.-matrix is used, that means the sum of the variances of the derived quantities. If the accuracy of the position of a point is considered, one has to look at the 2x2-v.c.-matrix Q_P of its coordinates. The point precision can be expressed by three measures. At first, the trace Q_P, which is to be understood as the mean variance (sum of the eigenvalues $\lambda_1 + \lambda_2$), leads to the mean point error defined by Helmert. Secondly, the determinant det Q_P as generalized variance (product of the eigenvalues $\lambda_1 \cdot \lambda_2$) gives the point error corresponding to Werkmeister's definition, which is to be interpreted as the area of the mean error ellipse. Thirdly, the quality of the point situation can be

expressed by the relation between the main axes of its error ellipse or by the quotient of the eigenvalues λ_1 / λ_2. In statistics, the minimal trace of Q_P is known as A-optimality, the minimal determinant of Q_P is called D-optimality, whereas $\lambda_1/\lambda_2 \rightarrow 1$ is characterizing E-optimality or isotropy, which means that the error ellipse is degenerating to a circle. In a network a global precision function can be formulated by combining the local accuracy measures of all 2x2 diagonal submatrices Q_{P_i} of the global v.c.-matrix Q_X of the complete network, with the particularity that

$$\sum_i \operatorname{tr} Q_{P_i} = \operatorname{tr} Q_X \ .$$

Grafarend and Harland (1973) show also how to combine different optimalities (A-, D- and E-optimality) to general precision functions. They are using them in the first order design of geodetic standard problems such as bisection and resection of one or more points.

Reliability aspects have found only hesitating acceptance until now in the objective functions of optimization processes, whereas a certain number of reliability measures are available. In the concept of "inner reliability" the boundary values $\nabla_0 l_i$ (Baarda's notation) are suited to indicate how well the observations can be checked. On the other hand, the "external reliability" of a network can be described by the following three parameters: the level of significance α_0 and the power β_0 of the applied statistical test and $\overline{\delta}_0$ as the maximal weighted norm of the values $(\nabla_0 x)_i$, which are to be understood as the effects of the boundary values $\nabla_0 l_i$ of the observations on the coordinates (van Mierlo, 1981).

4. Criterion matrices

Together with the great progress which has been made in the research about second order design solutions in the last decade, criterion matrices have had great success as alternatives to scalar risk functions. Criterion matrices have been introduced in geodesy by Baarda (1971) as artificial v.c.-matrices, which can be used for example as quality measures for networks and in the densification of networks. In the second order design problem they are to be understood as ideal v.c.-matrices, ideal means that they represent the optimal accuracy situation in the network to be designed.

The basic equation of the second order design with criterion matrices is

the well known relation

$$\left(A^T P A\right)^{-1} = Q_X \quad ,$$

in which A is the configuration or design matrix of the network, and Q_X is the given criterion matrix. With the restriction to uncorrelated observations (diagonal design), the unknown weights of the observations in P have to be estimated in such a way, that the criterion matrix is approximated as good as possible by the optimal design. In a second step, the inversion of P leads to the optimal variances and therewith to the necessary accuracy, with which the observations have to be carried out. The approximation of Q_X or, referring to the inverse equation

$$A^T P A = P_X \coloneqq Q_X^{-1} \quad ,$$

the approximation of P_X, can be realized by several mathematical strategies. Because the consideration of these different strategies is not the crucial point of this review, we only refer to the following publications: Grafarend (1975), Schaffrin et al. (1977), Schaffrin et al. (1981), Schmitt (1979) and Wimmer (1981).

In general, a geodetic network is said to be optimal if it has a homogeneous accuracy in all its parts. In two and three dimensions a second ideal property is added: isotropy. If both properties are happening in a two-dimensional network, all point error ellipses are degenerating into circles with equal radius. Moreover, we call the error situation completely isotropic if in addition to that all relative error ellipses are circles, too. Such an ideal error situation is coming up for example if the v.c.-matrix of a network degenerates to the identity matrix. But the identity matrix is unsuited to be introduced as criterion matrix in the second order design, because the postulation of zero-covariances between all coordinates is unrealizable. Therefore it is necessary to look for a general stochastic concept, in which an overall ideal covariance situation can be formulated. This concept is found in the interpretation of a geodetic point field as a stochastic process with the expectation values equal to zero and the covariance function with Taylor-Karman structure of the absolute coordinates, which is a homogeneous and isotropic structure. Covariance submatrices between two points can be computed from longitudinal and transversal correlation functions with the argument r, which is the planar distance between the points. The advantage of such criterion matrices is

that the observation plan of the network does not enter. The drawback is that in the underlying first order Markow model the characteristic distances of the correlation functions are free. By the way, a completely isotropic situation can be generated with identical longitudinal and transversal correlation functions. The complete theory is described in Grafarend and Schaffrin (1979), including the computation of derived v.c.-matrices of coordinate differences, azimuths, angles and distances.

The extensive discussions about the property of estimability in geodetic adjustments have caused a trend away from the point field concept to the construction of criterion matrices for estimable quantities. These criterion matrices have the advantage, that they can be computed without dealing with the problem of fixing the datum of a network. Schaffrin and Grafarend (1982) show how to formulate ideal v.c.-matrices for azimuths, angles and distances under the postulated of homogeneity and isotropy. This means for instance in distance networks, that the variances of distances and the covariances between distances of equal length are equal in all parts of the network, independent from their directions. Nevertheless, all the second order design procedures which are based upon the adjustment approach by variation of coordinates use as input a criterion matrix of absolute coordinates. Therefore, it must be done the step from the criterion matrix of the estimable quantities to the allocated criterion matrix of the coordinates (Schaffrin and Grafarend, 1982).

A second crucial point in the use of a criterion matrix Q_X is, that, if severe model errors shall be avoided, it must have the same rank as the configuration matrix A. Correlation functions of Taylor-Karman type are positive definite and lead to regular criterion matrices which are unsuited for free networks. The step from a regular criterion matrix to such a one with a defect corresponding to the adjustment model can be done by an arbitrary S-transformation. Schaffrin et al. (1981) propose to use as criterion matrix the v.c.-matrix of the projection $A^{-}Ax$ which is estimable in contrary to x itself.

The derived matrix is then

$$Q_{A^{-}Ax} = A^{-}A\,Q_X\,(A^{-}A)^{T} \quad ,$$

which has to be inverted to

$$P_x = (Q_{A^-AX})_r^-$$

$-( )_r^-$ is standing for a symmetric reflexive g-inverse (for example the pseudo-inverse $( )^+$) – in order to preserve the rank and the inverse eigenvalue properties for these optimization procedures, which are approximizing P_x. In both ways, the correctness of the model is reached by loosing the properties of homogeneity and isotropy of the point errors.

5. Conclusions

In order to save room for the text, the references are kept to an absolute minimum, which should give the reader an idea where to look for more information. A detailed list of references may be found in the first textbook about network optimization, written by Grafarend et al. (1979), further on within a review about that topic by Schmitt (1982).

REFERENCES

BAARDA, W. (1971): S-transformations and criterion matrices in the adjustment of geodetic networks. Computing Centre of the Delft Geodetic Institute, Delft, Netherlands.

GRAFAREND, E. (1974): Optimization of Geodetic Networks. Boll. Geodesia Sci. Aff. 33, 351-406.

GRAFAREND, E. (1975): Second order design of geodetic nets. ZfV 100, 158-168.

GRAFAREND, E. and HARLAND, P. (1973): Optimales Design geodätischer Netze I. DGK, Report A 74, München, FRG.

GRAFAREND, E., HEISTER, H., KELM, R., KROPFF, H. and SCHAFFRIN, B. (1979): Optimierung geodätischer Meßoperationen. Karlsruhe, FRG.

GRAFAREND, E. and SCHAFFRIN, B. (1979): Kriterion Matrizen I – Zweidimensionale homogene und isotrope geodätische Netze. ZfV 104, 133-149.

van MIERLO, J. (1981): Second order design: precision and reliability aspects. AVN 88, 95-101.

SCHAFFRIN, B., GRAFAREND, E. and SCHMITT, G. (1977): Kanonisches Design geodätischer Netze I. MG 2, 263-306.

SCHAFFRIN, B., KRUMM, F. and FRITSCH, D. (1981): Positiv-diagonale Genauigkeitsoptimierung von Realnetzen über den Komplementaritäts-Algorithmus. In: Conzett, Matthias and Schmidt: Ingenieurvermessung 80, Beiträge zum VIII. Internationalen Kurs für Ingenieurvermessung, Zürich, 1980, Bonn, FRG.

SCHAFFRIN, B. and GRAFAREND, E.W. (1982): Kriterion-Matrizen II - Zweidimensionale homogene und isotrope geodätische Netze. ZfV 107, 183-193.

SCHMITT, G. (1979): Zur Numerik des Designs zweiter Ordnung. DGK, Report C 256, München, FRG.

SCHMITT, G. (1982): Optimization of Geodetic Networks. Submitted to Reviews of Geophysics and Space Physics.

WIMMER, H. (1981): Ein Beitrag zur Gewichtsoptimierung geodätischer Netze. DGK, Report C 269, München, FRG.

WOLF, H. (1970): Polarität und Optimierung bei freien und eingeschalteten geodätischen Netzen. AVN 77, 291-298.

Abbreviations

AVN - Allgemeine Vermessungs-Nachrichten
DGK - Deutsche Geodätische Kommission
MG - Manuscripta Geodaetica
ZfV - Zeitschrift für Vermessungswesen

STATION ADJUSTMENT OF DIRECTIONS USING GENERALIZED INVERSES

Lars SJÖBERG
Uppsala and Gävle, Sweden

ABSTRACT

In a strict adjustment of directions at a triangle station no direction is arbitrarily chosen as "zero direction". Due to the unknown orientation of the set of directions the system of normal equations becomes singular, which implies that there is no unique least squares solution.

Furthermore, the directions are non-estimable in the sense of Rao. However, it is shown that all least squares estimates of directions provide identical unbiased angle estimates. The estimates and their covariance matrices are derived in the cases of directions observed in full rounds and incomplete rounds.

1. INTRODUCTION

The adjustment of a triangulation network is usually prepared by a least squares adjustment of observed directions at each station. As the orientation quantities of the directions are unknown, there is no unique least squares solution to the problem. Thus Bjerhammar (1973) and Ethrog (1981) suggested a generalized inverse approach. Bjerhammar (ibid.) and Holsen (1981) paid special attention to the covariance matrix of the adjusted directions and emphasized that for certain least squares estimates the covariance matrix is diagonal.

It is reasonable to restrict the acceptable solution space of directions to those providing unbiased angle estimates. Thus we start with deriving the set of least squares solutions and then we study the condition on unbiased angle estimation.

2. A CONDITION FOR UNBIASED ANGLE ESTIMATION

The general equation in the observation of a direction (X_i) is of the form

$$X_i - Z = \ell_i - \varepsilon_i \tag{2.1}$$

where Z is an unknown orientation quantity, ℓ_i is the observed direction relative to Z and ε_i is the random observation error. Each round (set) of observations adds one unknown orientation quantity to the system of equations. The system of observation equations may be written with obvious matrix notations:

$$\underset{(n,m)}{A}\ \underset{(m,1)}{X} + \underset{(n,k)}{B}\ \underset{(k,1)}{Z} = \underset{(n,1)}{L} - \underset{(n,1)}{\varepsilon} \tag{2.2}$$

or

$$\underset{(n,\ell)}{C}\ \boldsymbol{\gamma} = L - \varepsilon \tag{2.3}$$

where

$$C = [A, B]$$
$$\gamma^T = [X^T, Z^T]$$

and $\ell = m + k$

Here n is the total number of observations, m is the number of directions and k is the number of rounds. The normal equations corresponding to formula (2.3) are

$$C^T C \gamma = C^T L \tag{2.4}$$

with the general least squares solution

$$\tilde{\gamma} = \left(\underset{\sim}{C^T C}\right)^{-1} C^T L \tag{2.5}$$

minimizing $\varepsilon^T \varepsilon$. The explicit expression of the generalized inverse in (2.5) is (Bjerhammar 1973, p. 113, and the Appendix)

$$\left(\underset{\sim}{C^T C}\right)^{-1} = \left(C^T C\right)_0^{-1} + \left[I - \left(C^T C\right)_0^{-1} C^T C\right] M \tag{2.6}$$

where

$\left(C^T C\right)_0^{-1}$ = particular inverse

M = arbitrary inverse of dimension (ℓ, ℓ).

It follows from (2.3) that

$$E\{L\} = C \gamma \tag{2.7}$$

However, as can be seen from formula (2.5) the estimation of Y is neither unique nor estimable, i.e.

$$E\{\tilde{Y}\} \neq Y,$$

because we have no information on the absolute orientation of Y, but only on the interorientation among the directions and orientation quantities. Thus if one Z or X could be fixed the rank defect would vanish. However, a sufficient condition for practical purposes is that the least squares estimate of Y yields unbiased angle estimates. The vector of angle estimates ($\tilde{\mu}$) are related to the directions $\tilde{Y}$ by

$$\tilde{\mu} = G\,\tilde{Y} \tag{2.8a}$$

where

$$\underset{(m^*,\ell)}{G} = \underset{(m^*,m)}{[F} \quad \underset{(m^*,k)}{0]} \tag{2.8b}$$

$$F = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & & \vdots \\ 0 & 0 & -1 & 1 & 0 & & \\ \vdots & & & & & & \\ 0 & & \cdots & & 0 & -1 & 1 \end{bmatrix} \tag{2.8c}$$

$$m^* = m-1$$

In the case of unbiased estimation of μ we thus obtain

$$E\{\tilde{\mu}\} = \mu = G\gamma \tag{2.9}$$

Furthermore, from (2.8a), (2.5) and (2.7) we get

$$E\{\tilde{\mu}\} = G\left(\underset{\sim}{C}^T\underset{\sim}{C}\right)^{-1}\underset{\sim}{C}^T E\{L\} = G\left(\underset{\sim}{C}^T\underset{\sim}{C}\right)^{-1}\underset{\sim}{C}^T C\,Y \tag{2.10}$$

Finally, formulas (2.9) and (2.10) yield the following condition (c.f. Rao and Mitra 1971, p. 139):

$$G\left(\underset{\sim}{C}^T\underset{\sim}{C}\right)^{-1}\underset{\sim}{C}^T C = G \tag{2.11a}$$

or

$$G\,\underset{\sim}{C}^{-1} C = G \tag{2.11b}$$

Any inverse of $C^T C$ in (2.5) being consistent with (2.9) has to satisfy formula (2.11). This will be a fundamental condition for the estimation of directions in this paper.

3. APPLICATION OF THE CONDITION

3.1 The pseudo inverse solution

We now recall formulas (2.1) and (2.3). From these equations it follows that

$$C D^T = 0 \tag{3.1a}$$

if

$$D_{(1,\ell)} = [1, 1, \ldots, 1] \tag{3.1b}$$

As the rank defect of the general system (2.3) equals the rank of D (=1) one solution is given by the method of orthogonal bordering with D (Bjerhammar 1973, p. 103)

$$\hat{\gamma} = C_{II}^{-1} L \tag{3.2a}$$

where

$$C_{II}^{-1} = \left(C^T C + D^T D\right)^{-1} C^T \tag{3.2b}$$

This is the so-called pseudo inverse of C minimizing $\varepsilon^T \varepsilon$ and $\gamma^T \gamma$.

Let us now check whether C_{II}^{-1} yields unbiased angle estimates. From (3.2b) and the left member of (2.11b) we get

$$\begin{aligned} G C_{II}^{-1} C &= G\left(C^T C + D^T D\right)^{-1} C^T C = \\ &= G\left\{I - D^T D\left(C^T C + D^T D\right)^{-1}\right\} = G \end{aligned} \tag{3.3}$$

because, in view of (2.8b-c) and (3.1b)

$$G D^T = 0 \tag{3.4}$$

Hence C_{II}^{-1} satisfies (2.11b), which implies that the pseudo inverse solution of directions yields unbiased angle estimates.

3.2 The general least squares solution

The general least squares solution for Y was given in formulas (2.5) and (2.6). Selecting the pseudo inverse of C^TC as the particular inverse we get the general least squares inverse

$$\underset{\sim}{C}^{-1} = C_{II}^{-1} + \left[I - C_{II}^{-1} C\right] M C^T \tag{3.5}$$

Now we ask whether $\hat{\gamma}$ yields unbiased angle estimates. If the answer is affirmative the inverse (3.5) must satisfy the condition (2.11b) for any choice of M. In view of (3.3) we obtain from (3.5)

$$G \underset{\sim}{C}^{-1} C = G C_{II}^{-1} C + G\left(I - C_{II}^{-1} C\right) M C^T C = G \tag{3.6}$$

and we have proved that (2.11b) is satisfied, i.e. the general least squares solution of a station adjustment provides unbiased angle estimates. As a matter of fact we can even prove the following proposition.

Proposition: Each least squares estimate of direction yields identical angle estimates.

Proof: $\tilde{\mu} = G\tilde{X} = G\left[C_{II}^{-1} + \left(I - C_{II}^{-1} C\right) M C^T\right] L = G C_{II}^{-1} L$

The proof follows from (3.3).

Finally, if the observations are related to a weight matrix $P \neq I$ the pseudo inverse should be replaced by

$$C_{IP}^{-1} = \left(C^T P C + D^T D\right)^{-1} C^T P \qquad (3.7)$$

minimizing $\varepsilon^T P \varepsilon$.

4. ALTERNATIVE EXPRESSIONS FOR THE LEAST SQUARES SOLUTION

We now tackle the singular system of equations (2.2) in two different ways to obtain more explicit solutions for X and Z. The system of normal equations (2.4) can be written

$$A^T A X + A^T B Z = A^T L \qquad (4.1a)$$

$$B^T A X + B^T B Z = B^T L \qquad (4.1b)$$

As $A^T A$ has full rank (4.1a) yields

$$X = A_0^{-1} L - A_0^{-1} B Z \qquad (4.2)$$

where

$$A_0^{-1} = \left(A^T A\right)^{-1} A^T$$

Inserting this formula into (4.1b) we get

$$K^T K Z = K^T L \qquad (4.3)$$

where

$$K = \left(I - A^0\right) B \quad ; \quad A^0 = A A_0^{-1}$$

Formula (4.3) is a singular normal equation with the general solution

$$\tilde{Z} = K_{II}^{-1} L + \left(I - K_{II}^{-1} K\right) M K^T L \qquad (4.4)$$

where

$$K_{II}^{-1} = \left(B^T \left(I - A^0\right) B + D_1^T D_1\right)^{-1} B^T \left(I - A^0\right) \qquad (4.5)$$

$$K D_1^T = 0 \qquad (4.6)$$

In view of (3.1) it can easily be shown that

$$D_1 = [1, 1, \ldots, 1]$$

satisfies (4.6). Inserting (4.4) into (4.2) we finally arrive at

$$\tilde{X} = A_0^{-1} L - A_0^{-1} B K_{II}^{-1} L - A_0^{-1} B \left(I - K^0\right) M K^T L \qquad (4.7)$$

where

$$K^0 = K_{II}^{-1} K$$

In a completely analogous manner we could also start from (4.1b) to solve for Z:

$$Z = B_0^{-1} L - B_0^{-1} A X \qquad ; \qquad B_0^{-1} = \left(B^T B\right)^{-1} B^T \qquad (4.8)$$

Inserting this formula into (4.1a) we get the general solutions

$$\tilde{X} = \left(A^T \left(I - B^0\right) A\right)^{-1} A^T \left(I - B^0\right) L \qquad (4.9a)$$

$$\tilde{Z} = B_0^{-1} L - B_0^{-1} A \left\{A^T \left(I - \underset{\sim}{B^0}\right) A\right\}^{-1} A^T \left(I - B^0\right) L \qquad (4.9b)$$

Finally we note that in the case of a weight matrix P among the observations we should replace the previous definitions of A_0^{-1} and B_0^{-1} by

$$A_0^{-1} = \left(A^T P A\right)^{-1} A^T P$$

$$B_0^{-1} = \left(B^T P A\right)^{-1} B^T P$$

5. VARIANCE-COVARIANCE ESTIMATION

The random errors of the least squares estimates $\tilde{\gamma}$ are given by formula (2.5):

$$\tilde{\varepsilon}_\gamma = \left(C^T \underset{\sim}{C}\right)^{-1} C^T \varepsilon \tag{5.1}$$

where we assume that ε obeys the relation for random uncorrelated errors:

$$E\{\varepsilon\, \varepsilon^T\} = \sigma^2 I \ ,$$

σ^2 being the variance of unit weight. Formula (5.1) yields the covariance matrix for $\tilde{\gamma}$:

$$Q_{\gamma\gamma} = E\left\{\tilde{\varepsilon}_\gamma\, \tilde{\varepsilon}_\gamma^T\right\} = \sigma^2 \left(C^T \underset{\sim}{C}\right)^{-1} C^T C \left(C^T \underset{\sim}{C}\right)^{-1} \tag{5.2}$$

where σ^2 may be unbiasedly estimated by (Bjerhammar 1973, p. 128):

$$s^2 = \frac{L^T (L - C\tilde{\gamma})}{n - \text{rank}(C)} = \frac{L^T \left(L - A\tilde{X} - B\tilde{Z}\right)}{n - m - k + 1} \tag{5.3}$$

In particular, for the pseudo inverse solution we get

$$Q_{\gamma\gamma} = \sigma^2 \left(C^T C\right)^{-1}_{II} \tag{5.4}$$

In the same way we obtain the following covariance matrix for $\tilde{X}$ of formula (4.9a)

$$Q_{XX} = \sigma^2 \underset{\sim}{R}^{-1} R \underset{\sim}{R}^{-1} \tag{5.5}$$

where

$$R = A^T \left(I - B^0\right) A$$

6. PRACTICAL CHOICE OF SOLUTION

As we have shown all least squares solutions yield unbiased angle estimates. Thus the choice of solution becomes a question of convenience. The pseudo inverse solution (3.2) is an all-round solution both for direction and orientation quantity estimation. We distinguish between two cases of observation designs: full and incomplete rounds of observations.

6.1 Full rounds

An example of complete rounds of observations is given in section 9. In this case it can easily be shown that

$$B^T\left(I-A^0\right)B+D_1^T D_1 = B^T B = m\,I$$

where m is the number of directions. Hence,

$$K_{II}^{-1} = \frac{1}{m}\,B^T\left(I-A^0\right) \qquad (6.1)$$

Moreover,

$$A_0^{-1}\,B\,K_{II}^{-1} = A_0^{-1}\,B^0\left(I-A^0\right) = 0 \qquad (6.2)$$

Inserting (6.1) and (6.2) into (4.4) and (4.7) with M = 0 (pseudo inverse solution) we arrive at the following solutions (see also the Appendix):

$$\hat{Z} = B_0^{-1}\left(I-A^0\right)L = \frac{1}{m}\,B^T\left(I-A^0\right)L \qquad (6.3a)$$

and

$$\hat{X} = A_0^{-1}L = \frac{1}{k}\,A^T L \qquad (6.3b)$$

The covariance matrix $\hat{X}$ becomes

$$Q_{XX} = \sigma^2\left(A^T A\right)^{-1} = \sigma^2\,k^{-1}I \qquad (6.4)$$

with the following estimate of the variance of unit weight

$$s^2 = \frac{L^T\left(I-A^0\right)\left(I-B^0\right)L}{n-m-k+1} \qquad (6.5)$$

Similarly, by evaluating the pseudo inverse solution from formulas (4.9a-b) we obtain

$$\hat{X} = \frac{1}{m} A^T\left(I-B^0\right)L \qquad (6.6a)$$

and

$$\hat{Z} = B_0^{-1}\left\{I-\frac{1}{m} A A^T\left(I-B^0\right)\right\}L \qquad (6.6b)$$

We conclude that the simple formulas (6.3) and (6.4) should be preferred in the case of full rounds.

6.2 Incomplete rounds

In this case one or more rounds (or sets) of series of observations have not been completed. An example is given in section 9. This means that the simple formulas deduced in section 6.1 do not hold, and we have to apply one of the general formulas derived. Usually formulas (3.2a-b) are more practical than the special formulas (4.4)-(4.7) and (4.9a-b). However, if we merely care for an X-estimate (with no concern about error estimation) the pseudo inverse solution of (4.9a) might be advantageous, i.e.

$$\hat{X} = \left(A^T\left(I-B^0\right)A + D^T D\right)^{-1} A^T\left(I-B^0\right)L$$

where, again

$$D = c[1, 1, \ldots, 1]$$

c being an arbitrary constant.
Also formulas (4.4) and (4.7) with M = 0 should be of interest.

7. DISCUSSION

Ethrog (1981) recommends the use of the following least squares solution of directions and orientation quantities {cf. (4.9)}:

$$\hat{X} = A^{-1}_{I(I-B^0)} L \qquad (7.1a)$$

and

$$\hat{Z} = (B^T B)^{-1} B^T (I - A^0) L \qquad (7.1b)$$

where

$$A^0 = A A^{-1}_{I(I-B^0)}$$

$$A^{-1}_{I(I-B^0)} = K (K K)^{-1} K (K K)^{-1} K$$

$$K = A^T A - A^T B^0 A$$

As all least squares solutions satisfy (2.11) this holds as well for (7.1a). Thus it provides unbiased angle estimates. However, the computation of $\hat{X}$ by (7.1) is tedious and can hardly be justified from practical point of view. As an alternative we recommend the simple formulas (6.3a-b) in the case of complete rounds and formulas (3.2a-b) in the case of incomplete rounds. It should be noted that all these solutions are invariant with respect to origin and orientation of the coordinate system and no direction is arbitrarily chosen as a "zero direction".

Bjerhammar (1973, p. 218) shows that the simple formulas (6.3a-b) follow if the following condition can be satisfied:

$$A^T B \hat{Z} = 0$$

Obviously, this is the case only for completely observed rounds.

8. CONCLUSIONS

By using generalized matrix inverses we have solved for the complete set of least squares solutions of directions in the case of station adjustment of directions. It has been shown that all least squares solutions provide identical unbiased angle estimates. Practical solutions have been derived both in the case of complete and incomplete rounds of observations.

9. EXAMPLE

Consider three sets of complete observations of three targets. The system of observation equations becomes

$$A\,X + B\,Z = L - \varepsilon \qquad \text{with} \quad E\{\varepsilon\,\varepsilon^T\} = \sigma^2 I$$

$$A = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \\ 1 & & \\ & 1 & \\ & & 1 \\ 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \qquad B = \begin{bmatrix} -1 & & \\ -1 & & \\ -1 & & \\ & -1 & \\ & -1 & \\ & -1 & \\ & & -1 \\ & & -1 \\ & & -1 \end{bmatrix} \qquad L = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_9 \end{bmatrix}$$

The normal equations become the singular system

$$\begin{bmatrix} 3 & & & -1 & -1 & -1 \\ & 3 & & -1 & -1 & -1 \\ & & 3 & -1 & -1 & -1 \\ -1 & -1 & -1 & 3 & & \\ -1 & -1 & -1 & & 3 & \\ -1 & -1 & -1 & & & 3 \end{bmatrix} \begin{bmatrix} X \\ Z \end{bmatrix} = \begin{bmatrix} A^T & L \\ B^T & L \end{bmatrix}$$

Formulas (6.3a-b) yield

$$\hat{X} = \left(A^T A\right)^{-1} A^T L = \frac{1}{3} \begin{bmatrix} \ell_1 + \ell_4 + \ell_7 \\ \ell_2 + \ell_5 + \ell_8 \\ \ell_3 + \ell_6 + \ell_9 \end{bmatrix}$$

and

$$\hat{Z} = \left(B^T B\right)^{-1} B^T \left(I - A^0\right) L =$$

$$= \frac{1}{9}\begin{bmatrix} -2(\ell_1 + \ell_2 + \ell_3) + \ell_4 + \ell_5 + \ell_6 + \ell_7 + \ell_8 + \ell_9 \\ \ell_1 + \ell_2 + \ell_3 - 2(\ell_4 + \ell_5 + \ell_6) + \ell_7 + \ell_8 + \ell_9 \\ \ell_1 + \ell_2 + \ell_3 + \ell_4 + \ell_5 + \ell_6 - 2(\ell_7 + \ell_8 + \ell_9) \end{bmatrix}$$

The covariance matrix of $\hat{X}$ becomes

$$Q_{XX} = \frac{\sigma^2}{3}\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}$$

We now consider the same example, but with the last observation missing (incomplete rounds). Then we get the following system of normal equations

$$\begin{bmatrix} 3 & & & -1 & -1 & -1 \\ & 3 & & -1 & -1 & -1 \\ & & 2 & -1 & -1 & 0 \\ -1 & -1 & -1 & 3 & & \\ -1 & -1 & -1 & & 3 & \\ -1 & -1 & 0 & & & 2 \end{bmatrix} \begin{bmatrix} X \\ Z \end{bmatrix} = \begin{bmatrix} A^T & L \\ B^T & L \end{bmatrix} \qquad L = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_8 \end{bmatrix}$$

In this case we apply formulas (4.4)-(4.7) with M = 0. From the intermediate results

$$B^T \left(I - A^0\right) B = \frac{1}{6}\begin{bmatrix} 11 & 7 & -4 \\ -7 & 11 & -4 \\ -4 & -4 & 8 \end{bmatrix}$$

and

$$D_1 = [2 \quad 2 \quad 2]$$

we get

$$R = \left\{ B^T \left(I - A^0 \right) B + D_1^T D_1 \right\}^{-1} = 2 \begin{bmatrix} 5 & -1 & 0 \\ -1 & 5 & 0 \\ 0 & 0 & 4 \end{bmatrix}^{-1}$$

$$= \frac{1}{12} \begin{bmatrix} 5 & 1 & 0 \\ 1 & 5 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

$$K_{II}^{-1} = R B^T \left(I - A^0 \right) = \frac{1}{12} \begin{bmatrix} -3 & -3 & -2 & 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & -3 & -3 & -2 & 2 & 2 \\ 2 & 2 & 0 & 2 & 2 & 0 & -4 & -4 \end{bmatrix}$$

and

$$T = A_0^{-1} B K_{II}^{-1} = \frac{1}{12} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & -2 & -2 \end{bmatrix}$$

Hence,

$$\hat{X} = A_0^{-1} L - T L = \frac{1}{12} \begin{bmatrix} 4 & 0 & 0 & 4 & 0 & 0 & 4 & 0 \\ 0 & 4 & 0 & 0 & 4 & 0 & 0 & 4 \\ -1 & -1 & 6 & -1 & -1 & 6 & 2 & 2 \end{bmatrix} L =$$

$$= \begin{bmatrix} (\ell_1 + \ell_4 + \ell_7)/3 \\ (\ell_2 + \ell_5 + \ell_8)/3 \\ (\ell_3 + \ell_6)/2 + (\ell_7 + \ell_8)/6 - (\ell_1 + \ell_2 + \ell_4 + \ell_5)/12 \end{bmatrix}$$

$$\hat{Z} = K_{II}^{-1} L$$

and

$$Q_{XX} = \frac{\sigma^2}{36} \begin{bmatrix} 12 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 21 \end{bmatrix}$$

Thus the direction estimates are uncorrelated also in this case.

The angles are estimated by

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = F\,\hat{X} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \hat{X} =$$

$$= \begin{bmatrix} (\ell_2 - \ell_1 + \ell_5 + \ell_4 + \ell_8 - \ell_7)/3 \\ (\ell_3 + \ell_6 - \ell_8)/2 - (\ell_1 + 5\ell_2 + \ell_4 + 5\ell_5 - 2\ell_7)/12 \end{bmatrix}$$

This result shows that the angle estimates are unbiased.

Finally we obtain the covariance matrix of angles

$$Q_{\mu\mu} = F\,Q_{XX}\,F^T = \frac{\sigma^2}{36} \begin{bmatrix} 24 & -12 \\ -12 & 43 \end{bmatrix}$$

REFERENCES

BENNETT, G.G., 1965: The adjustment of incomplete direction observations. The Canadian Surveyor, Vol. XIX, No. 5, pp. 462-471, 1965.

BJERHAMMAR, A., 1973: Theory of Errors and Generalized Matrix Inverses. Elsevier Scientific Publ. Co., Amsterdam, 1973.

ETHROG, U., 1981: Station adjustment using the generalized least-squares method. Survey Review, Vol. 26, 202, pp. 180-191, 1981.

HELMERT, F.R., 1907: Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate. Teubner, Berlin, 1907.

HOLSEN, J., 1981: The equivalence between angles and directions and the adjustment of incomplete direction observations. Kart og Plan, 2-81, pp. 165-172, 1981.

RAO, C.R. and S.K. MITRA, 1971: Generalized Inverse of Matrices and Its Applications. J. Wiley & Sons, Inc., New York, 1971.

APPENDIX

For proof and further details we refer to Bjerhammar (1973).

Definition 1: A generalized inverse A^{-1} of a matrix A is defined by the condition

$$A A^{-1} A = A \tag{A.1}$$

The general solution to this equation is

$$\underset{\sim}{A}^{-1} = A_0^{-1} + N\left(I - AA_0^{-1}\right) + \left(I - A_0^{-1}A\right) M \tag{A.2}$$

where

A_0^{-1} = any particular inverse
N, M = arbitrary inverses of compatible dimensions
I = unit matrix

Definition 2: The pseudo inverse (Moore-Penrose inverse) of A, A_{II}^{-1}, satisfies the following equations

$$A A_{II}^{-1} A = A \tag{A.3a}$$

$$A_{II}^{-1} A A_{II}^{-1} = A_{II}^{-1} \tag{A.3b}$$

$$A A_{II}^{-1} = \left(A A_{II}^{-1}\right)^T \tag{A.3c}$$

$$A_{II}^{-1} A = \left(A_{II}^{-1} A\right)^T \tag{A.3d}$$

The pseudo inverse can be determined by

$$A_{II}^{-1} = A^T \left(A A^T\right)^{-1} A \left(A^T A\right)^{-1} A^T \tag{A.4}$$

or

$$A_{II}^{-1} = \left(A^T A + D^T D\right)^{-1} A^T =$$

$$= A^T \left(A A^T + C C^T\right)^{-1} \tag{A.5}$$

where

$$A D^T = 0 \quad \text{and} \quad A^T C = 0 \tag{A.6a}$$

and

$$\text{rank}(D) = \text{rank}(C) = \text{rank}(I - A A^{-1}) \tag{A.6b}$$

Theorem 1: The general solution for X of the consistent matrix equation

$$A X = L \tag{A.7}$$

is given by

$$\tilde{X} = A_0^{-1} L + \left(I - A_0^{-1} A\right) M \tag{A.8}$$

where A_0^{-1} is a particular inverse and M is an arbitrary vector of compatible dimension.

Theorem 2: The general least squares solution of the system

$$A X = L - \varepsilon \tag{A.9}$$

minimizing $\varepsilon^T P \varepsilon$, is given by

$$\tilde{X} = \left(A^T \underset{\sim}{P} A\right)^{-1} A^T P L \tag{A.10}$$

Corollary 1: There is a unique solution to (A.9) minimizing $\varepsilon^T P \varepsilon$ and $\tilde{X}^T \tilde{X}$. The solution is

$$\hat{X} = \left(A^T P A\right)_{II}^{-1} A^T P L \tag{A.11}$$

Corollary 2: If D satisfies (A.6) then

$$D \hat{X} = 0 \tag{A.12}$$

Proof: (A.11), (A.4) and (A.6a) yield

$$D\hat{X} = D\left(A^T P A\right)^{-1}_{II} A^T P L = D K^T \left(K K^T\right)^{-1} K\left(K^T K\right)^{-1} K^T A^T P L = 0$$

where

$$K = A^T P A$$

Corollary 3: For full rounds the solution to formula (4.2)

$$X = A_0^{-1} L - A_0^{-1} B Z$$

with

$$Z = \left(B^T\left(I - A^0\right) B + D^T D\right)^{-1}_{II} B^T\left(I - A^0\right) L$$

equals

$$\hat{X} = A_0^{-1} L$$

Proof: For full rounds

$$A_0^{-1} B = \underset{(m \times k)}{\frac{1}{k}\begin{bmatrix} D \\ D \\ \vdots \\ D \end{bmatrix}} \quad ; \quad D = [1,1,\ldots]$$

Thus it follows from (A.12) that

$$A_0^{-1} B Z = 0$$

NETWORK DENSIFICATION - PROBLEMS AND SOLUTIONS

W. M. WELSCH
Munich, Federal Republic of Germany

ABSTRACT

Problems of network densification circle around some aspects which crystallize clearly from the bibliography. These aspects are subject of the paper.

In contrasts to very large or free networks, with densification networks the boundary value problem of the given reference frame exists. Therefore, the theorist may interpret network densification as a general task of interpolation. From this point of view accuracy and reliability, partly also efficiency and practicability are criteria of the network structure or aims of optimization efforts.

However, in practice inhomogeneity and inaccuracy as well as unknown stochastic properties of the reference points cause difficult and delicate problems. On this account a great deal of the theoretical and practical treatises deal intensively with investigations how to get the densification points fitted into the given reference frame as smoothly as possible according to the respective purpose. This endeavor leads frequently to far-reaching consequences for the quality of the densification network, but results only exceptionally in an improvement or in a dynamic renovation of the existing frame network - a goal being essential from the theoretical point of view.

Thus, planning, performing, and evaluating network densification means considering the pros and cons of both theory and practice.

1. Definition and introduction

Network densification leads to densification networks. But this apparently trivial statement stands in contrast to some official instructions, according to which only at a certain order densification networks are regarded as such. In the following this definition will not be of any meaning.

As a definition, point intercalation, either single points or even whole networks, are considered densification networks, if and when their reference frame is externally fixed but not arbitrarily chosen.

This external reference system represents in some way the boundary value problem of densification networks. The most important point distinguishing between densification and free or absolute networks is to have to take into

consideration the external reference concerning its quantity as well as its quality. It influences the design and the observation but especially the evaluation of the intercalation measurements.

Some points should be emphasized.

The influence of redundant datum points is of great importance for the distortion of the configuration of the densification network, while the accuracy is strongly effected by the reference system being a stochastic but not a deterministic quantity. Most of the authors of treatises on network densification discuss especially questions resulting from these problems.

The practical realization of the theoretical results is impeded by many obstacles. A rigorous treatment of technical networks for special purposes can be achieved. However, the difficulties with the networks for general purposes established by the official surveying administrations cause the discussion of numerous technical, organizational and administrative problems.

First of all the crucial point is scientific. Nevertheless, the practical application of the suggested solutions may admittedly cause even more serious complications with far-reaching consequences. Therefore, planning, performing and evaluating network densification means considering the pros and cons of both theory and practice.

2. The stochastic reference frame

2.1 Rigorous solutions and dynamic networks

The theoretically easiest solution to the rigorous treatment of the observations which are added to the existing control network due to the network densification, doubtlessly is the combined new adjustment of all control and densification network observations in one step. This procedure is theoretically ingenious, too, because the densification observations which were added to the already adjusted control network, influence due to their additional information also the existing coordinates. They can be utilized to the improvement of the control points effecting their values as well as their accuracies.

A rigorous solution can be achieved also in two steps. BÄHR (1973) treats one solution which combines the results of the adjusted control networks as correlated observations with the observations carried out for the network

densification in one system of condition equations regarding the densification points as unknowns. This technique can also be applied if resulting from numerous configuration defects, the densification network (including the reference points) cannot be adjusted separately.

This pre-condition has to be fulfilled with a solution by PELZER (1978). He fits the separately adjusted densification network into the existing control network eliminating the discrepancies of the junction points by condition equations demanding the equality of the double coordinates. The covariance matrices are rigorously taken into account.

BÄHR (1982) aims at the same solution of connecting points without discrepancies by a generalized HELMERT-transformation. This transformation changes the coordinates of the junction points of both networks as well as all the other point coordinates so as to preserve rigorously the confirmity of the joint networks.

KUBÁČEK (1972) regards two networks as realizations of a stochastic process and informs of formulas (KUBÁČEK, 1977) for connecting stepwise any number of networks rigorously.

The strict concept of a dynamic control network (PELZER, 1980) cannot be easily put into practice since, in general, the calculation effort is enormous, the continuous change of the point coordinates is evoking many severe consequences, unsteadiness is created for instance in the cadastre system, and the problem of having permanently available the full variance-covariance matrix of the network is insurmountable.

The problem concerning the calculation effort is not invincible. The range of influence of the network densification on the existing control network can be analyzed (BÄHR, 1973) and be limited by statistical tests. Moreover, the stepwise solutions avoid the re-adjustment of the existing network which is required by the strict one-step solution. Until now the organizational, cadastral and legal problems of a dynamic control network and a dynamic cadastre have not yet been investigated. But the impact could possibly be limited by tolerances.

However, as far as the storage problem is concerned attempts were made to replace the actual rigorous covariance matrix of the control network by substitute matrices. PELZER (1978) concerning this matter, suggests storing only the diagonal elements as coordinate variances, while the covariances

are to be deduced from a model allocating higher correlations to adjacent points, lower to remote ones. Based on stochastic processes BUITEN (1975, 1978) provides even more generalized models of substitute matrices. Supported by those artificial covariance matrices it becomes possible to solve the task of network densification approximately rigorous with a justifiable effort.

Summarizing it can be stated that the time has not yet come for dynamic control networks.

2.2 "Best" network fittings

Those solutions which attempt to fit the densification net into the control network by some kind of transformation, withdraw a step further from the rigorous solution. Different aspects can be relevant.

Following the line of strict solutions, the first possibility sets those elements of the variance-covariance matrix to zero which are related to the points not being directly involved in the network densification. In consequence these points remain unchanged and do not need to be taken into account any longer. The junction points achieve definite coordinates, discrepancies do not appear.

This idea provides the basis for solutions by COOPER and LEHAY (1977, 1978) who point out that a weak reference system will be improved, while a good one will not be influenced negatively.

However, the risk is taken to disturb the neighbouring accuracy between the points of the reference system, since the change of coordinates is effective only to the junction points.

The usual approximate net-connections by redundant similarity-(HELMERT-)transformations transfer the area of disturbed neighbouring accuracy or, so to speak, the inconsistency of neighbouring accuracy to the connecting points themselves, because they obtain double coordinates after the fitting-in of the densification network. This dissatisfying fact cannot be abolished, not even by utilizing all available stochastic information, i.e. the covariance matrices of the network to be connected as weight matrices (CONZETT, 1975; SCHMID, 1980; SCHERRER, 1980). The removal of those discrepancies is, therefore, treated with special attention.

The easiest method would be to do nothing leaving the ambiguity as it is. But as a matter of fact, this solution cannot be accepted although exactly that is the usual practice.

A rigorous possibility for a weighted mean could be achieved by block adjustments (WOLF, 1978) reducing the normal equations of the two networks down to the junction points. But naturally, this procedure is far too complicated and would, nevertheless, only mean an approximate solution, provided the recalculation from the connecting points to the blocks would be neglected (KLINGENBERG et al., 1977). A mean with weights taken from the variances of the coordinates of the connecting points is possible but represents an approximate solution, too. In both cases the problems of disturbed neighbouring accuracy remain existing.

At this point the very important phenomenon of the datum-dependence of the covariance matrices which are used for all those calculations, has to be mentioned. The joint processing of covariance matrices related to different networks makes it indispensable to refer to a - possibly arbitrary - but yet common geodetic datum. Since BAARDA (1973) datum- or S-transformations are well-known.

The rigorous solutions were officially rejected in practice because they can hardly be realized for large networks, at least not at the present. In search of the "best" approximative solution, the fitting of networks by collocation provides a recommendable possibility as seen not only from the practical but also from the theoretical point of view. Already WOLF (1979) pointed at the interrelations between adjustment and collocation. He defined in this sense the point-intercalation as a problem of prediction. Consequently the collocations has been used in some variety to overcome the problem of discrepancies in a manner suitable to practice.

HALMOS et al. (1974) describes the problem as a local filter-problem. KLINGENBERG (1977) distributes the coordinate discrepancies in the sense of an interpolation - collocation to the vicinity of the junction points. Based on the work of KRAUS (1970), CAROSIO (1980) deals with the fitting of old triangulation networks of 4. order into newer and better networks of higher order. In a transition zone overlapping the old and new triangulations, the point coordinates are changed in such way that the fitting is as smooth as possible. The range of this transition zone can be varied and adapted to the actual problem by setting up suitable covariance functions.

BUITEN (1978) describes the procedure of network-fitting by collocation methods in detail. Preferably a similarity transformation is taken as trend-function. On it a signal is superposed which removes systematic parts from the discrepancies as regional effects. The remaining discrepancies are due to the noise. Its contribution is small but still troublesome. To get rid of it, the covariance matrix of the noise parameters is set zero. Therefore this procedure represents the special collocation case of noisefree prediction. The final values of the coordinates of the densification network points which are not nodal points are calculated from the trend and signal by interpolation. The covariance matrices of the signal are substitute matrices following the conditions of stochastic processes.

Referring to this method an approximation solution is obtained by removing the discrepancies with the help of weights depending on point intervals (RUOPP, 1972; BLACHNITZKY, 1974; OVERHOFF, 1981). Pointwise, several transformations are carried out. The weights for the mean of the results are taken from the mutual distances between the respective points. The procedure is not rigorous, of course, but practicable and conserving neighbouring accuracies.

Summarizing, the above mentioned methods attempt to approach a rigorous solution by taking the following measures: evaluating the network densification also the reference points are effected, at least the junction points; fitting-in the densification network as smoothly as possible is aimed at; the neighbouring accuracies, particularly at the network margins, are attempted to be conserved. This endeavour is supported by utilizing the covariance matrices of both the densification and the control network.

2.3 Hierarchical structure of networks

A further step of simplifying the mathematical model of network-fitting is achieved by regarding the coordinates of the junction points of the reference frame as unchangeable, and that from the very beginning of the set-up. If, additionally, the covariance matrix of the reference frame is assumed to vanish, the classical concept of hierarchical network structures is put into existence (PELZER, 1980). The strictness of this structure is a bit released tolerating, that variances and covariances related to the junction points effect only the densification network but not the reference points. Unchangeable control point coordinates are of great advantage. However, the constraints to the intercalation network are sometimes hardly bearable,

especially if a weak and badly distorted control network has to serve a precise densification network as reference frame.

Already GOTTHARDT (1957) points at the misjudgement of the accuracies of the intercalated points if the covariance matrix of the reference points is neglected. Subsequently the effect of the stochastic properties of the reference frame has frequently been investigated; see (GLEINSVIK (1963), ACKERL (1965), LENZMANN (1981), MIERLO (1982) et al. The conclusion is always the same: the use of reference points as deterministic quantities leads to overestimating the accuracies of the densification points. Considering the stochastic properties of the reference points, a strict solution for the covariance matrix, i.e. for the accuracy of the densification net, is achievable (WELSCH and KELM, 1980). If an actual covariance matrix of the reference system is not available, a substitute matrix can be simulated and utilized. BUITEN and RICHARDUS (1982) go into the details of network densification by adjustment, a method widely applied to official network densification in the Netherlands.

As seen from those investigations, a strictly hierarchical network structure is the more risky, the more stages the hierarchy has. Therefore, analyses were made (HILGER, 1966) whether certain orders of densification can be skipped without loss of accuracy. Others examined if the traditional principle "from major to minor" should be given up or even be reversed (GERKE and PELZER, 1970). Theoretical considerations support this idea. However, it has not succeeded in practice.

3. Planning of network densifications

Some aspects of the establishment and structure of reference- and densification networks have already been given. AUGATH (1976) has carried out extensive reflections on that topic. He starts from the various requirements of different users coming from cartography, cadastre and natural science. On grounds of the changed and precise measurement techniques (PELZER, 1981) a set-up of networks in two stages may be sufficient on principle leading to a distortionless frame- and densification multipurpose network satisfying most of the requirements.

The arrangement of the observations, namely that one of the densification network, results from criteria of minimizing costs and maximizing accuracy and reliability (AUGATH, 1977 and 1980); FÖRSTNER, 1981). The results can

definitely be not in coincidence with each other, since uniformly distributed and weighted observations are essential for the optimization or reliability, while depending on local effects varying weights can be achieved as an optimum result for accuracies.

An overview on planning aspects of densification networks according to criteria of network qualities is given by AUGATH (1982) and HEUS (1982).

For the general mathematical treatment of optimization tasks it is referred here to the special bibliography.

4. Observation and calculation procedures, criteria for qualification

The changes of the geodetic measurement and evaluation techniques are treated by PELZER (1981) in detail. With terrestrial network densification in spite of the high accuracy achieved by electronic range finders, the method of triangulation has not been pushed away. Thus, for the intercalation of network-units combined observation procedures have become characteristic. In some cases densification by traverse networks is preferred, too. For extended surveys of details by electronic tacheometers free-stationing has asserted. This has been in coincidence with the development of the tacheometer systems themselves including their good storage facilities, and with the generation of comfortable computer program systems. Those development trends have been pursued by the OBERKOCHEN (1970 - 1981) courses on electronic tacheometer techniques.

Along with the terrestrial methods of network densification also photogrammetric techniques have been developed. Photogrammetry has reached a level (GRÜN, 1979) which enables analytical methods to compete with terrestrial ones according to the same criteria of accuracy, reliability and economy. Even if this is in general not unrestrictedly the case (HVIDEGAARD, 1979; ADLER and GREENFELD, 1981) the discussion (LEUPIN and MONVOISIN, 1981; NEISECKE, 1981) proves that for many purposes photogrammetric methods can support and complete terrestrial techniques as a partner of equal quality.

As far as the data processing by computer programs is concerned (WELSCH, 1981a), some common features can be stated. While the common characteristic of some earlier program systems was the endeavour to establish rigorous functional models, the most modern programs reflect the following trends: automatical calculation of approximate coordinates, continuous testing of the reliability of the observations and data-snooping, partly a posteriori

variance estimation, working up network points in large numbers, stating of the accuracies of all adjustment results etc. The attempt has been become obvious to integrate the network adjustment into the general geodetic data flow, to process also large networks in medium-sized computers by subtle storage techniques, to combine simple handling with mathematical and statistical strictness. Some program systems intending to those characteristics may be mentioned: MEISSL and STUBENVOLL (1977), KATRIN (BENNING, 1979), TRINA (FÖRSTNER, 1979), TANA (GRÜNDIG, 1980).

The theory and the techniques applied are, however, of general validity. They are not restricted to densification networks. Therefore, it seems to be advisible not to discuss here date processing and qualification criteria. A review on all those procedures is given by ALBERDA (1980).

However, the only proposal made here is to quantify and to qualify by strain analysis techniques (WELSCH, 1981b) the distortions which a densification network undergoes if it is fitted into a reference frame by constraints.

5. Official densification networks

The explanations of the last chapters have been related to theoretical problems which are essential to densification networks according to their definition. The set-up of some solutions may be difficult as seen from the mathematical point of view but it can be mastered. Strict solutions can be applied to local and possibly regional networks. There are no difficulties on principle.

However, this is different with the extended official networks established by state or national surveying administrations. Those networks show great inhomogeneities due to great differences of their development, structure, quality and status of updating caused by various technical, administrative and political reasons. Further difficulties are added due to their extension and continuous and interlocking changes. Of course, the circumstances vary from one country to another. The paper of BLACHNITZKY (1982) might give a vital impression of the today's situation within the Federal Republic of Germany.

The existing national control networks are not sufficient for all requirements. Therefore, a systematic renewal should be aimed at (PELZER, 1981). The today achieved standard of observation and evaluation techniques is so good as to accomplish anational control network free of tensions. The ad-

vantages of such a network with accuracies better than the uncertainty of the physical point definition are obvious. Many suggestions have been made for a systematic renewal. Although many individuals (KNOOP, 1972; KRIEFALL, 1975; PÖTZSCHNER, 1977; PELZER, 1981; and others) as well as the official surveying administrations have been anxious, the final success can be questioned for political, organizational and economical reasons.

By all means, the official control networks are essential for many technical and scientific problem solutions. However, their purpose above all is to be the basis for the realization of official tasks. Consequently, their treatment has to conform possibly also to the latest state of knowledge but largely to other specific criteria.

6. Conclusions

The purely theoretical and technical problems connected with network densification are solved to a great deal. The problem of the stochastic reference frame can be treated rigorously or with efficient approximate solutions meeting practical requirements. The recent developments of instrumentation techniques have lead to observation methods which guarantee point accuracies within the range of the point definition. The observations can be optimized according to criteria such as economy, accuracy and reliability. They can be processed by computer programs even in large numbers. Due to the progress of photogrammetry analytical methods compete with terrestrial ones.

However, The existing official control networks do not satisfy the today's requirements of engineering and natural science. Consequently, technical networks for special purposes have become necessary. They can meet the demands since the full theoretical and technical knowledge can be applied. For these cases the official networks serve only as a general reference. Systematic network renewals which could nationwide provide tensionfree and precise geodetic fundamentals, cannot be expected in the near future. Consequently, official networks are restrained to tasks of sovereignty.

REFERENCES

ACKERL, F., 1965: Der Einfluß fehlerhafter Festpunkte auf trigonometrisch abgeleitete Neupunkte. Zeitschrift für Vermessungswesen 90 (1965), S. 279-283.

ADLER, R., GREENFELD, J., 1981: Experiences in applying photogrammetric methods as alternative to ground control surveys. XVI. FIG-Congress, paper 507.6, Montreux 1982.

ALBERDA, J.E., 1980: A review of analysis techniques for engineering survey control schemes. Industrial and Engineering Survey Conference, London 1980, pp. 6.2/1-42.

AUGATH, W., 1976: Untersuchungen zum Aufbau geodätischer Lagenetze. Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie, Photogrammetrie und Kartographie an der TU Hannover, Heft 72, Hannover 1976.

AUGATH, W., 1977: Über den Aufbau geodätischer Lagenetze nach Zuverlässigkeitskriterien. International Symposium on Optimization of Design and Computation of Control Networks, Sopron 1977.

AUGATH, W., 1980: Optimierungskriterien und deren Auswirkung in Netzen der Landesvermessung. In: Pelzer (ed.): Geodätische Netze in Landes- und Ingenieurvermessung. Wittwer, Stuttgart 1980, S. 479-498.

AUGATH, W., 1982: Accuracy and reliability measures concerning design and qualification of densification networks. Meeting of FIG-Study Group 5 B on Survey Control Networks, Aalborg 1982.

BAARDA, W., 1973: S-Transformations and criterion matrices. Netherland Geodetic Commission, New Series, Vol. 5, No. 1, Delft 1973.

BÄHR, H.-G., 1973: Zum Ausgleichungsverfahren bei der Ergänzung bestehender Netze. Deutsche Geodätische Kommission, Heft C 187, München 1973.

BÄHR, H.-G., 1982: Eine verallgemeinerte Lösung für die überbestimmte Koordinatentransformation zwischen ausgeglichenen geodätischen Netzen. Zeitschrift für Vermessungswesen 107 (1982), S. 194-206.

BENNING, W., 1979: Komplexe Auswertung trigonometrischer Netze - Das Programmsystem KATRIN. Nachrichten des öffentlichen Vermessungsdienstes Nordrhein-Westfalen 12 (1979) 2, S. 59-75.

BLACHNITZKY, K., 1974: Punktspezifische, gewichtete Helmerttransformation. Unveröffentlicht.

BLACHNITZKY, K., 1982: Status and problems of official horizontal densification networks as shown by the Federal Republic of Germany. Meeting of FIG-Study Group 5 B on Survey Control Networks, Aalborg 1982.

BUITEN, H.J., 1975: The junction of nets by adjustment. Nederlands Geodetisch Tijdschrift 10 (1975) 5, pp. 181-191.

BUITEN, H.J., 1978: Junction of nets by collocation. Manuscripta geodaetica, Vol. 3 (1978), pp. 253-297.

BUITEN, H.J., RICHARDUS, P., 1982: Junctions of control survey nets by adjustment compared with coordinate transformation. Meeting of FIG-Study Group 5 B on Survey Control Networks, Aalborg 1982.

CAROSIO, A., 1980: Anwendung von Interpolationsverfahren in der Landestriangulation. Vermessung, Photogrammetrie und Kulturtechnik 78 (1980) 10, S. 405-407.

CONZETT, R., 1975: Über lineare Transformationen von Triangulationsnetzen. Vermessung, Photogrammetrie und Kulturtechnik 73 (1975) 314,S. 265-268.

COOPER, M.A.R., LEAHY, F.J., 1977: Phased adjustment of engineering control surveys. XV. FIG-Congress, Stockholm 1977.

COOPER, M.A.R., LEAHY, F.J., 1978: An adjustment of a second-order network. Survey Review XXIV (1978) 87, pp. 224-233.

FÖRSTNER, W., 1979: Das Rechenprogramm TRINA für geodätische Lagenetze in der Landesvermessung. Nachrichten des öffentlichen Vermessungsdienstes Nordrhein-Westfalen 12 (1979 2, S. 125-166.

FÖRSTNER, W., 1981: On the reliability of 3rd and 4th order network densifications. VI. International Symposium on Geodetic Networks and Computations, München 1981.

GERKE, K., PELZER, H., 1970: Netzverdichtung oder Netzaufbau aus Funktionen gemessener Größen? Allgemeine Vermessungsnachrichten 77 (1970) 8, S. 309-316.

GLEINSVIK, P., 1963: Über die Genauigkeit der trigonometrischen Punktbestimmung mit besonderer Berücksichtigung der Einwirkung der Festpunkte. Zeitschrift für Vermessungswesen 88 (1963), S. 441-446.

GOTTHARDT, E., 1957: Zum Genauigkeitsabfall bei der getrennten Ausgleichung von Netzteilen. Zeitschrift für Vermessungswesen 82 (1957) 3, S. 79-85.

GRÜN, A., 1979: Zur Anwendung der modernen Präzisionsphotogrammetrie in der Netzverdichtung und Katastervermessung. Zeitschrift für Vermessungswesen 104 (1979) 3, S. 86-97.

GRÜNDIG, L., 1980: Programmsystem TANA zur Tachymeternetzausgleichung. VIII. Internationaler Kurs für Ingenieurvermessung, paper B 1, Zürich 1980.

HALMOS, F., KÁDÁR, J., KARSAY, F., 1974: Local adjustment by least squares filtering. Bulletin Géodésique (1974) 111, pp. 21-51.

HEUS, H.M. de, 1982: Quality related problems of densification networks. Meeting of FIG-Study Group 5 B on Survey Control Networks, Aalborg 1982.

HILGER, E., 1966: Die netzweise Ausgleichung einzelner oder mehrerer Triangulationsordnungen in einem Guß. Deutsche Geodätische Kommission, Heft C 98, Frankfurt 1966.

HVIDEGAARD, J., 1979: Densification of trigonometric nets - practical experiences with bundle adjustment. Photogrammetria 35 (1979) 1, pp. 29-36.

KLINGENBERG, F., 1977: Zur Fixpunktverdichtung in schon gemessenem Gebiet. Vermessung, Photogrammetrie und Kulturtechnik 75 (1977) 6, S. 195-198.

KLINGENBERG, F., LEUENBERGER, Ch., 1977: Erneuerung der Triangulation IV. Ordnung auf dem Gebiet der Stadt Zürich. Vermessung, Photogrammetrie und Kulturtechnik 75 (1977) 2, S. 69-74.

KNOOP, H., 1972: Einpassung der Meßergebnisse in vorhandene Netze. In: Technik der elektro-optischen Tachymetrie. Wichmann, Neue Folge, Karlsruhe (1972) 19, S. 115-131.

KRAUS, K., 1970: Interpolation nach kleinsten Quadraten in der Photogrammetrie. Zeitschrift für Vermessungswesen 95 (1970) 9, S. 387-390.

KRIEFALL, H.-J., 1975: Die Verbindung der modernen Landesvermessung mit Kataster- und Ingenieurvermessungen. In: Geodätische Woche Köln 1975. Wittwer, Stuttgart 1976, S. 66-74.

KUBÁČEK, L., 1972: The Connecting of Nets with respect to their statistical properties. Mitteilungen der geodätischen Institute der TH Graz, Graz (1972) 11.2, S. 291-296.

KUBÁČEK, L., 1977: Sequential connecting of networks. International Symposium on Optimization of Design and Computation of Control Networks, Sopron 1977.

LENZMANN, L., 1981: Zur Korrelation bei Punkteinschaltung und Netzausgleichung. Zeitschrift für Vermessungswesen 106 (1981) 5, S. 253-259.

LEUPIN, M., MONVOISIN, Ph., 1981: Grundsätzliche Probleme bei der Anwendung der Photogrammetrie in der Netzverdichtung. Zeitschrift für Vermessungswesen 106 (1981) 6, S. 316-321.

MEISSL, P., STUBENVOLL, K., 1977: Ein Computer-Programmsystem zur Verdichtung trigonometrischer Netze. Mitteilungen der geodätischen Institute der TU Graz, Graz (1977) 25.

MIERLO, J. van, 1982: Design and adjustment of densification networks. VII. Conference of Southern African Surveyors, Johannisburg 1982.

NEISECKE, O., 1981: Vergleich terrestrischer und photogrammetrisch gewonnener Lagenetze. Zeitschrift für Vermessungswesen 106 (1981) 11, S. 597-606. Entgegnung und Stellungnahme in Zeitschrift für Vermessungswesen 107 (1982) 6, S. 243-252.

OBERKOCHEN, 1970 - 1981: Veröffentlichungen zu den Oberkochener Instrumentenkursen. Sammlung Wichmann, Neue Folge, Karlsruhe 1970 - 1981.

OVERHOFF, W., 1981: Zum Problem der Umformung mit Abstandsgewichten (Restklaffenbeseitigung). Nachrichten des öffentlichen Vermessungsdienstes Nordrhein-Westfalen, Heft 2, 1981.

PELZER, H., 1978: Über die Fortführung und Verdichtung geodätischer Festpunktfelder. In: Wissenschaftliche Arbeiten der Lehrstühle Geodäsie, Photogrammetrie, Kartographie der TU Hannover, Hannover (1978) 83, S. 97-110.

PELZER, H., 1980: Dynamische oder hierarchische Netze? In: Pelzer (ed.): Geodätische Netze in Landes- und Ingenieurvermessung. Wittwer, Stuttgart 1980, S. 389-411.

PELZER, H., 1981: Wandlungen in der geodätischen Aufnahme- und Auswertetechnik. Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, Heft 100, Hannover 1981, S. 63-69.

PÖTZSCHNER, W., 1977: Die Anpassung des Lagefestpunktfeldes an die heutigen Genauigkeitsansprüche. Zeitschrift für Vermessungswesen 102 (1977) 4, S. 157-165.

RUOPP, M., 1972: Unmittelbares Abstecken nach Landeskoordinaten mit elektronischem Tachymeter und Tischkomputer. In: Technik der elektro-optischen Tachymetrie. Wichmann, Karlsruhe (1972) 19, S. 132-140.

SCHERRER, R., 1980: Auffelderung eines Netzes unter der Bedingung minimalisierter Klaffungen. Mitteilungen des Institutes für Geodäsie und Photogrammetrie der ETH Zürich, Zürich (1980) 29, S. 52-69.

SCHMID, H.H., 1980: Vom freien zum gelagerter Netz. Mitteilungen des Institutes für Geodäsie und Photogrammetrie der ETH Zürich, Zürich (1980) 29, S. 1-51.

WELSCH, W., 1981a: Present state of activities in the field of control survey networks. XVI. FIG-Congress, paper 501.3, Montreux 1981.

WELSCH, W., 1981b: Description of homogeneous horizontal strains and some remarks to their analysis. VI. International Symposium on Geodetic Networks and Computations, München 1981.

WELSCH, W., KELM, R., 1980: Numerische Studie über den Einfluß der Kovarianzmatrix der Altpunkte auf die Genauigkeit der Einschaltung eines Neupunktes. In: Pelzer (ed.): Geodätische Netze in Landes- und Ingenieurvermessung. Wittwer, Stuttgart (1980), S. 227-238.

WOLF, H., 1969: Prädiktion und Punktausgleichung. Zeitschrift für Vermessungswesen 94 (1969) 5, S. 165-169.

WOLF, H., 1978: The Helmert Block Method - its origin and development. 2nd International Symposium on Problems Related to the Redefinition of North American Geodetic Networks, Washington 1978.

THE INCREMENT METHOD

G. ZLATANOV
Sofia, Bulgaria

1. Introduction

The increment method was proposed by the author of this paper in 1969, and was later improved. The essence of the increment method consists in the modification of the method of I. Pranis-Praniewich [1], but can be applied also within the frame of non-linear version of the least squares method, allowing together with the evaluation of the unknowns, also the preparation of the information necessary for the computation of the mean errors both of all unknowns and also of the functions of groups of unknowns. It is developed with a view to using a computer with an internal storage which is insufficient to include the entire system of normal equations.

2. Linear Model

First we consider the increment method in its linear version and then we carry out an additional specialization for non-linear case.

We assume that the whole system of correction equations is divided into t groups, whereby the i-group of equations is presented in the matrix form

$$V_i = A_{i,i-1} X_{i-1,i} + A_{ii} X_{ii} + A_{i,i+1} X_{i,i+1} - L_i \qquad (1)$$

$$\left(i = \overline{1,t}\right)$$

where: V_i, L_i are the vectors of the corrections, respectively the free terms of the i-system of equations,

X_{ii} is the vector of the partially independent unknowns in the i-group,

$X_{i,i+1}$ is the vector of the unknowns, connecting the i-group with the (i+1)-group,

$A_{i,j}$ are the design matrices.

For the sake of simplicity we assume that equations for normed corrections (reduced to unit weight) are presented in (1). Hence the requirement of the least squares method, applied to equation (1) appears in the form

$$\sum_{i=1}^{t} (V_i, V_i) = \min . \qquad (2)$$

The essence of the increment method consists in solving the i-group of equations with a view to the condition

$$\left(V_{i-1,i}, V_{i-1,i}\right) + (V_i, V_i) = \min , \qquad (3)$$

where $V_{i-1,i}$ is a vector of equivalent corrections of the (i-1)-system of equations.

The computation process in the increment method is conducted for <u>direct</u> and <u>revers</u> courses. In the case of direct course the final solution including the variance-covariance matrix for the unknowns, is obtained only for the last group. For all the remaining groups provisional values of the unknowns are obtained. In the revers course corrections to the unknowns are calculated, as well as the variance-covariance matrix for the unknowns, belonging to the separate groups of equations.

<u>2.1. Direct Computation Course</u>

For the sake of simplicity we assume that t=2. Hence, taking (1) into account, the correction equations are to be presented in the form

$$\begin{aligned} V_1 &= A_{11} X_{11} + A_{12} X_{12} - L_1 \\ V_2 &= A_{21} X_{21} + A_{22} X_{22} - L_2 . \end{aligned} \qquad (4)$$

To begin with, the first group is solved, proceeding from the consideration

$$V_1' = A_{11} X_{11}' + A_{12} X_{12}' - L_1 \qquad (5)$$

$$(V_1', V_1') = \min . \qquad (5a)$$

The solution of (5), carrying (5a) in mind, leads to a system of normal equations with the following extended matrix

$$N' = \begin{bmatrix} A_{11}^t A_{11} & A_{11}^t A_{12} & -A_{11}^t L_1 \\ A_{12}^t A_{11} & A_{12}^t A_{12} & -A_{12}^t L_1 \\ -L_1^t A_{11} & -L_1^t A_{12} & -L_1^t L_1 \end{bmatrix} . \qquad (6)$$

After extracting a matrix root of N' we obtain the triangular matrix

$$R' = \begin{bmatrix} R'_{11} & R'_{12} & S'_1 \\ & r_{12} & s_{12} \\ & & W' \end{bmatrix} \qquad (7)$$

where $R'^t_{11} R'_{11} = A^t_{11} A_{11}$, $R'_{12} = R'^{t^{-1}}_{11} (A^t_{11} A_{12})$, $S'_1 = -R'^{t^{-1}}_{11} (A^t_{11} L_1)$

(7a)

$$r^t_{12} r_{12} = A^t_{12} A_{12} - R'^t_{12} R'_{12}, \quad s_{12} = -r^{t^{-1}}_{12} (A^t_{12} L_1 + R'^t_{12} S'_1)$$

$$W' = \sqrt{L^t_1 L_1 - S'^t_1 S'_1 - s^t_{12} s_{12}} = \sqrt{V'^t_1 V'_1} .$$

The matrix to the right of (7) corresponds to the following elimination system

$$R'_{11} X'_{11} + R'_{12} X'_{12} + S'_1 = 0$$

$$r_{12} X'_{12} + s_{12} = 0 . \qquad (8)$$

The joint solution of the two groups of correction equations from (4), which satisfies the condition

$$(V_1, V_1) + (V_2, V_2) = \min \qquad (9)$$

leads to the following system of normal equations

$$N_{11} X_{11} + N_{12} X_{12} + N_{13} X_{22} + F_1 = 0$$

$$N^t_{12} X_{11} + N_{22} X_{12} + N_{23} X_{22} + F_2 = 0 \qquad (10)$$

$$N^t_{13} X_{11} + N^t_{23} X_{12} + N_{33} X_{22} + F_3 = 0$$

where

$$N_{11} = A_{11}^t A_{11} \ , \quad N_{12} = A_{11}^t A_{12} \ , \quad N_{13} = 0 \ , \quad F_1 = -A_{11}^t L_1$$

$$N_{22} = A_{12}^t A_{12} + A_{21}^t A_{21} \ , \quad N_{23} = A_{21}^t A_{22} \ , \quad F_2 = -A_{12}^t L_1 + A_{21}^t L_2 \tag{10a}$$

$$N_{33} = A_{22}^t A_{22} \ , \quad F_3 = -A_{22}^t L_2 \ .$$

Excluding X_{11} from (10), we obtain

$$\begin{aligned} N_{22.1} X_{12} + N_{23.1} X_{22} + F_{2.1} &= 0 \\ N_{23.1}^t X_{12} + N_{33.1} X_{22} + F_{3.1} &= 0 \end{aligned} \tag{11}$$

where

$$\begin{aligned} N_{ik.1} &= N_{ik} - N_{1i}^t N_{11}^{-1} N_{1k} \\ F_{i.1} &= F_i - N_{1i}^t N_{11}^{-1} F_1 \ . \end{aligned} \tag{11a}$$

The more detailed presentation of the matrix $N_{ik.1}$ and $F_{i.1}$ from (11) leads to

$$N_{22.1} = A_{12}^t A_{12} + A_{21}^t A_{21} - A_{12}^t A_{11} \left(A_{11}^t A_{11}\right)^{-1} A_{11}^t A_{12}$$

$$N_{23.1} = A_{21}^t A_{22}$$

$$F_{2.1} = -A_{12}^t L_1 - A_{21}^t L_2 + A_{12}^t A_{11} \left(A_{11}^t A_{11}\right)^{-1} A_{11}^t L_1 \tag{12}$$

$$N_{33.1} = A_{22}^t A_{22}$$

$$F_{3.1} = A_{22}^t L_2 \ .$$

<u>Statement I</u> The system (11) can be obtained also proceeding from the equations

$$\begin{aligned} V_{12} &= r_{12} X_{12} + s_{12} \\ V_2 &= A_{21} X_{12} + A_{22} X_{22} - L_2 \end{aligned} \tag{13}$$

by solving them under the condition

$$(V_{12},V_{12})+(V_2,V_2)=\min \ . \qquad (13a)$$

And indeed the problem (13), (13a) leads to the following normal equations system

$$\begin{aligned} M_{11}\,X_{12}+M_{12}\,X_{22}+G_1 &= 0 \\ M_{12}^t\,X_{12}+M_{22}\,X_{22}+G_2 &= 0 \end{aligned} \qquad (14)$$

where with a view to (7a) and (12) we have

$$\begin{aligned} M_{11} &= r_{12}^t\,r_{12}+A_{21}^t\,A_{21}=N_{22.1} \\ M_{12} &= A_{21}^t\,A_{22}=N_{23.1} \\ G_1 &= r_{12}^t\,s_{12}-A_{21}^t\,L_2=-A_{12}^t\,L_1-A_{21}^t\,L_2-R'^t_{12}\,S'_1= \\ &= -A_{12}^t\,L_1-A_{21}^t\,L_2+A_{12}^t\,A_{11}\left(A_{11}^t\,A_{11}\right)^{-1}A_{11}^t\,L_1= \\ &= F_{2.1} \\ M_{22} &= A_{22}^t\,A_{22}=N_{33.1} \\ G_2 &= -A_{22}^t\,L_2=F_{3.1} \ . \end{aligned} \qquad (15)$$

Formulae (15) prove statement I.

Statement I expresses the method of connecting the two groups correction equations and is the base of direct computation course after the increment method.

For the extended matrix of the system of normal equations (14), we have, taking into account (15)

$$N'' = \begin{bmatrix} M_{11} & M_{12} & G_1 \\ M_{12}^t & M_{22} & G_2 \\ G_1^t & G_2^t & s_{12}^t s_{12}+L_2^t L_2 \end{bmatrix} . \qquad (16)$$

After extracting a matrix root from N" we obtain

$$R'' = \begin{bmatrix} R''_{11} & R''_{12} & S''_1 \\ & R''_{22} & S''_2 \\ & & W'' \end{bmatrix} \tag{17}$$

where

$$R''^{t}_{11} R''_{11} = M_{11} = r^{t}_{12} r_{12} + A^{t}_{21} A_{21}$$

$$R''_{12} = \left(R''^{-1}_{11}\right)^{t} M_{12} = \left(R''^{-1}_{11}\right)^{t} A^{t}_{21} A_{22}$$

$$S''_1 = \left(R''^{-1}_{11}\right)^{t} G_1 = \left(R''^{-1}_{11}\right)^{t} \left(r^{t}_{12} s_{12} - A^{t}_{21} L_2\right)$$

$$\begin{aligned} R''^{t}_{22} R''_{22} &= M_{22} - R''^{t}_{12} R''_{12} = \\ &= A^{t}_{22} A_{22} - \left(A^{t}_{21} A_{22}\right)^{t} \left(r^{t}_{12} r_{12} + A^{t}_{21} A_{21}\right)^{-1} \cdot \\ &\quad \cdot \left(A^{t}_{21} A_{22}\right) \end{aligned} \tag{17a}$$

$$\begin{aligned} S''_2 &= \left(R''^{-1}_{22}\right)^{t} \left(G_2 - R''^{t}_{12} S''_1\right) = -\left(R''^{-1}_{22}\right)^{t} \left(A^{t}_{22} L_2 + \right. \\ &\quad \left. + R''^{t}_{12} S''_1\right) \end{aligned}$$

$$\begin{aligned} W'' &= \sqrt{L^{t}_2 L_2 + s^{t}_{12} s_{12} - S''^{t}_1 S''_1 - S''^{t}_2 S''_2} = \\ &= \sqrt{V^{t}_{12} V_{12} + V^{t}_2 V_2} \ . \end{aligned}$$

By making use of statement I we can draw the entire elimination scheme (common for the two groups). We have

$$\begin{aligned} &R'_{11} X_{11} + R'_{12} X_{12} + S'_1 = 0 \\ &\text{---------------------------------} \\ &R''_{11} X_{12} + R''_{12} X_{22} + S''_1 = 0 \\ &R''_{22} X_{22} + S''_2 = 0 \ . \end{aligned} \tag{18}$$

Under these circumstances the quantity $W^2 = V^{t}_1 V_1 + V^{t}_2 V_2$ can be written

$$\begin{aligned} W^2 &= V^{t}_1 V_1 + V^{t}_2 V_2 = \\ &= L^{t}_1 L_1 + L^{t}_2 L_2 - S'^{2}_1 - S''^{2}_1 - S''^{2}_2 \ . \end{aligned} \tag{19}$$

From the last equation of (7a) and the last equation of (17a) we obtain the relation

$$W'^2 + W''^2 = V'^t_1 V'_1 + V^{1t}_{12} V_{12} + V^t_2 V_2 =$$
$$= L^t_1 L_1 - S'^t_1 S'_1 - s^t_{12} s_{12} + L^t_2 L_2 + s^t_{12} s_{12} - \quad (20)$$
$$- S''^t_1 S''_1 - S''^t_2 S''_2 = W^2 \ .$$

Formula (20) is a check of the correct connection of the two groups of equations.

The computations for the direct course and, as we shall see below, for the reverse course as well, can be simplified if in the computation of the second group, X_{12} is assumed as a vector of the provisional (constant) values.

Thus be setting

$$X_{12} = X'_{12} + dX_{12} \quad (21)$$

equations (13) appear in the form

$$V_{12} = r_{12}\, dX_{12}$$
$$V_2 = A_{21}\, dX_{12} + A_{22}\, X_{22} - L'_2 \quad (22)$$

where

$$L'_2 = L_2 - A_{21}\, X'_{12} \quad (22a)$$

and we have taken into account the validity of the second equation in (8).

Under these circumstances by introducing in (18) the relations (21) and

$$X_{11} = X'_{11} + dX_{11} \quad (23)$$

and taking into account the first equation in (8) we have

$$
\begin{aligned}
R'_{11}\, dX_{11} + R'_{12}\, dX_{12} &= 0 \\
\text{---------------------------------} & \\
R''_{11}\, dX_{12} + R''_{12}\, X_{22} + \overline{S''}_1 &= 0 \\
R''_{22}\, X_{22} + S''_2 &= 0
\end{aligned}
\tag{24}
$$

where

$$
\overline{S''}_1 = S''_1 + R''_{11}\, X'_{12} \ . \tag{24a}
$$

Therefore, for the direct computation course, provisional values of the unknowns are obtained from each elimination system (with the exception of the elimination system for the last group of equation). The computations can also be checked for each group, applying the formula

$$
V^t_{i-1,1}\, V + V'^t_i\, V' = W_i^2 \tag{25}
$$

whereby the computation of the corrections (the left part of (25)) is carried out with the provisional values of the unknowns $\left(X_{i-1,i}, X_{ii}, X_{i,i+1}\right)$.

2.2 Reverse Computation Course

As already mentioned the corrections of the provisional values of the unknowns and the values required for the evaluation of the accuracy are determined in the reverse course. For this purpose, from the setended matrix of the elimination system for the last group (in this case R") in accordance with the algorithm of Banachiewicz [2] we go over to the matrix

$$
\widetilde{R''} = \begin{bmatrix} R''_{11} & R''_{12} & \overline{S''}_1 \\ & R''_{22} & S''_s \\ & & 1 \end{bmatrix} . \tag{26}
$$

The inverse of (26) is

$$
\widetilde{R''}^{-1} = \begin{bmatrix} R''^{11} & R''^{12} & dX_{12} \\ & R''^{22} & dX_{22} \\ & & 1 \end{bmatrix} \tag{27}
$$

where R''^{11} and R''^{22} are triangular matrices. The matrix $\widetilde{R''}^{-1}$ contains both the values of the unknowns and the information necessary for the variance-covariance matrix for these unknowns. For the latter we have

$$Q'' = \begin{bmatrix} R''^{11} & R''^{12} \\ . & R''^{22} \end{bmatrix} \cdot \begin{bmatrix} R''^{11} & R''^{12} \\ . & R''^{22} \end{bmatrix}^t . \tag{28}$$

The knowledge of Q", respectively of $\widetilde{R''}^{-1}$ answers the question of the accuracy of arbitrary functions of unknowns belonging to the second group.

As can be seen nothing of unknown nature was met at the computation of the last (second) group. However the case with the other group is different. There the matrix of the coefficients in front of the unknowns in the elimination system, namely

$$R' = \begin{bmatrix} R'_{11} & R'_{12} \\ . & r'_{12} \end{bmatrix} \tag{29}$$

is not suitable for obtaining the variance-covariance matrix Q' in the same easy way, as with the second group. It is obvious, that in R' the matrix R'_{22}, of the same type, is to be put in the place of r_{12}, but so as

$$R_{22}^{t\,-1} \cdot R_{22}^{-1} = Q'_{12}, R'_{22} = R_{22}^t \tag{30}$$

where Q'_{12} is the variance-covariance matrix for dX_{12}, which can be obtained of the computation of the second group, applying the formula

$$Q'_{12} = R''^{11} \cdot R''^{11^t} + R''^{12} \cdot R''^{12^t} . \tag{31}$$

Since R'_{22} is to be of the type of r_{12}, then obviously its inverse is to be of the same type. By applying equation (30) we see immediately that the evaluation of R_{22}^{-1} is reduced to finding the lower triangular matrix root.

Statement II If Q is a symmetrical and positively determined matrix, then it possesses a lower triangular matrix root K, so that

$$K^t \cdot K = Q \tag{32}$$

designating

$$K = \sqrt{Q} \quad . \tag{32a}$$

The existence of lower triangular matrix root is proved just as this is done for the existence of the elementary matrix root.

The computation itself of the lower triangular root is conducted after the formulas

$$k_{ii} = \sqrt{q_{ii} - \sum_{h=i+1}^{n} k_{hi}^2} \quad , \; i = n, n-1, \ldots, 1$$

$$k_{ij} = \left(q_{ij} - \sum_{h=j+1}^{n} k_{hi}\, k_{hj} \right) : k_{ii} \tag{33}$$

$$i = n, n-1, \ldots, 1 \quad , \quad j = i-1, i-2, \ldots, 1$$

$$k_{ij} = 0 \text{ by } i < j$$

where n is the order of Q.

Therefore, the transformed R' appears as

$$R' = \begin{bmatrix} R'_{11} & R'_{12} \\ \cdot & R'_{22} \end{bmatrix} . \tag{34}$$

Formula (34) shows that after we obtain R_{22}^{-1} as lower triangular root of Q'_{12} we have to find its transposed inverse as well. In practice this is however not necessary. Instead of R' we can draw up the matrix

$$R'_0 = \begin{bmatrix} R'_{11} & R'_{12} \\ \text{----------} & \\ \cdot & R'^{22} \end{bmatrix} \tag{35}$$

where the designation $R'^{22} = R_{22}^{-1^t}$ is introduced.

In R'_0 we seem to have begun the inversion, obtaining the last cell of the inverse matrix. It remains to "complete" the inversion, so that

$$R'^{-1} = \begin{bmatrix} R'^{11} & R'^{12} \\ . & R'^{22} \end{bmatrix} \tag{36}$$

in order to obtain the unknowns of the first group, it is suitable to carry out the transformation on the matrix R' from (7) by setting

$$R'_0 = \begin{bmatrix} R'_{11} & R'_{12} & S'_1 \\ \cdot & R'^{22} & dX_{12} \\ . & \cdot & 1 \end{bmatrix} . \tag{37}$$

"Completing" the inversion of R'_0 we obtain

$$R'^{-1} = \begin{bmatrix} R'^{11} & R'^{12} & dX_{11} \\ \cdot & R'^{22} & dX_{12} \\ . & \cdot & 1 \end{bmatrix} . \tag{38}$$

This "completing" of the inversion R'_0 is conducted after the formulae

$$\begin{aligned} r^{ii} &= 1/r_{ii} \qquad \left(i = \overline{1,m}\right) \\ r^{ij} &= -r^{ii} \sum_{h=i+1}^{j} r_{ih}\, r^{hj} \\ i &= m, m-1, \ldots, 1 \qquad \text{by} \quad j = \overline{m+1, m+n} \\ i &= j-1, j-2, \ldots, 1 \quad \text{by} \quad j = \overline{1,m} \\ j &= m+n, m+n-1, \ldots, m, m-1, \ldots, 1 \\ m &= \text{order}\,(R'_{11}), \quad n = \text{order}\,(R'_{22}) + 1 \ . \end{aligned} \tag{39}$$

By introducing this computation process in a computer, it is entirely pos-

sible to use one hand the same data file for the matrices N' (without the elements under the main diagonal), R' and R'^{-1}, which lead to maximum saving of memory.

3. Non-Linear Model

The correction equations, for example for the first group of observations, can be presented also in the following non-linear model

$$v_i = F_i(x_1, x_2, \ldots, x_m ; y_1, y_2, \ldots, y_n) - l_i \qquad (40)$$

$$\left(i = \overline{1,q} \; ; \; q \geq m+n\right)$$

where $x_j \left(j = \overline{1,m}\right)$ are the partially independent unknowns, and $y_k \left(k = \overline{1,n}\right)$ the connecting unknowns.

By introducing the vectors

$$\begin{aligned} \overline{v} &= |v_1, v_2, \ldots, v_q| \\ \overline{l} &= |l_1, l_2, \ldots, l_q| \\ \overline{x} &= |x_1, x_2, \ldots, x_m| \\ \overline{y} &= |y_1, y_2, \ldots, y_n| \end{aligned} \qquad (41)$$

and the vector-function

$$\overline{F} = |F_1, F_2, \ldots, F_q| \qquad (42)$$

equations (40) can be replaced by the functional

$$\overline{v} = \overline{F}(\overline{x};\overline{y}) - \overline{l} \; . \qquad (43)$$

Let us assume that the approximate s-solution $\overline{x}^s, \overline{y}^s$ is found for (43), which satisfies the condition

$$(\overline{v}^s, \overline{v}^s) = \min \; . \qquad (43a)$$

For the s+1 solution we can write

$$\overline{x}^{s+1} = \overline{x}^{s} + d\overline{x}^{s+1}$$
$$\overline{y}^{s+1} = \overline{y}^{s} + d\overline{y}^{s+1} \quad . \tag{44}$$

Assuming that the functions $F_i \left(i = \overline{1,q}\right)$ are continuously differentiable in some convex field, containing $\overline{x}^{s}$, $\overline{x}^{s+1}$, $\overline{y}^{s}$ and $\overline{y}^{s+1}$ desintegrate the functional $\overline{F}(\overline{x};\overline{y})$ in the powers of the small norm of the vectors $d\overline{x}$ and $d\overline{y}$, limiting ourselves to the linear members

$$\overline{F}\left(\overline{x}^{s+1};\overline{y}^{s+1}\right) = \overline{F}\left(\overline{x}^{s};\overline{y}^{s}\right) + \overline{F}'_{x}\left(\overline{x}^{s};\overline{y}^{s}\right) d\overline{x}^{s+1} + \overline{F}'_{y}\left(\overline{x}^{s};\overline{y}^{s}\right) d\overline{y}^{s+1} \quad . \tag{45}$$

In formula (45) $\overline{F}'_{x}$ and $\overline{F}'_{y}$ are matrices in a sense similar to the matrices of Jacobi. By introducing (45) in (43) we obtain

$$\overline{v}^{s+1} = \overline{F}'_{x}\left(\overline{x}^{s};\overline{y}^{s}\right) d\overline{x}^{s+1} + \overline{F}'_{y}\left(\overline{x}^{s};\overline{y}^{s}\right) d\overline{y}^{s+1} + \overline{F}\left(\overline{x}^{s};\overline{y}^{s}\right) - \overline{l} \quad . \tag{46}$$

Equation (46) is solved under the condition

$$\left(\overline{v}^{s+1},\overline{v}^{s+1}\right) = \min \quad . \tag{47}$$

The process of consecutive approximations, reminding of the method of Newton, is to continue until the norm of the two vectors dx and dy becomes acceptably small (so as not to influence the values of the elements of the matrices $\overline{F}'_{x}$ and $\overline{F}'_{y}$). So ends the direct course of computation of the first group of correction equations, represented by the equations (40) and (43) respectively.

No corrections are introduced in the computation of the second group of vectors $\overline{y}^{w}$ although such corrections are to be calculated in the consecutive approximations at the computations of the second group of equations. In the last computation of the second group, the corrections of $\overline{y}^{w}$ are to be taken into account and they give the final values of these unknowns calculated in the direct course.

The aforesaid shows that the non-linear model of the increment method does not lead to basic differences from the linear model; therefore everything that was said for the connection of the groups and the computation of the variance-covariance matrices of the unknowns in the linear model remains valid also in the non-linear model.

REFERENCES

1. PRANIS-PRANIEWICH, I.J., 1956: Handbook for the adjustment of triangulations. Geodezizdat, Moscow, 1956 (in Russian)
2. BANACHIEWICH, T., 1959: Krakowian calculation. PWN, Warsaw, 1959 (in Polish)
3. ZLATANOV, G., 1968: Krakowian calculation. "Technika", Sofia, 1968 (in Bulgarian)
4. ZLATANOV, G., 1969: The increment method. Official bull., No. 3, Sofia, 1969 (in Bulgarian)
5. ZLATANOV, G., 1970: Accuracy evaluation in the increment method. Transactions of the GUGK, No. 2, Sofia, 1970 (in Bulgarian)
6. ZLATANOV, G., 1979: Computer technique in geodesy. "Technika", Sofia, 1979 (in Bulgarian)

Schriftenreihe

Wissenschaftlicher Studiengang Vermessungswesen

Hochschule der Bundeswehr München

Bisher erschienene Hefte:

(Die Hefte erscheinen in zwangloser Folge)

Nr. 1/78 A. Schödlbauer :

Curriculum für den wissenschaftlichen Studiengang Vermessungswesen der Hochschule der Bundeswehr München

Nr. 2/78 A. Chrzanowski and E. Dorrer :

Proceedings "Standards and Specifications for Integrated Surveying and Mapping Systems", Workshop held in Munich, 1-2 June 1977

Nr. 3/78 W. Caspary und A. Geiger :

Untersuchungen zur Leistungsfähigkeit elektronischer Neigungsmesser

Nr. 4/79 E. Baumann, W. Caspary, H. Dupraz / W. Niemeier / H. Pelzer, E. Kuntz / G. Schmitt, W. Welsch :

Seminar über Deformationsmessungen

Nr. 5/81 K. Torlegård :

Accuracy Improvement in Close Range Photogrammetry

Nr. 6/82 W. Caspary und W. Welsch :

Beiträge zur großräumigen Neutrassierung

Nr. 7/82 K. Borre and W.M. Welsch :

Proceedings "Survey Control Networks", Meeting of FIG-Study Group 5B, Aalborg, 7 - 9 July 1982

Bezugsnachweis: Studiengang Vermessungswesen
im Fachbereich Bauingenieur- und Vermessungswesen
Hochschule der Bundeswehr München
Werner-Heisenberg-Weg 39
8014 Neubiberg