Regression Analysis for Clusters in Gene-Environment Networks based on Ellipsoidal Calculus and Optimization

In Commemoration of our Teacher and Friend
Professor Alexander Rubinov

Erik Kropat[1], Gerhard-Wilhelm Weber[2] and Jan-J. Rückmann[2]

[1]Institute for Theoretical Computer Science, Mathematics and Operations Research
Universität der Bundeswehr München, Neubiberg, Germany
E-mail: erik.kropat@unibw.de

[2]Institute of Applied Mathematics, Middle East Technical University, Turkey;
Faculty of Economics, Business and Law, University of Siegen, Germany;
Center for Research on Optimization and Control, University of Aveiro, Portugal;
Faculty of Science, Universiti Teknologi Malaysia (UTM), Skudai, Malaysia
E-mail: gweber@metu.edu.tr

[3]School of Mathematics
University of Birmingham, United Kingdom
E-mail: J.Ruckmann@bham.ac.uk

Abstract. In this paper, we discuss regression models for gene-environment networks under ellipsoidal uncertainty. Functionally related groups of genes and environmental factors are identified by clustering techniques and the corresponding uncertain states are represented in terms of ellipsoids. The time-dependent expression values are determined by a regulatory system where the interactions between the clusters are defined by (affine-) linear coupling rules. Ellipsoidal calculus is applied to determine explicit representations of the uncertain multivariate states of the system. Various regression models are introduced for an estimation of the unknown system parameters which depend on uncertain (ellipsoidal) measurement data. Herewith, we offer an *Elliptic Operations Research*, in which we analyze the structure of the optimization problems obtained, especially, in view of their solvability by semidefinite programming and interior point methods, we discuss the structural frontiers and research challenges, and we conclude with an outlook.

Keywords. Gene-Environment Networks – Ellipsoidal OR – Exact Clustering – Semidefinite Programming – Dynamical Systems – Computational Statistics – Uncertainty – Identification – Systems Biology – Computational Biology

AMS (MOS) subject classification: 93A30, 92D10, 90C34

1 Introduction

Genetic systems are often characterized by the presence of a high number of variables and parameters resulting in an extraordinary complexity of the underlying

regulatory networks. The technological progress of the last decades and the development of high throughput technologies resulted in a generation of massive quantities of data that can be used to enlighten the hidden network structure. Along with these developments, the availability of extremely large data sets also challenged mathematics and bioinformatics and led to new methods for an analysis of highly interconnected systems. However, for many years, mathematical studies focussed on genetic items, neglecting the importance of the regulating effects of additional *environmental factors*. In this context, the concept of *gene-environment networks* was developed [42, 43, 48, 49, 50, 53, 54, 56, 57], supported by twin research and various other studies. These networks mathematically express and assess the importance of factors such as, e.g., poison, pollution, stress, but also measurements in education, health care, environmental protection and in the development of the financial markets, for an estimation of the biological states. In fact, this impact also holds true in the reverse direction, and as a special class of gene-environment networks we mention *eco-finance networks* [24, 51].

Gene-environment networks are composed of two distinct groups of variables. The first group consists of the genes (proteins or other cell components), the most relevant data items to be observed. In addition, many other cell components and factors like toxins, transcription factors, radiation, etc. play an important role in genetic networks. This additional group of environmental items exerts a strong influence on the behaviour of the genes. However, the complexity of the data, its volume and variability, demands for highly developed methods that allow to detect relationships between changes in gene expression and in additional environmental factors. In recent years, gene-environment networks under various kinds of uncertainty have been analyzed. These models were based on interval arithmetics, but also spline regression and stochastic differential equations. We note that this concept encompasses a broader scope of scientific inquiry than when genetic networks were initially considered. Many other examples from system biology and life sciences refer to the so-called *target-environment networks*, where environmental effects have to be considered. Among them are, e.g., *metabolic networks* [13, 34, 54], *immunological networks* [19], networks in *toxicogenomics* [28], but also *social-* and *ecological networks* [18]. We refer to [17, 22, 38, 39, 40, 59, 60] for applications, practical examples and numerical calculations.

Gene-environment networks are generally built on measurement data from microarray experiments and environmental measurements which are always effected by errors and random noise. Error intervals can capture the deviations from specific data values by imposing individual bounds on each variable. Various regression models have been developed and studied with the help of *generalized Chebychev approximation* and, equivalently to that approximation, *semi-infinite* and even *generalized semi-infinite optimization* [45, 46, 47, 48, 49, 50, 54, 55, 57]. However, error intervals referring to single variables do not reflect correlations of the multivariate data within specific clusters of genetic and environmental items. A natural extension is a representation of uncertainty in terms of ellipsoidal sets which are directly related to multivariate Gaussian distributions and the corresponding covariance matrices. Ellipsoids have proved to be very flexible with respect to correlations

of the data, whereas intervals and parallelpipes usually come from a perspective where stochastic dependencies among any two of the errors made in the measurements of the expression values of genes and environmental levels are not taken into account explicitly [6]. Moreover, paraxial sets are usually smaller than the ellipsoids and their orthogonal projections into the 2-dimensional Cartesian planes, respectively [6]. Indeed, those confidence ellipsoids are obtained with respect to stochastic dependencies of the error variables. Those dependencies are the case in reality, e.g., in microarray experiments and in environmental studies as well. In reverse, any ellipsoid can be inscribed into a sufficiently large parallelpipe which, in addition, could be suitably located and directed in space around its eigenaxes.

Clustering and classification allow to identify certain clusters of data commonly exerting influence on other groups of genes and/or environmental factors. The uncertain expression values of these clusters are represented by ellipsoidal states. The multiple interactions of genetic and environmental items are determined by (affine-) linear coupling rules which define a time-discrete *regulatory model*. The future states and predictions of the linear model can be calculated by *ellipsoidal calculus* and explicit representations of these predictions in terms of centers and shape matrices can be determined by an *iterative procedure*. Various *regression models* are introduced for an estimation of the unknown parameters of the regulatory model, which heavily depend on the cluster structure. These models compare the actual measurements and the (ellipsoidal) predictions of the linear model; however, in a set-theoretic sense. Beside a notion of distance between centers of ellipsoids, nonnegative criteria functions associated with the shape matrix of the ellipsoid are used in the objective function of the regression model. The trace and the determinant are examples of such measures and they establish the basis for the definition of regression models for parameter estimation of the regulatory system. *Semi-definite programming* as well as *conic programming* and *interior point methods* can then be applied for solution.

The paper is organized as follows: In Section 2, we review basic facts about ellipsoidal calculus. In Section 3, the time-discrete model under ellipsoidal uncertainty is introduced. After formulating an algorithm for predictions of ellipsoidal future states of genetic and environmental factors, we turn to regression analysis in Section 4. Various regression models are introduced. Their solvability and the related optimization methods are discussed in Section 5.

2 Ellipsoidal Calculus

The parameter-dependent time-discrete model for gene-environment networks of this paper and the corresponding regression models will be based on *ellipsoidal calculus*. In this first section, we shortly review the basic operations needed to deal with ellipsoidal uncertainty such as *sums*, *intersections (fusions)* and *affine-linear transformations* of ellipsoids. The family of ellipsoids in $\mathbb{R}^p$ is closed with respect to affine-linear transformations but neither the sum nor the intersection is generally ellipsoidal, so both must be approximated by ellipsoidal sets.

2.1 Ellipsoidal Descriptions

An *ellipsoid* in $\mathbb{R}^p$ will be parameterized in terms of its center $c \in \mathbb{R}^p$ and a symmetric non-negative definite *configuration (or shape) matrix* $\Sigma \in \mathbb{R}^{p\times p}$ as

$$\mathcal{E}(c,\Sigma) = \{\Sigma^{1/2}u + c \,|\, \|u\| \leq 1\},$$

where $\Sigma^{1/2}$ is any matrix square root satisfying $\Sigma^{1/2}(\Sigma^{1/2})^T = \Sigma$. When Σ is of full rank, the non-degenerate ellipsoid $\mathcal{E}(c,\Sigma)$ may be expressed as

$$\mathcal{E}(c,\Sigma) = \{x \in \mathbb{R}^p \,|\, (x-c)^T\Sigma^{-1}(x-c) \leq 1\}.$$

The eigenvectors of Σ point in the directions of principal semiaxes of $\mathcal{E}$. The lengths of the semiaxes of the ellipsoid $\mathcal{E}(c,\Sigma)$ are given by $\sqrt{\lambda_i}$, where λ_i are the eigenvalues of Σ for $i = 1,\ldots,p$. The volume of the ellipsoid $\mathcal{E}(c,\Sigma)$ is given by $\operatorname{vol}\mathcal{E}(c,\Sigma) = V_p\sqrt{\det(\Sigma)}$, where V_p is the volume of the unit ball in $\mathbb{R}^p$, i.e.,

$$V_p = \begin{cases} \dfrac{\pi^{p/2}}{(p/2)!} & \text{, for even } p, \\ \dfrac{2^p\pi^{(p-1)/2}((p-1)/2)!}{p!} & \text{, for odd } p. \end{cases}$$

2.2 Affine Transformations

The family of ellipsoids is closed with respect to *affine transformations*. Given an ellipsoid $\mathcal{E}(c,\Sigma) \subset \mathbb{R}^p$, matrix $A \in \mathbb{R}^{m\times p}$ and vector $b \in \mathbb{R}^m$ we get $A\mathcal{E}(c,\Sigma) + b = \mathcal{E}(Ac+b, A\Sigma A^T)$. Thus, ellipsoids are preserved under affine transformation. If the rows of A are linearly independent (which implies $m \leq p$), and $b = 0$, the affine transformation is called *projection* [27].

2.3 Sums of K Ellipsoids

Given K bounded ellipsoids of $\mathbb{R}^p$, $\mathcal{E}_k = \mathcal{E}(c_k,\Sigma_k)$, $k = 1,\ldots,K$, their *geometric (Minkowksi) sum* $\mathcal{E}_1 + \mathcal{E}_1 = \{z_1 + z_2 \,|\, z_1 \in \mathcal{E}_1,\ z_2 \in \mathcal{E}_2\}$ is not generally an ellipsoid. However, it can be tightly approximated by parameterized families of external ellipsoids. We adapt the notion of the minimal trace ellipsoid from [16] and introduce the outer ellipsoidal approximation $\mathcal{E}(\sigma,P) = \oplus_{k=1}^K\mathcal{E}_k$ containing the sum $\mathcal{S} = \sum_{k=1}^K \mathcal{E}_k$ of ellipsoids which is defined by

$$\sigma = \sum_{k=1}^K c_k$$

and

$$P = \left(\sum_{k=1}^K \sqrt{\operatorname{Tr}\Sigma_k}\right)\left(\sum_{k=1}^K \frac{\Sigma_k}{\sqrt{\operatorname{Tr}\Sigma_k}}\right).$$

2.4 Intersection of Ellipsoids

The intersection of two ellipsoids is generally not an ellipsoid. For this reason we replace this set by the outer ellipsoidal approximation of minimal volume and adapt the notion of *fusion* of ellipsoids from [36]. Given two non-degenerate ellipsoids $\mathcal{E}(c_1, \Sigma_1)$ and $\mathcal{E}(c_2, \Sigma_2)$ in $\mathbb{R}^p$ with $\mathcal{E}(c_1, \Sigma_1) \cap \mathcal{E}(c_2, \Sigma_2) \neq \emptyset$ we define an ellipsoid

$$\begin{aligned}\mathcal{E}_\lambda(c_0, \Sigma_0) := \{x \in \mathbb{R}^p \mid \lambda(x - c_1)^T \Sigma_1^{-1}(x - c_1) \\ + (1 - \lambda)(x - c_2)^T \Sigma_2^{-1}(x - c_2) \leq 1\},\end{aligned}$$

where $\lambda \in [0, 1]$. The ellipsoid $\mathcal{E}_\lambda(c_0, \Sigma_0)$ coincides with $\mathcal{E}(c_1, \Sigma_1)$ and $\mathcal{E}(c_2, \Sigma_2)$ for $\lambda = 1$ and $\lambda = 0$, respectively. In order to determine a tight external ellipsoidal approximation $\mathcal{E}_\lambda(c_0, \Sigma_0)$ of the intersection of $\mathcal{E}(c_1, \Sigma_1)$ and $\mathcal{E}(c_2, \Sigma_2)$, we introduce

$$\mathcal{X} := \lambda\Sigma_1^{-1} + (1 - \lambda)\Sigma_2^{-1}$$

and

$$\tau := 1 - \lambda(1 - \lambda)(c_2 - c_1)^T \Sigma_2^{-1} \mathcal{X}^{-1} \Sigma_1^{-1}(c_2 - c_1).$$

The ellipsoid $\mathcal{E}_\lambda(c_0, \Sigma_0)$ is given by the center

$$c_0 = \mathcal{X}^{-1}(\lambda\Sigma_1^{-1}c_1 + (1 - \lambda)\Sigma_2^{-1}c_2)$$

and shape matrix

$$\Sigma_0 = \tau\mathcal{X}^{-1}.$$

The *fusion* of $\mathcal{E}(c_1, \Sigma_1)$ and $\mathcal{E}(c_2, \Sigma_2)$, whose intersection is a nonempty bounded region, is defined as the ellipsoid $\mathcal{E}_\lambda(c_0, \Sigma_0)$ for the value $\lambda \in [0, 1]$ that minimizes its volume [36]. The fusion of $\mathcal{E}(c_1, \Sigma_1)$ and $\mathcal{E}(c_2, \Sigma_2)$ is $\mathcal{E}(c_1, \Sigma_1)$, if $\mathcal{E}(c_1, \Sigma_1) \subset \mathcal{E}(c_2, \Sigma_2)$; or $\mathcal{E}(c_2, \Sigma_2)$, if $\mathcal{E}(c_2, \Sigma_2) \subset \mathcal{E}(c_1, \Sigma_1)$; otherwise, it is $\mathcal{E}_\lambda(c_0, \Sigma_0)$ defined as above where λ is the only root in $(0, 1)$ of the following polynomial of degree $2p - 1$:

$$\begin{aligned}&\tau(\det \mathcal{X}) \operatorname{Tr}(\operatorname{co}(\mathcal{X})(\Sigma_1^{-1} - \Sigma_2^{-1})) - p(\det \mathcal{X})^2 \\ &\times (2c_0^T\Sigma_1^{-1}c_1 - 2c_0^T\Sigma_2^{-1}c_2 + c_0^T(\Sigma_2^{-1} - \Sigma_1^{-1})c_0 - c_1^T\Sigma_1^{-1}c_1 + c_2^T\Sigma_2^{-1}c_2) = 0.\end{aligned}$$

Here, $\operatorname{co}(\mathcal{X})$ denotes the matrix of cofactors of $\mathcal{X}$. Since $\mathcal{X}^{-1} = \operatorname{co}(\mathcal{X})/\det \mathcal{X}$, we represent this polynomial as

$$\begin{aligned}&\tau(\det \mathcal{X})^2 \operatorname{Tr}(\mathcal{X}^{-1}(\Sigma_1^{-1} - \Sigma_2^{-1})) - p(\det \mathcal{X})^2 \\ &\times (2c_0^T\Sigma_1^{-1}c_1 - 2c_0^T\Sigma_2^{-1}c_2 + c_0^T(\Sigma_2^{-1} - \Sigma_1^{-1})c_0 - c_1^T\Sigma_1^{-1}c_1 + c_2^T\Sigma_2^{-1}c_2) = 0.\end{aligned}$$

We note that it is also possible to define an inner ellipsoidal approximation. The method of finding the internal ellipsoidal approximation of the intersection of two ellipsoids is described in [44].

3 Gene-Environment Systems under Ellipsoidal Uncertainty

The previous preparations enable us to go on with our *Ellipsoidal Operations Research* via the dynamics which we present subsequently.

3.1 Time-Discrete Gene-Environment Systems

We consider time-discrete gene-environment regulatory systems with n genes and m environmental factors. Often, certain groups of genes and environmental items can be identified which exert a more or less regulating influence on other groups of data items. For this reason, the set of genes is divided in R disjoint clusters $C_r \subset \{1,\ldots,n\},\ r = 1,\ldots,R$ in a preprocessing step of clustering. Similarly, the set of all environmental items is divided in S disjoint clusters $D_s \subset \{1,\ldots,m\},\ s = 1,\ldots,S$. Here, motivated by the application in the genetic context, we focus on non-overlapping clusters and assume a strict sub-division of the variables, so that the relations $C_{r_1} \cap C_{r_2} = \emptyset$ for all $r_1 \neq r_2$ and $D_{s_1} \cap D_{s_2} = \emptyset$ for all $s_1 \neq s_2$ are fulfilled. We note that our method is also applicable for overlapping clusters. This particular case will be addressed in a forthcoming paper [25]. There exist many clustering techniques in the fields of unsupervised learning and data mining and we do not focus on a specific method. One reasonable approach consists in *nearest neighbor methods* [21] in combination with the discriminant adaptive nearest neighbor metric that allow an identification of ellipsoidal neighborhoods [20]. Also clustering methods based on *minimal spanning trees* [8] or methods from *nonsmooth optimization* [7] are suited for the ellipsoidal approach.

Each cluster corresponds to a functionally related group of genes or environmental factors and the uncertain states of these cluster are represented in terms of ellipsoids

$$X_r = \mathcal{E}(\mu_r, \Sigma_r) \subset \mathbb{R}^{|C_r|}, \quad E_s = \mathcal{E}(\rho_s, \Pi_s) \subset \mathbb{R}^{|D_s|}.$$

The dynamics and interactions between the various clusters of genetic and environmental items are given by the linear model

$$\left.\begin{aligned} X_j^{(\kappa+1)} &= \xi_{j0} + \left(\bigoplus_{r=1}^{R} \mathcal{A}_{jr}^{GG} X_r^{(\kappa)}\right) + \left(\bigoplus_{s=1}^{S} \mathcal{A}_{js}^{EG} E_s^{(\kappa)}\right) \\ E_i^{(\kappa+1)} &= \zeta_{i0} + \left(\bigoplus_{r=1}^{R} \mathcal{A}_{ir}^{GE} X_r^{(\kappa)}\right) + \left(\bigoplus_{s=1}^{S} \mathcal{A}_{is}^{EE} E_s^{(\kappa)}\right) \end{aligned}\right\} \quad (EC)$$

with $\kappa \geq 0$ and $j = 1,2,\ldots,R$, $i = 1,2,\ldots,S$.

In this time-discrete system, four types of cluster interactions and regulating effects are involved: (GG) gene $\rightarrow$ gene, (EG) environment $\rightarrow$ gene, (GE) gene $\rightarrow$ environment, (EE) environment $\rightarrow$ environment. The system (EC) is defined by (affine) linear coupling rules, what implies that all future states of genetic and environmental clusters are ellipsoids themselves. In particular, the sums $\bigoplus_{r=1}^{R} \mathcal{A}_{jr}^{GG} X_r^{(\kappa)}$

and $\bigoplus_{s=1}^{S} \mathcal{A}_{js}^{EG} E_s^{(\kappa)}$ describe the *cumulative effects* of all genetic and environmental clusters exerted on the elements of cluster C_j in a set theoretic or ellipsoidal sense. In the same way, the (ellipsoidal) sums $\bigoplus_{r=1}^{R} \mathcal{A}_{ir}^{GE} X_r^{(\kappa)}$ and $\bigoplus_{s=1}^{S} \mathcal{A}_{is}^{EE} E_s^{(\kappa)}$ refer to the additive genetic and environmental effects on cluster E_i. The *degree of connectivity* between the individual clusters is given by the (unknown) interactions matrices $\mathcal{A}_{jr}^{GG} \subset \mathbb{R}^{|C_j|\times|C_r|}$, $\mathcal{A}_{js}^{EG} \subset \mathbb{R}^{|C_j|\times|D_s|}$, $\mathcal{A}_{ir}^{GE} \subset \mathbb{R}^{|D_i|\times|C_r|}$, and $\mathcal{A}_{is}^{EE} \subset \mathbb{R}^{|D_i|\times|D_s|}$. These matrices are in turn sub-matrices of the general interaction matrices $\mathcal{A}^{GG} \in \mathbb{R}^{n\times n}$, $\mathcal{A}^{EG} \in \mathbb{R}^{n\times m}$, $\mathcal{A}^{GE} \in \mathbb{R}^{m\times n}$, $\mathcal{A}^{EE} \in \mathbb{R}^{m\times m}$. Since we assume disjoint clusters, the aforementioned sub-matrices are distinct building blocks of the general interactions matrices where the elements of the corresponding clusters define the position within the general interactions matrices. The intercepts $\xi_{j0} \in \mathbb{R}^{|C_j|}$ and $\zeta_{i0} \in \mathbb{R}^{|D_i|}$ are disjoint components of the vectors $\xi_0 = (\xi_{10}, \ldots, \xi_{R0})^T \in \mathbb{R}^n$ and $\zeta_0 = (\zeta_{10}, \ldots, \zeta_{S0})^T \in \mathbb{R}^m$, respectively. We note that the initial values of the linear system (EC) can be defined by the first genetic and environmental measurements, i.e., $X_j^{(0)} = \overline{X}_j^{(0)}$ and $E_i^{(0)} = \overline{E}_i^{(0)}$.

Remark: We note that the interconnections between the various clusters of genes and environmental factors can be represented in terms of a highly interconnected *gene-environment network*. The nodes of this network can be identified with the clusters, weighted by the (time-dependent) ellipsoidal states. The branches between the nodes (or clusters) are weighted by the matrices and intercept vectors of the linear coupling rules of model (EC). Although the parameters and the weights of the branches are static, the evolution of the ellipsoidal states turns the gene-environment network in a dynamic graph model. Hereby, network analysis and concepts from discrete mathematics become applicable and features like connectedness, cycles and shortest paths can be investigated [23].

3.2 Algorithm

The linear system (EC) introduced in Section 3.1 refers to the time-discrete behaviour of the ellipsoidal states of clusters and environmental variables. By applying ellipsoidal calculus from Section 2, the centers and configuration matrices of the predictions $X_j^{(\kappa+1)}$ and $E_s^{(\kappa+1)}$ of (ellipsoidal) genetic and environmental cluster states can be calculated iteratively. Throughout this section we assume $\kappa \geq 0$. The state of the genetic cluster C_j, $j = 1, 2, \ldots, R$, is given by the ellipsoid

$$X_j^{(\kappa+1)} = \mathcal{E}\left(\mu_j^{(\kappa+1)}, \Sigma_j^{(\kappa+1)}\right)$$

with center

$$\mu_j^{(\kappa+1)} = \xi_{j0} + \sum_{r=1}^{R} A_{jr}^{GG} \mu_r^{(\kappa)} + \sum_{s=1}^{S} A_{js}^{EG} \rho_s^{(\kappa)}$$

and configuration matrix

$$\Sigma_j^{(\kappa+1)} = \left(\sqrt{\mathrm{Tr}\,\mathcal{G}_j^{(\kappa)}} + \sqrt{\mathrm{Tr}\,\mathcal{H}_j^{(\kappa)}}\right) \cdot \left(\frac{\mathcal{G}_j^{(\kappa)}}{\sqrt{\mathrm{Tr}\,\mathcal{G}_j^{(\kappa)}}} + \frac{\mathcal{H}_j^{(\kappa)}}{\sqrt{\mathrm{Tr}\,\mathcal{H}_j^{(\kappa)}}}\right),$$

where

$$\mathcal{G}_j^{(\kappa)} = \left(\sum_{r=1}^{R} \sqrt{\mathrm{Tr}\left(A_{jr}^{GG}\Sigma_r^{(\kappa)}(A_{jr}^{GG})^T\right)}\right) \cdot \left(\sum_{r=1}^{R} \frac{A_{jr}^{GG}\Sigma_r^{(\kappa)}(A_{jr}^{GG})^T}{\sqrt{\mathrm{Tr}\left(A_{jr}^{GG}\Sigma_r^{(\kappa)}(A_{jr}^{GG})^T\right)}}\right),$$

$$\mathcal{H}_j^{(\kappa)} = \left(\sum_{s=1}^{S} \sqrt{\mathrm{Tr}\left(A_{js}^{EG}\Pi_s^{(\kappa)}(A_{js}^{EG})^T\right)}\right) \cdot \left(\sum_{s=1}^{S} \frac{A_{js}^{EG}\Pi_s^{(\kappa)}(A_{js}^{EG})^T}{\sqrt{\mathrm{Tr}\left(A_{js}^{EG}\Pi_s^{(\kappa)}(A_{js}^{EG})^T\right)}}\right).$$

Similarly, the states of the environmental cluster D_i, $i = 1, 2, \ldots, S$, can be represented in terms of the ellipsoid

$$E_i^{(\kappa+1)} = \mathcal{E}\left(\rho_i^{(\kappa+1)}, \Pi_i^{(\kappa+1)}\right)$$

with center

$$\rho_i^{(\kappa+1)} = \zeta_{i0} + \sum_{r=1}^{R} A_{ir}^{GE}\mu_r^{(\kappa)} + \sum_{s=1}^{S} A_{is}^{EE}\rho_s^{(\kappa)}$$

and configuration matrix

$$\Pi_i^{(\kappa+1)} = \left(\sqrt{\mathrm{Tr}\,\mathcal{M}_i^{(\kappa)}} + \sqrt{\mathrm{Tr}\,\mathcal{N}_i^{(\kappa)}}\right) \cdot \left(\frac{\mathcal{M}_i^{(\kappa)}}{\sqrt{\mathrm{Tr}\,\mathcal{M}_i^{(\kappa)}}} + \frac{\mathcal{N}_i^{(\kappa)}}{\sqrt{\mathrm{Tr}\,\mathcal{N}_i^{(\kappa)}}}\right),$$

where

$$\mathcal{M}_i^{(\kappa)} = \left(\sum_{r=1}^{R} \sqrt{\mathrm{Tr}\left(A_{ir}^{GE}\Sigma_r^{(\kappa)}(A_{ir}^{GE})^T\right)}\right) \cdot \left(\sum_{r=1}^{R} \frac{A_{ir}^{GE}\Sigma_r^{(\kappa)}(A_{ir}^{GE})^T}{\sqrt{\mathrm{Tr}\left(A_{ir}^{GE}\Sigma_r^{(\kappa)}(A_{ir}^{GE})^T\right)}}\right),$$

$$\mathcal{N}_i^{(\kappa)} = \left(\sum_{s=1}^{S} \sqrt{\mathrm{Tr}\left(A_{is}^{EE}\Pi_s^{(\kappa)}(A_{is}^{EE})^T\right)}\right) \cdot \left(\sum_{s=1}^{S} \frac{A_{is}^{EE}\Pi_s^{(\kappa)}(A_{is}^{EE})^T}{\sqrt{\mathrm{Tr}\left(A_{is}^{EE}\Pi_s^{(\kappa)}(A_{is}^{EE})^T\right)}}\right).$$

4 Regression Analysis Under Ellipsoidal Uncertainty

4.1 The Regression Problem

The unknown system parameters of model (EC) are given by the entries of the interaction matrices $\mathcal{A}_{jr}^{GG}$, $\mathcal{A}_{js}^{EG}$, $\mathcal{A}_{ir}^{GE}$, and $\mathcal{A}_{is}^{EE}$ and the intercepts ξ_{j0} and ζ_{i0}. For an estimation of these parameters and, thus, for regression analysis, we have to compare the predictions of the linear model (EC) and the observations from genetic and environmental measurements. The (ellipsoidal) measurements

$$\overline{X}_r^{(\kappa)} = \mathcal{E}\big(\overline{\mu}_r^{(\kappa)}, \overline{\Sigma}_r^{(\kappa)}\big) \subset \mathbb{R}^{|C_r|}, \quad \overline{E}_s^{(\kappa)} = \mathcal{E}\big(\overline{\rho}_s^{(\kappa)}, \overline{\Pi}_s^{(\kappa)}\big) \subset \mathbb{R}^{|D_s|},$$

with $r = 1, 2, \ldots, R$, $s = 1, 2, \ldots, S$ and $\kappa = 0, 1, \ldots, T$ are taken at sampling times $t_0 < t_1 < \ldots < t_T$ and the first T predictions of the linear model are given by the ellipsoids

$$\widehat{X}_j^{(\kappa+1)} = \mathcal{E}\big(\widehat{\mu}_j^{(\kappa+1)}, \widehat{\Sigma}_j^{(\kappa+1)}\big) := \xi_{j0} + \left(\bigoplus_{r=1}^{R} \mathcal{A}_{jr}^{GG} \overline{X}_r^{(\kappa)}\right) + \left(\bigoplus_{s=1}^{S} \mathcal{A}_{js}^{EG} \overline{E}_s^{(\kappa)}\right),$$

$$\widehat{E}_i^{(\kappa+1)} = \mathcal{E}\big(\widehat{\rho}_i^{(\kappa+1)}, \widehat{\Pi}_i^{(\kappa+1)}\big) := \zeta_{i0} + \left(\bigoplus_{r=1}^{R} \mathcal{A}_{ir}^{GE} \overline{X}_r^{(\kappa)}\right) + \left(\bigoplus_{s=1}^{S} \mathcal{A}_{is}^{EE} \overline{E}_s^{(\kappa)}\right),$$

with $j = 1, 2, \ldots, R$, $i = 1, 2, \ldots, S$ and $\kappa = 0, 1, \ldots, T-1$.

The main idea of our ellipsoidal regression analysis is to maximize the overlap of the predictions and measurement values (both ellipsoids). For this reason, we introduce the ellipsoids

$$\Delta X_r^{(\kappa)} := \widehat{X}_r^{(\kappa)} \cap \overline{X}_r^{(\kappa)} \quad \text{and} \quad \Delta E_s^{(\kappa)} := \widehat{E}_s^{(\kappa)} \cap \overline{E}_s^{(\kappa)},$$

with $r = 1, 2, \ldots, R$, $s = 1, 2, \ldots, S$ and $\kappa = 1, \ldots, T$, where $\cap$ denotes the fusion of ellipsoids introduced in Subsection 2.4. In addition, the centers of the ellipsoids are adjusted, so that their distance becomes minimized (cf. Figure 1). This leads us to the following regression problem:

$$(R) \quad \text{Maximize} \quad \sum_{\kappa=1}^{T} \left\{ \sum_{r=1}^{R} \left\| \Delta X_r^{(\kappa)} \right\|_* - \left\| \widehat{\mu}_r^{(\kappa)} - \overline{\mu}_r^{(\kappa)} \right\|_2^2 + \sum_{s=1}^{S} \left\| \Delta E_s^{(\kappa)} \right\|_* - \left\| \widehat{\rho}_s^{(\kappa)} - \overline{\rho}_s^{(\kappa)} \right\|_2^2 \right\}.$$

Here, $\|\cdot\|_*$ denotes a measure that reflects the geometrical size of the intersections (fusions) and we assume that $\|\Delta X_r^{(\kappa)}\|_* = 0$, if $\Delta X_r^{(\kappa)} = \emptyset$ and $\|\Delta E_s^{(\kappa)}\|_* = 0$, if $\Delta E_s^{(\kappa)} = \emptyset$. There exist various measures related to the shape of the intersections, e.g., the *volume* (which corresponds to the ellipsoid matrix determinant) or the *sum of squares of semiaxes* (which corresponds to the trace of the configuration matrix).

These examples lead to specific formulations of the regression problem (R) and they depend on the shape matrices of the fusions $\Delta X_r^{(\kappa)}$ and $\Delta E_s^{(\kappa)}$. For further details on geometrical (ellipsoidal) measures and the related regression problems we refer to [25].

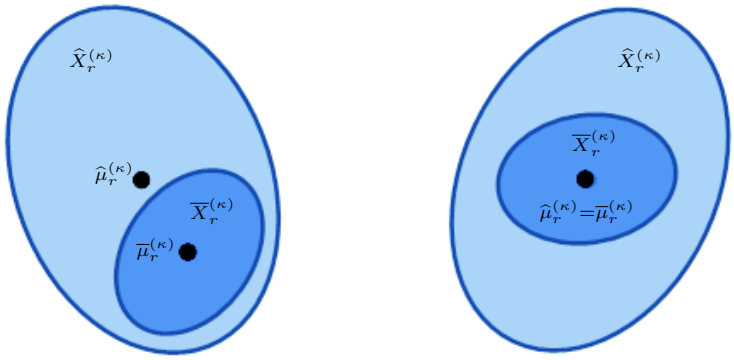

Figure 1: Overlap of ellipsoids: The intersections of the two ellipsoids $\widehat{X}_r^{(\kappa)}$ and $\overline{X}_r^{(\kappa)}$ have the same geometrical size with the same measure of fusions on the left and the right side. On the right side, the centers $\widehat{\mu}_r^{(\kappa)}$ and $\overline{\mu}_r^{(\kappa)}$ are adjusted in order to minimize the difference between the centers of ellipsoids.

For a deeper analysis of the regression problem (R), explicit representations of the fusions $\Delta X_r^{(\kappa)}$ and $\Delta E_s^{(\kappa)}$ are required. The fusion $\Delta X_r^{(\kappa)} = \widehat{X}_r^{(\kappa)} \cap \overline{X}_{C_r}^{(\kappa)}$ is an ellipsoid $\mathcal{E}\big(\Delta\mu_r^{(\kappa)},\, \Delta\Sigma_r^{(\kappa)}\big)$ with center

$$\Delta\mu_r^{(\kappa)} = \big[\mathcal{X}_r^{(\kappa)}\big]^{-1}\big(\lambda\big[\widehat{\Sigma}_r^{(\kappa)}\big]^{-1}\widehat{\mu}_r^{(\kappa)} + (1-\lambda)\big[\overline{\Sigma}_r^{(\kappa)}\big]^{-1}\overline{\mu}_r^{(\kappa)}\big)$$

and shape matrix

$$\Delta\Sigma_r^{(\kappa)} = \xi_r^{(\kappa)}\,\big[\mathcal{X}_r^{(\kappa)}\big]^{-1},$$

where

$$\mathcal{X}_r^{(\kappa)} := \lambda\big[\widehat{\Sigma}_r^{(\kappa)}\big]^{-1} + (1-\lambda)\big[\overline{\Sigma}_r^{(\kappa)}\big]^{-1}$$

and

$$\xi_r^{(\kappa)} := 1 - \lambda(1-\lambda)\big(\overline{\mu}_r^{(\kappa)} - \widehat{\mu}_r^{(\kappa)}\big)^T\big[\overline{\Sigma}_r^{(\kappa)}\big]^{-1}\big[\mathcal{X}_r^{(\kappa)}\big]^{-1}\big[\widehat{\Sigma}_r^{(\kappa)}\big]^{-1}\big(\overline{\mu}_r^{(\kappa)} - \widehat{\mu}_r^{(\kappa)}\big).$$

The parameter λ is the only root in $(0,1)$ of the following polynomial of degree $2|C_r|-1$:

$$\begin{aligned}&\xi_r^{(\kappa)}\big(\det \mathcal{X}_r^{(\kappa)}\big)^2 \operatorname{Tr}\Big(\big[\mathcal{X}_r^{(\kappa)}\big]^{-1}\Big(\big[\widehat{\Sigma}_r^{(\kappa)}\big]^{-1}-\big[\overline{\Sigma}_r^{(\kappa)}\big]^{-1}\Big)\Big)-|C_r|\big(\det \mathcal{X}_r^{(\kappa)}\big)^2\\ &\times\Big(2\big[\Delta\mu_r^{(\kappa)}\big]^T\big[\widehat{\Sigma}_r^{(\kappa)}\big]^{-1}\widehat{\mu}_r^{(\kappa)}-2\big[\Delta\mu_r^{(\kappa)}\big]^T\big[\overline{\Sigma}_r^{(\kappa)}\big]^{-1}\overline{\mu}_r^{(\kappa)}\\ &+\big[\Delta\mu_r^{(\kappa)}\big]^T\Big(\big[\overline{\Sigma}_r^{(\kappa)}\big]^{-1}-\big[\widehat{\Sigma}_r^{(\kappa)}\big]^{-1}\Big)\Delta\mu_r^{(\kappa)}-\big[\widehat{\mu}_r^{(\kappa)}\big]^T\big[\widehat{\Sigma}_r^{(\kappa)}\big]^{-1}\widehat{\mu}_r^{(\kappa)}\\ &+\big[\overline{\mu}_r^{(\kappa)}\big]^T\big[\overline{\Sigma}_r^{(\kappa)}\big]^{-1}\overline{\mu}_r^{(\kappa)}\Big)=0.\end{aligned}$$

Similarly, the fusion $\Delta E_s^{(\kappa)} = \widehat{E}_s^{(\kappa)} \cap \overline{E}_s^{(\kappa)}$ is an ellipsoid $\mathcal{E}\big(\Delta\rho_s^{(\kappa)}, \Delta\Pi_s^{(\kappa)}\big)$ with center

$$\Delta\rho_s^{(\kappa)} = \big[\mathcal{Y}_s^{(\kappa)}\big]^{-1}\big(\lambda\big[\widehat{\Pi}_s^{(\kappa)}\big]^{-1}\widehat{\rho}_s^{(\kappa)}+(1-\lambda)\big[\overline{\Pi}_s^{(\kappa)}\big]^{-1}\overline{\rho}_s^{(\kappa)}\big)$$

and shape matrix

$$\Delta\Pi_s^{(\kappa)} = \eta_s^{(\kappa)}\,\big[\mathcal{Y}_s^{(\kappa)}\big]^{-1},$$

where

$$\mathcal{Y}_s^{(\kappa)} := \lambda\big[\widehat{\Pi}_s^{(\kappa)}\big]^{-1}+(1-\lambda)\big[\overline{\Pi}_s^{(\kappa)}\big]^{-1}$$

and

$$\eta_s^{(\kappa)} := 1-\lambda(1-\lambda)\big(\overline{\rho}_s^{(\kappa)}-\widehat{\rho}_s^{(\kappa)}\big)^T\big[\overline{\Pi}_s^{(\kappa)}\big]^{-1}\big[\mathcal{Y}_s^{(\kappa)}\big]^{-1}\big[\widehat{\Pi}_s^{(\kappa)}\big]^{-1}\big(\overline{\rho}_s^{(\kappa)}-\widehat{\rho}_s^{(\kappa)}\big).$$

The parameter λ is the only root in $(0,1)$ of the following polynomial of degree $2|D_s|-1$:

$$\begin{aligned}&\eta_s^{(\kappa)}\big(\det \mathcal{Y}_s^{(\kappa)}\big)^2 \operatorname{Tr}\Big(\big[\mathcal{Y}_s^{(\kappa)}\big]^{-1}\Big(\big[\widehat{\Pi}_s^{(\kappa)}\big]^{-1}-\big[\overline{\Pi}_s^{(\kappa)}\big]^{-1}\Big)\Big)-|D_s|\big(\det \mathcal{Y}_s^{(\kappa)}\big)^2\\ &\times\Big(2\big[\Delta\rho_s^{(\kappa)}\big]^T\big[\widehat{\Pi}_s^{(\kappa)}\big]^{-1}\widehat{\rho}_s^{(\kappa)}-2\big[\Delta\rho_s^{(\kappa)}\big]^T\big[\overline{\Pi}_s^{(\kappa)}\big]^{-1}\overline{\rho}_s^{(\kappa)}\\ &+\big[\Delta\rho_s^{(\kappa)}\big]^T\Big(\big[\overline{\Pi}_s^{(\kappa)}\big]^{-1}-\big[\widehat{\Pi}_s^{(\kappa)}\big]^{-1}\Big)\Delta\rho_s^{(\kappa)}-\big[\widehat{\rho}_s^{(\kappa)}\big]^T\big[\widehat{\Pi}_s^{(\kappa)}\big]^{-1}\widehat{\rho}_s^{(\kappa)}\\ &+\big[\overline{\rho}_s^{(\kappa)}\big]^T\big[\overline{\Pi}_s^{(\kappa)}\big]^{-1}\overline{\rho}_s^{(\kappa)}\Big)=0.\end{aligned}$$

We now return to the objective function of regression problem (R), where the geometrical size of the fusions $\Delta X_r^{(\kappa)}$ and $\Delta E_s^{(\kappa)}$ is measured. As a measure for the size of a p-dimensional ellipsoid $\mathcal{E}(0,Q)$ nonnegative-valued criteria functions $\psi(\mathcal{E}(0,Q))$ defined on the set of all nondegenerate ellipsoids can be applied. These functions are monotonous by increasing with respect to inclusion, i.e., $\psi(\mathcal{E}_1) \le \psi(\mathcal{E}_2)$ if $\mathcal{E}_1 \subseteq \mathcal{E}_2$. Such measures are, e.g.,

(a) *the trace of* Q,

$$\psi_T(\mathcal{E}(0,Q)) := \operatorname{Tr} Q = \lambda_1+\ldots+\lambda_p,$$

where λ_i are the eigenvalues of Q (i.e., $\operatorname{Tr} Q$ is equal to the sum of the squares of the semiaxes),

(b) *the determinant of* Q,

$$\psi_{Det}(\mathcal{E}(0,Q)) := \det Q = \lambda_1 \cdot \ldots \cdot \lambda_p,$$

which is equal to the product of eigenvalues and proportional to the volume

$$\operatorname{vol} \mathcal{E}(0,Q) = \pi^{\frac{p}{2}} \, (\det Q)^{\frac{1}{2}} \big(\Gamma\big(\tfrac{p}{2}+1\big)\big)^{-1}$$

of the ellipsoid, where Γ stands for the Gamma-function.

For further details on criteria functions we refer to [26], p. 101. The measures stated above lead to different representations of the regression problem (R) and we study them now in more detail.

4.2 The Trace Criterion

By measuring the size of the ellipsoids $\Delta X_r^{(\kappa)}$ and $\Delta E_s^{(\kappa)}$ in terms of their (squared) lengths of semiaxes and, thus, the traces of the shape matrices, we can state the following regression problem:

$$(R_{Tr}) \qquad \text{Maximize} \qquad \sum_{\kappa=1}^{T} \Bigg\{ \sum_{r=1}^{R} \Bigg[\operatorname{Tr}\big(\Delta\Sigma_r^{(\kappa)}\big) - \sum_{j=1}^{|C_r|} \big(\widehat{\mu}_{r,j}^{(\kappa)} - \overline{\mu}_{r,j}^{(\kappa)}\big)^2 \Bigg] + \sum_{s=1}^{S} \Bigg[\operatorname{Tr}\big(\Delta\Pi_s^{(\kappa)}\big) - \sum_{i=1}^{|D_s|} \big(\widehat{\rho}_{s,i}^{(\kappa)} - \overline{\rho}_{s,i}^{(\kappa)}\big)^2 \Bigg] \Bigg\}.$$

As the trace of the shape matrix of an ellipsoid is equal to the sum of the squares of the semiaxes, the regression problem takes the form

$$(R'_{Tr}) \qquad \text{Maximize} \qquad \sum_{\kappa=1}^{T} \Bigg\{ \sum_{r=1}^{R} \Bigg[\sum_{j=1}^{|C_r|} \lambda_{r,j}^{(\kappa)} - \big(\widehat{\mu}_{r,j}^{(\kappa)} - \overline{\mu}_{r,j}^{(\kappa)}\big)^2 \Bigg] + \sum_{s=1}^{S} \Bigg[\sum_{i=1}^{|D_s|} \Lambda_{s,i}^{(\kappa)} - \big(\widehat{\rho}_{s,i}^{(\kappa)} - \overline{\rho}_{s,i}^{(\kappa)}\big)^2 \Bigg] \Bigg\},$$

where $\lambda_{r,j}^{(\kappa)}$ and $\Lambda_{s,i}^{(\kappa)}$ are the eigenvalues of $\Delta\Sigma_r^{(\kappa)}$ and $\Delta\Pi_s^{(\kappa)}$, respectively.

4.3 The Determinant Criterion

Another objective function of the regression model can be build with the determinants of the configuration matrices of the ellipsoids $\Delta X_r^{(\kappa)}$ and $\Delta E_s^{(\kappa)}$:

$$(R_{Det}) \qquad \text{Maximize} \qquad \sum_{\kappa=1}^{T} \Bigg\{ \sum_{r=1}^{R} \Bigg[\det\big(\Delta\Sigma_r^{(\kappa)}\big) - \sum_{j=1}^{|C_r|} \big(\widehat{\mu}_{r,j}^{(\kappa)} - \overline{\mu}_{r,j}^{(\kappa)}\big)^2 \Bigg] + \sum_{s=1}^{S} \Bigg[\det\big(\Delta\Pi_s^{(\kappa)}\big) - \sum_{i=1}^{|D_s|} \big(\widehat{\rho}_{s,i}^{(\kappa)} - \overline{\rho}_{s,i}^{(\kappa)}\big)^2 \Bigg] \Bigg\}.$$

Equivalent formulations of (R_{Det}) can be given in terms of the eigenvalues of the configuration matrices

$$(R'_{Det}) \qquad \text{Maximize} \qquad \sum_{\kappa=1}^{T}\Bigg\{\sum_{r=1}^{R}\Bigg[\Bigg(\prod_{j=1}^{|C_r|}\lambda_{r,j}^{(\kappa)}\Bigg)-\sum_{j=1}^{|C_r|}\big(\widehat{\mu}_{r,j}^{(\kappa)}-\overline{\mu}_{r,j}^{(\kappa)}\big)^2\Bigg] + \sum_{s=1}^{S}\Bigg[\Bigg(\prod_{i=1}^{|D_s|}\Lambda_{s,i}^{(\kappa)}\Bigg)-\sum_{i=1}^{|D_s|}\big(\widehat{\rho}_{s,i}^{(\kappa)}-\overline{\rho}_{s,i}^{(\kappa)}\big)^2\Bigg]\Bigg\},$$

and the volumes of the ellipsoids $\Delta X_r^{(\kappa)}$ and $\Delta E_s^{(\kappa)}$

$$(R''_{Det}) \qquad \text{Maximize} \qquad \sum_{\kappa=1}^{T}\Bigg\{\sum_{r=1}^{R}\Bigg[\Bigg(\pi^{\frac{2}{|C_r|}}\,\Gamma\Bigg(\frac{|C_r|}{2}+1\Bigg)\,\mathrm{vol}\left(\Delta X_r^{(\kappa)}\right)\Bigg)^2 - \sum_{j=1}^{|C_r|}\big(\widehat{\mu}_{r,j}^{(\kappa)}-\overline{\mu}_{r,j}^{(\kappa)}\big)^2\Bigg] + \sum_{s=1}^{S}\Bigg[\Bigg(\pi^{\frac{2}{|D_s|}}\,\Gamma\Bigg(\frac{|D_s|}{2}+1\Bigg)\,\mathrm{vol}\left(\Delta E_s^{(\kappa)}\right)\Bigg)^2 - \sum_{i=1}^{|D_s|}\big(\widehat{\rho}_{s,i}^{(\kappa)}-\overline{\rho}_{s,i}^{(\kappa)}\big)^2\Bigg]\Bigg\}.$$

5 Optimization Methods

We now turn to optimization methods for the regression models of the previous subsections. These models depend on the shape matrices $\Sigma_r^{(\kappa)}$ and $\Pi_s^{(\kappa)}$ and the distance of the centers $\mu_r^{(\kappa)}$ and $\rho_s^{(\kappa)}$ of the fusions $\Delta X_r^{(\kappa)}$ and $\Delta E_s^{(\kappa)}$. In this situation, methods from semidefinite programming [14] can be applied, because the objective functions of these volume-related programming problems depend on, e.g., the determinant or eigenvalues of symmetric positive semidefinite matrices. In addition, the squared Euclidean norm is conic representable ([10], p. 88) and, thus, semidefinite representable ([10], p. 132). However, some of the regression models introduced above have to be slightly modified in order to obtain positive semidefinite representable objective functions [11]. For example, the objective function of the regression model (R_{Det}) depends directly on the determinant of shape matrices, but the determinant $\det(M)$ considered as a function of symmetric positive semidefinite $n \times n$-matrices M (short: $M \succcurlyeq 0$) is neither a convex nor a concave function of M (if $n \geq 2$). However, if p is a rational number with $0 \leq p \leq \frac{1}{n}$, then

$$f(M) = \begin{cases} -\det{}^p(M) & , M \succcurlyeq 0, \\ \infty & , \text{otherwise}, \end{cases}$$

is positive semidefinite representable ([11], p. 81). For this reason, we introduce the regression model

$$(\widetilde{R}_{Det}) \qquad \text{Maximize} \qquad \sum_{\kappa=1}^{T}\Bigg\{-\sum_{r=1}^{R}\left[\det{}^{p}\left(\Delta\Sigma_r^{(\kappa)}\right)+\sum_{j=1}^{|C_r|}\left(\widehat{\mu}_{r,j}^{(\kappa)}-\overline{\mu}_{r,j}^{(\kappa)}\right)^2\right] - \sum_{s=1}^{S}\left[\det{}^{q}\left(\Delta\Pi_s^{(\kappa)}\right)+\sum_{i=1}^{|D_s|}\left(\widehat{\rho}_{s,i}^{(\kappa)}-\overline{\rho}_{s,i}^{(\kappa)}\right)^2\right]\Bigg\},$$

where the rational numbers p, q fulfill the conditions $0 \leq p \leq \frac{1}{|C_r|}$ and $0 \leq q \leq \frac{1}{|D_s|}$. Since $\det(M) = \prod_{i=1}^{n}\lambda_i(M)$, where $\lambda_i(M)$ are the eigenvalues of M, we can replace (R'_{Det}) by

$$(\widetilde{R}'_{Det}) \qquad \text{Maximize} \qquad \sum_{\kappa=1}^{T}\Bigg\{-\sum_{r=1}^{R}\left[\left(\prod_{j=1}^{|C_r|}\lambda_{r,j}^{(\kappa)}\right)^{p}+\sum_{j=1}^{|C_r|}\left(\widehat{\mu}_{r,j}^{(\kappa)}-\overline{\mu}_{r,j}^{(\kappa)}\right)^2\right] - \sum_{s=1}^{S}\left[\left(\prod_{i=1}^{|D_s|}\Lambda_{s,i}^{(\kappa)}\right)^{q}+\sum_{i=1}^{|D_s|}\left(\widehat{\rho}_{s,i}^{(\kappa)}-\overline{\rho}_{s,i}^{(\kappa)}\right)^2\right]\Bigg\},$$

and instead of (R''_{Det}) we suggest

$$(\widetilde{R}''_{Det}) \qquad \text{Maximize} \sum_{\kappa=1}^{T}\Bigg\{\sum_{r=1}^{R}\left[\left(\pi^{\frac{2}{|C_r|}}\,\Gamma\left(\frac{|C_r|}{2}+1\right)\text{vol}\left(\Delta X_r^{(\kappa)}\right)\right)^{2p} - \sum_{j=1}^{|C_r|}\left(\widehat{\mu}_{r,j}^{(\kappa)}-\overline{\mu}_{r,j}^{(\kappa)}\right)^2\right] + \sum_{s=1}^{S}\left[\left(\pi^{\frac{2}{|D_s|}}\,\Gamma\left(\frac{|D_s|}{2}+1\right)\text{vol}\left(\Delta E_s^{(\kappa)}\right)\right)^{2q} - \sum_{i=1}^{|D_s|}\left(\widehat{\rho}_{s,i}^{(\kappa)}-\overline{\rho}_{s,i}^{(\kappa)}\right)^2\right]\Bigg\}.$$

We note that in case of positive definite shape matrices $\Delta\Sigma_r^{(\kappa)}$ and $\Delta\Pi_s^{(\kappa)}$ negative powers of the determinant can be used. If p is a positive rational, the function

$$f(M) = \begin{cases}\det{}^{-p}(M) & , M \succ 0,\\ \infty & , \text{otherwise,}\end{cases}$$

of the symmetric $n \times n$-matrix M is positive semidefinite representable ([11], p. 83). Here, $M \succ 0$ means that M is positive semidefinite. Now, with two positive rationals p, q we obtain a further regression model

$$(R'''_{Det}) \qquad \text{Maximize} \qquad \sum_{\kappa=1}^{T}\Bigg\{\sum_{r=1}^{R}\left[\det{}^{-p}\left(\Delta\Sigma_r^{(\kappa)}\right)+\sum_{j=1}^{|C_r|}\left(\widehat{\mu}_{r,j}^{(\kappa)}-\overline{\mu}_{r,j}^{(\kappa)}\right)^2\right] + \sum_{s=1}^{S}\left[\det{}^{-q}\left(\Delta\Pi_s^{(\kappa)}\right)+\sum_{i=1}^{|D_s|}\left(\widehat{\rho}_{s,i}^{(\kappa)}-\overline{\rho}_{s,i}^{(\kappa)}\right)^2\right]\Bigg\}.$$

In the regression model (R'_{Tr}), sums of all eigenvalues of the configuration matrices $\Delta\Sigma_r^{(\kappa)}$ and $\Delta\Pi_s^{(\kappa)}$ are considered, which can also be regarded as positive semidefinite representable functions ([11], p. 80). In general, *interior point methods* can applied which have a moderate complexity [29, 30, 31, 33]. Alternatively, for regression problems that consider sums of eigenvalues in the objective function, asscociated *bilevel problems* can be introduced which could be solved by *gradient methods*. In fact, in [32] structural frontiers of conic programming are discussed with other optimization methods compared, and future applications in machine learning and data mining prepared. However, we would like to underline that in the areas regression and classification of statistical learning (cf. e.g., [21, 37]), our optimization based methods provided and further promise very good and competitive results [15, 22, 41, 52, 58].

REMARK *(Future Research)*. We represented genetic and environmental states with the help of ellipoids. This model addressed both data and predicted variables, and it served as an input into design or system matrices. Since we finally arrived at conic optimization problems, we may also state the corresponding dual problems. Here we will in future investigate, and also invite the reader to this, which roles the ellipsoids then play in terms of the outputs of the interactions. This could establish a *Dual Ellipsoidal OR*.

Moreover, in [49] we expressed that we can *navigate* between entries and entire block matrices, e.g., *quadrants*, of the interaction and design matrices, a control or regulatory measurements, e.g., in the context of Kyoto Protocol [35, 50]. A closer understanding of this navigating in the presence of ellipsoidal uncertainty in the state variables and, what is more, in the blockwise (clustered) entries of our system matrices, is subject of future research, too.

6 Conclusion

In this paper, we analyzed a time-discrete regression model for gene-environment regulatory networks under ellipsoidal uncertainty. The ellipsoidal approach - complemented by powerful optimization methods of semidefinite programming and the efficiency of interior point methods - offers a new perspective for the analysis of regulatory systems in systems biology and life sciences, and it could even establish an *Ellipsoidal OR*. The representation of the dynamic states of the underlying system in terms of ellipsoids was motivated by our and our colleagues' studies on gene-environment networks and eco-finance networks, where errors and uncertainty are modeled by intervals [58, 59, 60]. Here, we further extended this approach by replacing the intervals by more general uncertainty sets. Ellipsoids are more flexible than intervals and parallelpipes what also refers to stochastic dependencies between the various genes and environmental factors. Moreover, ellipsoids are directly related to covariance matrices and they provide good approximations of convex sets. In particular, models based on Gaussian random noise refer to the ellipsoidal approach. However, Gaussian random distributions are often used as simplifications,

and in many applications non-Gaussian probability distributions have to be applied. In future works, we will further extend our models based on ellipsoidal calculus and we will turn to a more set-theoretic representation of uncertainty based on semi-algebraic sets [9, 12]. We will combine this new perception with refined optimization methods, and by this we will offer a further avenue for the analysis of regulatory systems, particularly with regard to applications and real-world data. Furthermore, we refined collaborative game theory under interval uncertainty [1, 2, 3, 4] by our ellipsoidal calculus [5], herewith including correlations, as being expressed by subcoalitions found by clustering. In terms if, e.g., *joint implementation*, collaborative game theory serves for an new interpretation of the TEM model, which is some dynamical model related with the Kyoto Protocol and a part of our gene-environment networks [35, 48, 49]. With this study, we aim scientific advances and the improvement of living conditions on earth.

References

[1] S.Z. Alparslan Gök, Cooperative Interval Games, PhD Thesis, Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey, 2009.

[2] S.Z. Alparslan Gök, R. Branzei, and S. Tijs, Convex interval games, Preprint at IAM, Middle East Technical University, Ankara, Turkey, and Center for Economic Research, Tilburg University, The Netherlands, 2008.

[3] S.Z. Alparslan Gök, R. Branzei, and S. Tijs, Airport interval games and their Shapley value, to appear in *Operations Research and Decisions*, 1, 2009.

[4] S.Z. Alparslan Gök, S. Miquel, and S. Tijs, Cooperation under interval uncertainty, *Math. Methods Oper. Res.*, **69**, (2009), pp. 99–109.

[5] S.Z. Alparslan Gök, and G.-W. Weber, Cooperative games under ellipsoidal uncertainty, to appear in the *Proceedings of PCO 2010, 3rd Global Conference on Power Control and Optimization*, February 2-4, 2010, Gold Coast, Queensland, Australia.

[6] A. Aster, B. Borchers, and C. Thurber, Parameter estimation and inverse problems, Academic Press, 2004.

[7] A.M. Bagirov, and J. Yearwood, A new nonsmooth optimization algorithm for minimum sum-of-squares clustering problems, *European J. Oper. Res.*, **170**, 2 (2006), pp. 578–596.

[8] Z. Barzily, Z.V. Volkovich, B. Akteke-Öztürk, and G.-W. Weber, Cluster stability using minimal spanning trees, *ISI Proceedings of 20th Mini-EURO Conference Continuous Optimization and Knowledge-Based Technologies*, Neringa, Lithuania, May 20-23, (2008), pp. 248–252.

[9] R. Benedetti, Real algebraic and semi-algebraic sets, Hermann, Ed. des Sciences et des Arts, Paris, 1990.

[10] A. Ben-Tal, and A. Nemirovski, Convex Optimization in Engineering - Modeling, Analyis, Algorithms. Faculty of Industrial Engineering and Management Technion-City, Haifa 32000, Israel, available at www.isa.ewi.tudelft.nl/ roos/courses/WI4218/tud00r.pdf.

[11] A. Ben-Tal, Conic and robust optimization, Lecture notes (2002) Available at http://iew3.technion.ac.il/Home/Users/morbt.phtml.

[12] J. Bochnak, M. Coste, and M.-F. Roy, Real algebraic geometry, Springer 1998.

[13] E. Borenstein, and M.W. Feldman, Topological signatures of species interactions in metabolic networks, *J. Comput. Biol.*, **16**, 2, (2009), pp. 191–200, DOI: 10.1089/cmb.2008.06TT.

[14] S. Boyd, and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[15] E. Büyükbebeci, Comparison of MARS, CMARS and CART in Predicting Default Probabilities for Emerging Markets, MSc. Term Project Report/Thesis in Financial Mathematics, at IAM, METU, Ankara, August 2009.

[16] P. Durieu, É. Walter, and B. Polyak, Multi-input multi-output ellipsoidal state bounding, *J. Optim. Theory Appl.*, **111**, 2, (2001), pp. 273–303.

[17] J. Gebert, M. Lätsch, E.M.P. Quek, and G.-W. Weber, Analyzing and optimizing genetic network structure via path-finding, *Journal of Computational Technologies*, **9**, 3, (2004), pp. 3–12.

[18] A. Gökmen, S. Kayalgil, G.-W. Weber, I. Gökmen, M. Ecevit, A. Sürmeli, T. Bali, Y. Ecevit, H. Gökmen, and D.J. DeTombe, Balaban Valley Project: Improving the Quality of Life in Rural Area in Turkey, *International Scientific Journal of Methods and Models of Complexity*, **7**, 1, (2004).

[19] J.R. Harris, W. Nystad, and P. Magnus, Using genes and environments to define asthma and related phenotypes: applications to multivariate data, *Clinical and Experimental Allergy*, **28**, 1, (1998), pp. 43–45.

[20] T. Hastie, and R. Tibshirani, Discriminant adaptive nearest neighbor classification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, 6, (1996), pp. 607–616.

[21] T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning, Springer (2001).

[22] A. Işcanoğlu, G.-W. Weber, and P. Taylan, Predicting default probabilities with generalized additive models for emerging markets, invited lecture, Graduate Summer School on New Advances in Statistics, METU (2007).

[23] E. Kropat, S. Pickl, A. Rössler, and G.-W. Weber, On theoretical and practical relations between discrete optimization and nonlinear optimization, in special issue Colloquy Optimization Structure and Stability of Dynamical Systems (at the occasion of the colloquy with the same name, Cologne, October 2000) of *Journal of Computational Technologies*, **7**, 2002), pp. 27–62.

[24] E. Kropat, G.-W. Weber, and B. Akteke-Öztürk, Eco-finance networks under uncertainty, in J. Herskovits, A. Canelas, H. Cortes, and M. Aroztegui, eds., *Proceedings of the International Conference on Engineering Optimization* (ISBN 978857650156-5, CD), EngOpt 2008, Rio de Janeiro, Brazil (2008).

[25] E. Kropat, G.-W. Weber, and C.S. Pedamallu, Regulatory networks under ellipsoidal uncertainty - optimization theory and dynamical systems, Preprint at IAM, METU, Ankara, Turkey, 2009. Submitted to *SIAM Journal on Optimization*.

[26] A.B. Kurzhanski, and I. Vályi, Ellipsoidal Calculus for Estimation and Control, Birkhäuser, 1997.

[27] A.A. Kurzhanski, and P. Varaiya, Ellipsoidal Toolbox Manual, EECS Department, University of California, Berkeley, 2008.

[28] W.B. Mattes, S.D. Pettit, S.A. Sansone, P.R. Bushel, and M.D. Waters, Database development in toxicogenomics: issues and efforts, *Environmental Health Perspectives*, **112**, (4), (2004), pp. 495-505.

[29] A. Nemirovski, Five Lectures on Modern Convex Optimization, C.O.R.E. Summer School on Modern Convex Optimization (2002), Available at http://iew3.technion.ac.il/Labs/Opt/opt/LN/Final.pdf.

[30] A. Nemirovski, Lectures on Modern Convex Optimization, Israel Institute of Technology (2002), available at http://iew3.technion.ac.il/Labs/Opt/opt/LN/Final.pdf.

[31] A. Nemirovski, Interior Point Polynomial Time Algorithms in Convex Programming, Lecture Notes (2004), available at https://itweb.isye.gatech.edu.

[32] A. Nemirovski, Modern Convex Optimization, Lecture at PASCAL Workshop, Thurnau, Germany, March 16-18 (2005).

[33] Y.E. Nesterov, and A.S. Nemirovskii, Interior Point Polynomial Algorithms in Convex Programming, SIAM, 1994.

[34] M. Partner, N. Kashtan, and U. Alon, Environmental variability and modularity of bacterial metabolic network, *BMC Evolutionary Biology*, **7**, (2007), 169doi:10.1186/1471-2148-7-169.

[35] S. Pickl, Der τ-value als Kontrollparameter - Modellierung und Analyse eines Joint-Implementation Programmes mithilfe der dynamischen kooperativen Spieltheorie und der diskreten Optimierung, Thesis, Darmstadt University of Technology, Department of Mathematics, 1998.

[36] L. Ros, A. Sabater, and F. Thomas, An ellipsoidal calculus based on propagation and fusion, *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, **32**, 4 (2002), pp. 430–442.

[37] Y. She, Sparse regression with exact clustering, PhD Thesis, Department of Statistics, Stanford University, USA, 2008.

[38] M. Taştan, Analysis and prediction of gene expression patterns by dynamical systems, and by a combinatorial algorithm, MSc Thesis, Institute of Applied Mathematics, METU, Turkey, 2005.

[39] M. Taştan, T. Ergenç, S.W. Pickl, and G.-W. Weber, Stability analysis of gene expression patterns by dynamical systems and a combinatorial algorithm, in *HIBIT – Proceedings of International Symposium on Health Informatics and Bioinformatics*, Turkey '05, Antalya, Turkey, 2005, pp. 67–75.

[40] M. Taştan, S.W. Pickl, and G.-W. Weber, Mathematical modeling and stability analysis of gene-expression patterns in an extended space and with Runge-Kutta discretization, in *Proceedings of Operations Research 2005*, Bremen, September 2005, Springer, pp. 443–450.

[41] P. Taylan, G.-W. Weber, and A. Beck, New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and techology, in the special issue in honour of Prof. Dr. Alexander Rubinov, Burachik, B., Yang, X. (guest eds.) *Optimization*, **56**, 5-6, (2007), pp. 1–24.

[42] Ö. Uğur, S.W. Pickl, G.-W. Weber, and R. Wünschiers, Operational research meets biology: An algorithmic approach to analyze genetic networks and biological energy production, Preprint no. 50, Institute of Applied Mathematics, METU, 2006, submitted for the special issue of *Optimization* at the occasion of the 5th Ballarat Workshop on Global and Non-Smooth Optimization: Theory, Methods, and Applications (2006).

[43] Ö. Uğur, and G.-W. Weber, Optimization and dynamics of gene-environment networks with intervals, in the special issue at the occasion of the 5th Ballarat Workshop on Global and Non-Smooth Optimization: Theory, Methods and Applications, November 28-30, 2006, of *J. Ind. Manag. Optim.*, **3**, 2, (2007), pp. 357–379.

[44] A.Y. Vazhentsev, On internal ellipsoidal approximations for problems of control synthesis with bounded coordinates, *J. Comput. System Sci. Int.*, **39**, 3, (2000), 399.

[45] F.G. Vázques, J.-J. Rückmann, O. Stein, and G. Still, Generalized semi-infinite programming: A tutorial, *Journal of computational and applied mathematics*, **217**, 2, (2008), pp. 394–419.

[46] G.-W. Weber, Charakterisierung struktureller Stabilität in der nichtlinearen Optimierung, in Aachener Beiträge zur Mathematik 5, H.H. Bock, H.T. Jongen, and W. Plesken, eds., Augustinus publishing house (now: Mainz publishing house) Aachen (1992).

[47] G.-W. Weber, Generalized semi-infinite optimization and related topics, Heldermann Publishing House, Research and Exposition in Mathematics 29, Lemgo, K.H. Hofmannn and R. Wille, R., eds. (2003).

[48] G.-W. Weber, S.Z. Alparslan-Gök, and N. Dikmen, Environmental and life sciences: gene-environment networks - optimization, games and control - a survey on recent achievements, Preprint at Institute of Applied Mathematics, invited paper, to appear in the special issue of *Journal of Organisational Transformation and Social Change*, **5**, 3, (2008), pp. 197–233 .

[49] G.-W. Weber, S.Z. Alparslan-Gök, and B. Söyler, A new mathematical approach in environmental and life sciences: gene-environment networks and their dynamics, *Environmental Modeling & Assessment*, **14**, 2, April 2009, pp. 267-288.

[50] G.-W. Weber, E. Kropat, B. Akteke-Öztürk, and Z.-K. Görgülü, A survey on OR and mathematical methods applied on gene-environment networks, Special Issue on "Innovative Approaches for Decision Analysis in Energy, Health, and Life Sciences" of *Central European Journal of Operations Research (CEJOR)* at the occasion of EURO XXII 2007 (Prague, Czech Republic, July 8-11, 2007), **17**, 3, (2009), pp. 315–341.

[51] G.-W. Weber, E. Kropat, A. Tezel, and S. Belen, Optimization applied on regulatory and eco-finance networks - survey and new developments, to appear in *Pac. J. Optim.*, **6**, 3, Special Issue in memory of Professor Alexander Rubinov, M. Fukushima, M., et al., guest eds. (2010).

[52] G.-W. Weber, I. Batmaz, G. Köksal, P. Taylan and F. Yerlikaya-Özkurt, CMARS: A new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimisation, Preprint at IAM, METU, Ankara, submitted to *Advances in Computational and Applied Mathematics*.

[53] G.-W. Weber, S. Özögür-Akyüz, and E. Kropat, A review on data mining and continuous optimization applications in computational biology and medicine, *Embryo Today, Birth Defects Research (Part C)*, **87**, (2009), pp. 165–181.

[54] G.-W. Weber, P. Taylan, S.-Z. Alparslan-Gök, S. Özöğür, and B. Akteke-Öztürk, Optimization of gene-environment networks in the presence of errors and uncertainty with Chebychev approximation, *TOP*, the Operational

Research journal of SEIO (Spanish Statistics and Operations Research Society) 16, 2 (2008), pp. 284-318.

[55] G.-W. Weber, and A. Tezel, On generalized semi-infinite optimization of genetic networks, *TOP*, **15**, 1, (2007), pp. 65–77.

[56] G.-W. Weber, A. Tezel, P. Taylan, A. Soyler, and M. Çetin, Mathematical contributions to dynamics and optimization of gene-environment networks, in Special Issue: In Celebration of Prof. Dr. Dr. Hubertus Th. Jongen's 60th Birthday, D. Pallaschke, O. Stein, guest eds., of *Optimization*, **57**, 2, (April 2008), pp. 353-377.

[57] G.-W. Weber, Ö. Uğur, P. Taylan, and A. Tezel, On optimization, dynamics and uncertainty: a tutorial for gene-environment networks, in the Special Issue *Networks in Computational Biology* of *Discrete Appl. Math.*, **157**, 10, May 2009, pp. 2494–2513.

[58] F. Yerlikaya, A new contribution to nonlinear robust regression and classification with MARS and its applications to data mining for quality control in manufacturing, Thesis, Middle East Technical University, Ankara, Turkey, 2008.

[59] F.B. Yılmaz, A mathematical modeling and approximation of gene expression patterns by linear and quadratic regulatory relations and analysis of gene networks, MSc Thesis, Institute of Applied Mathematics, METU, Ankara, Turkey, 2004.

[60] F.B. Yılmaz, H. Öktem, and G.-W. Weber, Mathematical modeling and approximation of gene expression patterns and gene networks, in *Operations Research Proceedings*, F. Fleuren, D. den Hertog, and P. Kort, eds., 2005, pp. 280–287.