Modeling, Inference and Optimization of Regulatory Networks Based on Time Series Data

Gerhard-Wilhelm Weber[a,1], Ozlem Defterli[b], Sırma Zeynep Alparslan Gök[c], Erik Kropat[d,2,*]

[a] *Middle East Technical University, Institute of Applied Mathematics, Departments of Financial Mathematics, Scientific Computing and Actuarial Sciences, 06531 Ankara, Turkey*
[b] *Çankaya University, Faculty of Arts and Science, Department of Mathematics and Computer Science, Balgat, 06530 Ankara, Turkey*
[c] *Süleyman Demirel University, Faculty of Arts and Sciences, Department of Mathematics, 32 260 Isparta, Turkey*
[d] *Universität der Bundeswehr München, Department of Informatics, Institute for Theoretical Computer Science, Mathematics and Operations Research Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany*

Abstract

In this survey paper, we present advances achieved during the last years in the development and use of OR, in particular, optimization methods in the new *gene-environment* and *eco-finance networks*, based on usually finite data series, with an emphasis on *uncertainty* in them and in the interactions of the model items. Indeed, our networks represent models in the form of *time-continuous* and *time-discrete dynamics*, whose unknown parameters we estimate under constraints on complexity and regularization by various kinds of optimization techniques, ranging from linear, mixed-integer, spline, semi-infinite and robust optimization to conic, e.g., semidefinite programming. We present different kinds of uncertainties and a new time-discretization technique, address aspects of data preprocessing and of stability, related

*Corresponding author
Email addresses: gweber@metu.edu.tr (Gerhard-Wilhelm Weber), defterli@cankaya.edu.tr (Ozlem Defterli), zeynepalparslan@yahoo.com (Sırma Zeynep Alparslan Gök), erik.kropat@unibw.de (Erik Kropat)
[1]Honorary positions: University of Aveiro, Portugal; University of Siegen, Germany; Universiti Teknologi Malaysia, Skudai, Malaysia
[2]Fax: +49 (89) 6004 3561, Phone: +49 (89) 6004 2350

aspects from game theory and financial mathematics, we work out structural frontiers and discuss chances for future research and OR application in our real world.

Keywords: nonlinear programming, uncertainty modeling, computational biology, environment, games, data mining
2010 MSC: 90C30, 92C42, 91B76

---

1. Introduction

Like very few other disciplines only, the modeling and prediction of genetic data is requesting mathematics nowadays to deeply understand its foundations. This need is even forced by the rapid changes in a world of globalization. Such a study has to include aspects of stability and tractability; the still existing limitations of modern technology in terms of measurement errors and uncertainty have to be taken into account. In Weber et al. (2008b), the important role played by the environment is introduced into the biological context and connected with employing the theories of optimization and dynamical systems.

According to the classical understanding of stability from science, technology and medicine (cf. Guckenheimer and Holmes (1997)), which is mostly positive in terms of some local order, a coming to a rest (recovering) or as the robust behaviour of a small or large bio-system against attacks (e.g., by epidemics), there is also the negative meaning of inflexibility. In fact, in medicine, an organism, living being or bio-system which is unable to adapt to a changing environment is seriously threatened by infections, radiation and other kinds of attacks. What is more, thirdly, a stability analysis on our various expression levels can also lead to an acceptance or a rejection (and following improvement) for the model. If a component of the model behaves unbounded, then this is in contradiction with the natural-technical limitation of the gene- (environment-) expression levels which lie in trusted regions of bounded intervals (see Weber et al. (2008b)).

Further, Weber et al. (2008a) survey and closer explain recent advances in understanding the mathematical foundations and interdisciplinary implications of the newly introduced *gene-environment networks*. They integrate the important theme of environmental protection by joint international projects into the our context of networks and their dynamics. As an example of environmental protection, they study CO_2 emissions, their implications for global

warming by greenhouse effect, the reduction of both and the joint implementation requested for this purpose by Kyoto protocol (cf. Kyoto (1997a,b,c)). Moreover, in the survey of Weber et al. (2009a), the important theme carbon dioxide emission reduction is integrated into the context of the networks and their dynamics.

In Defterli et al. (2009), time-discrete gene-environment and eco-finance networks are investigated as a subclass of target-environment regulatory systems. A different time discretization scheme is derived for a set of dynamical models representing these networks. Rarefication of such kind of regulatory network is studied together with the corresponding mixed-integer regression problem, and a further relaxation could be obtained by means of continuous optimization. Therefore, the paper of Defterli et al. (2009) extends the existing mathematical toolbox by introducing and applying a new method of time-discretization into the study and discussing its potential of improvement. In addition, a method of problem regularization, called rarefication, is recalled and applied and integrated into this research. With this, it is aimed to reach an improved modeling and prediction of our networks, and for a better service in the real-world application areas like health care, environmental protection and sustainable development, economy and society, financial sector, and the living conditions of the people (Pickl and Weber (2001, 2002); Gökmen et al. (2004); Taylan et al. (2007); Weber et al. (2008a, 2009a)).

In Operational Research, *eco-finance networks* under *uncertainty* became introduced in Weber and Uğur (2007). We can find the closely related *eco-finance networks* presented in Kropat et al. (2008) and all these networks are regarded and analyzed as *regulatory networks*. Compared with the more classical genetic networks (which now appear as a subclass), the new nodes are environmental items such as, e.g., poison in soil, groundwater, air or food, emissions, radiation, but also the state of development of the financial markets, the welfare and living conditions, temperature (concerning, e.g., global warming) and, finally, education and campaigns for a healthy lifestyle. *Environmental items* themselves and how they exercise effects - often in mutually catalyzing or multiplicative ways, are becoming very important, the more so as we are living in a time of globalization, of rapid information exchange, of mobility and multicausalities in all kinds of bio- and social systems, communities and societies as well (see Alparslan Gök and Weber (2009)).

Clustering, but also *classification*, provide an insight into the structure of the data and allow to identify groups of model items which are considered to jointly act on other clusters of items of the regulatory model. The uncertain

states of these clusters are represented by ellipsoids, and ellipsoidal calculus is applied to model the dynamics of the system (Kropat et al. (2009a,b)).

As we explained, the ellipsoid-valued model groups the usually big set of items, in the game theoretical context we say: players or actors, into clusters. Let us think of the family of nations which signed the Treaty of Kyoto (Kyoto (1997a,b,c)) or the participants in a worldwide auction (cf. Córdoba Bueno (2006)), etc..

Regulatory networks are often characterized by the presence of a large number of variables and parameters resulting in a complexity which is beyond man's everyday perception. The development of high throughput technologies has resulted in a generation of massive quantities of data. This technological progress has been accompanied by the development of new mathematical methods for the analysis of such highly interconnected systems that allows to gain deeper insights in the dynamic behaviour of complex regulatory systems in biology, finance and engineering sciences. In Kropat et al. (2009b) a special class of so-called *TE-regulatory systems* (Target-Environment regulatory systems) is addressed.

There is a rich list of roles and performances delivered which are associated and assigned to ellipsoids. They include: *(i)* encompassing of objects, *(ii)* inner or outer approximation of shapes and bodies, of discrete or continuous kinds of sets, *(iii)* support for classification of objects and discrimination of different objects, *(iv)* defining critical points or contours which mark tails of higher dimensional and connected versions of tails that describe neighbourhoods of infinity, usually with small values of small probabilities assigned, *(v)* set-valued generalizations of numbers, and generalizations of balls with a reduced wealth of symmetries but still highly symmetrical, *(vi)* geometrical representation of linear mappings which execute certain expansions and contractions (herewith, deformation; e.g., applied to a ball) and rotations, with respect to axes in an orthogonal system of coordinates, *(vi)* geometrical representation of some symmetry breakings, compared with balls, *(vii)* geometrical representation of dependencies, especially, of variances and correlations, *(viii)* easy measurability and support for an approximate measuring of other sets and subsets.

The paper is organized as follows. In Section 2, we introduce and analyze time-continuous and time-discrete gene-environment networks, as a subclass of target-environment regulatory systems, without emphasizing the concept of uncertainty. A different discretization scheme is applied in the study of time-discrete networks and nonlinear mixed-integer programming is used to

calculate the unknown parameters in the optimization problem. Finally, the corresponding values of the gene-expression levels at discrete time points t_k $(k = 0, 1, \ldots, l)$ are obtained and compared with the given data to see the performance of the new method. The need and importance of studying gene-environment networks under the concept of uncertainty is explained and these networks are studied under the interval and ellipsoidal uncertainty in Section 3. Basic notions of interval calculus and ellipsoidal calculus, and interval games are also presented in this section. Section 4 gives some examples from the theory of cooperative interval games and discuss the ellipsoidal extension, then *Technology-Emission-Means* (TEM) model and its interval-valued version is investigated. Our remarks and future works are presented in Section 5.

2. Gene-Environment Networks

2.1. A class of dynamical models

The subject of genetic network is one of the important and promising research areas for most of the branches of modern science. A genetic network can be expressed as a weighted directed graph containing nodes (vertices) that represent the genes, and the arcs with functional weights describing the influences of each gene onto the other genes in the network. The aim is to find and predict these interactions between genes. In the literature, many analytic and numerical methods are constructed to understand the structure and evolution of this kind of networks (Weber et al. (2008c); Ahuja et al. (1993); Chen et al. (1999); DeRisi et al. (1997); Gebert et al. (2004a,b,c, 2006, 2007); Huang (1999); Özcan et al. (2005); Özöğür et al. (2005); Taştan (2005); Taştan et al. (2006); Uğur et al. (2009); Weber and Tezel (2007); Weber et al. (2008b, 2009b); Dress et al. (2009)) and also for the extended versions of gene-environment networks where the new nodes in the network represent environmental items (Uğur and Weber (2007); Weber et al. (2008c); Uğur et al. (2009); Weber and Tezel (2007); Weber et al. (2008b, 2009b)).

The dynamics of gene-environment networks were firstly described by the time-autonomous ordinary differential equations having the following form

$$\dot{\mathbb{E}} = \mathbb{F}(\mathbb{E}), \tag{1}$$

where $\mathbb{E} = (\mathbb{E}_1, \mathbb{E}_2, \ldots, \mathbb{E}_d)^T$ $(\mathbb{E} = \mathbb{E}(t), t \in I)$ is the d-vector of positive concentration levels of proteins and of certain levels of the environmental

factors. $\dot{\mathbb{E}}\,(=\frac{d\mathbb{E}}{dt})$ represents the continuous change in the gene-expression data with respect to time and $\mathbb{F}_i : \mathbb{R}^d \to \mathbb{R}$ are nonlinear coordinate functions of $\mathbb{F}$ (Chen et al. (1999); Uğur et al. (2009); Hoon et al. (2003); Sakamoto and Iba (2001)). From microarray experiments and environmental measurements, the experimental data vectors $\bar{\mathbb{E}}$ are obtained at the sample times. These data are used in the evaluation of parameters which are associated to and contained in the definition of $\mathbb{F}$. The data vector $\mathbb{E}(t_0) = \mathbb{E}_0$ contains the initial values of the gene-expression levels where $\mathbb{E}_0 = \bar{\mathbb{E}}_0$. Moreover, the gene-expression level (concentration rate) of the ith gene at time t is denoted by $E_i(t)$, and $E_i(t)$ stands for any of the first n coordinates in the d-vector $\mathbb{E}$ of genetic and environmental states noting that $G := \{1, 2, \ldots, n\}$ is the set of genes in the network.

There are many types of equation (1) which are defined and studied in the literature, where these equations represent the dynamics of the gene networks as follows (Weber et al. (2008c); Chen et al. (1999); Gebert et al. (2004a, 2006, 2007); Taştan (2005); Yılmaz (2004); Yılmaz et al. (2005); Sakamoto and Iba (2001); Taştan et al. (2005)):
(i) In the dynamical equation

$$\dot{E} = ME, \tag{2}$$

the matrix M is an $(n \times n)$-constant matrix and E is the $(n \times 1)$-vector of the gene-expression levels. The next equation,
(ii)

$$\dot{E} = M(E)E \tag{3}$$

represents and the dynamical system of the n genes and their interaction alone so that the matrix M is an $(n \times n)$-matrix with entries as functions of polynomials, exponential, trigonometric, splines or wavelets containing some parameters to be optimized.
(iii) An extended version of the model in *(ii)* is defined and studied in Yılmaz (2004) and Yılmaz et al. (2005) in order to include nonlinear interactions such as disturbances and genetic changes caused by the environment in long and in short term. In this extension, an affine linear shift term $C(E)$ is added and we get the following system (see Weber et al. (2008c); Taştan (2005); Taştan et al. (2006); Uğur et al. (2009); Taştan et al. (2005); Yılmaz (2004); Yılmaz et al. (2005); Weber et al. (2008b, 2009b) for details and other derivations of the extended models):

$$\dot{E} \;=\; M(E)E + C(E). \tag{4}$$

The initial value problem in the extended space, can be written in a multiplicative form as

$$\dot{\mathbb{E}} = \mathbb{M}(\mathbb{E})\mathbb{E}, \qquad \mathbb{E}_0 = \mathbb{E}(t_0) = \begin{bmatrix} E_0 \\ \check{E}_0 \end{bmatrix}, \tag{5}$$

where

$$\dot{\mathbb{E}} := \begin{bmatrix} E \\ \check{E} \end{bmatrix}, \qquad \mathbb{M}(\mathbb{E}) := \begin{pmatrix} M(E) & \check{M}(E) \\ 0 & 0 \end{pmatrix}, \tag{6}$$

are an $(n+m)$-vector and $(n+m)\times(n+m)$-matrix, respectively, since m environmental factors are included in the network.

The models given by the continuous dynamical equations in $(i), (ii)$ and (iii) can be written in general as follows (Uğur and Weber (2007); Weber et al. (2008c,b, 2009b)):

$$\dot{\mathbb{E}} = \mathbb{M}(\mathbb{E})\mathbb{E}, \tag{7}$$

with the initial value $\mathbb{E}_0 = \mathbb{E}(t_0)$, where $\mathbb{M}(\mathbb{E})$ is a $(d\times d)$-matrix, $\mathbb{E}$ and $\mathbb{E}_0$ are $(d\times 1)$-vectors.

2.2. Time discretization

The time-continuous models and equations presented in Subsection 2.1 in items (i), (ii) and (iii), can be approximated by applying discretization methods to evaluate the numerical solution at discrete time points. At this point, the importance of the choice of the method arise. The Euler's method was firstly used in the time-discretization for the gene-expression patterns and it has been seen that Euler's method gives inaccurate results and slow convergence (see Dubois and Kalisz (2004) for further information). As a next step, Runge-Kutta methods were introduced in Ergenç and Weber (2004) and, specifically, the 2nd order Heun's method was studied in Taştan (2005) and Taştan et al. (2006, 2005). Runge-Kutta methods have advantages in terms of truncation error, stability (which is closer to the stability of the time-continuous model), and implementation (Heath (2002)) when it is compared with the Euler's method. Therefore, the choice of the method in the numerical study plays an important role. In the work of Defterli et al. (2009), the 3rd-order Heun's method is used in the discretization of the time-continuous models to improve the rate of convergence and accuracy.

Consider the general case described by the model in (7) and applying the 3rd-order Heun's numerical method, we formulate the corresponding time-

discrete model as follows (Defterli et al. (2009)):

$$\begin{aligned}
\mathbb{E}^{(k+1)} &= \mathbb{E}^{(k)} + \frac{h_k}{4}(k_1 + 3k_3), \qquad (8)\\
k_1 &= \mathbb{M}(\mathbb{E}^{(k)})\mathbb{E}^{(k)},\\
k_2 &= \mathbb{M}(\mathbb{E}^{(k)} + \frac{h_k}{3}k_1)(\mathbb{E}^{(k)} + \frac{h_k}{3}k_1),\\
k_3 &= \mathbb{M}(\mathbb{E}^{(k)} + \frac{2h_k}{3}k_2)(\mathbb{E}^{(k)} + \frac{2h_k}{3}k_2);
\end{aligned}$$

which can be rewritten as

$$\begin{aligned}
\mathbb{E}^{(k+1)} &= \mathbb{E}^{(k)} + \frac{h_k}{4}\mathbb{M}(\mathbb{E}^{(k)})\mathbb{E}^{(k)} + \mathbb{M}(\mathbb{E}^{(k)} + \frac{2h_k}{3}\mathbb{M}(\mathbb{T}^{(k)})\mathbb{T}^{(k)})\\
&\quad \times\{\frac{3h_k}{4}\mathbb{E}^{(k)} + \frac{h_k^2}{2}\mathbb{M}(\mathbb{T}^{(k)})\mathbb{E}^{(k)} + \frac{h_k^3}{6}\mathbb{M}(\mathbb{T}^{(k)})\mathbb{M}(\mathbb{E}^{(k)})\mathbb{E}^{(k)}\}, \qquad (9)
\end{aligned}$$

where $\mathbb{T}^{(k)} = \mathbb{E}^{(k)} + \frac{h_k}{3}\mathbb{M}(\mathbb{E}^{(k)})\mathbb{E}^{(k)}$. The time-discrete equation is obtained as

$$\mathbb{E}^{(k+1)} = \mathbb{M}^{(k)}\mathbb{E}^{(k)}, \qquad (10)$$

with the matrix $\mathbb{M}^{(k)}$ defined in below

$$\begin{aligned}
\mathbb{M}^{(k)} &:= \mathbb{I} + \frac{h_k}{4}\mathbb{M}(\mathbb{E}^{(k)}) + \mathbb{M}(\mathbb{E}^{(k)} + \frac{2h_k}{3}\mathbb{M}(\mathbb{T}^{(k)})\mathbb{T}^{(k)})\\
&\quad \times\{\frac{3h_k}{4}\mathbb{I} + \frac{h_k^2}{2}\mathbb{M}(\mathbb{T}^{(k)}) + \frac{h_k^3}{6}\mathbb{M}(\mathbb{T}^{(k)})\mathbb{M}(\mathbb{E}^{(k)})\}. \qquad (11)
\end{aligned}$$

The approximate values of the gene-expression data at the next time level can be obtained from the previous time level by using the iterative formula in (10). The vector $\bar{\mathbb{E}}^{(\kappa)}$ $(\kappa = 0, 1, \ldots, l-1)$ in the extended space denotes the DNA microarray experimental data and the environmental items obtained at the time-level t_k. The approximations obtained by (10) are expressed by $\widehat{\mathbb{E}}^{(\kappa)}$ $(\kappa = 0, 1, \ldots, l-1)$, and set $\widehat{\mathbb{E}}^{(0)} = \mathbb{E}^{(0)}$ for the initial values. The kth approximation or prediction, $\widehat{\mathbb{E}}^{(k)}$, is calculated as $\widehat{\mathbb{E}}^{(k)}(:= \mathbb{E}^{(k)}) = \mathbb{M}^{(k-1)}((\cdots\mathbb{M}^{(1)}(\mathbb{M}^{(0)}\mathbb{E}^{(0)})))$, where $k \in \mathbb{N}_0$. Hence, the time-discrete dynamics of the gene-environment networks can be obtained by using formula (10). In the matrix $\mathbb{M}^{(k)}$ in above derivations, the entry $m_{ij}^{(k)}$ stands for the coefficient of proportionality (i.e., multiplied by $\mathbb{E}_j^{(k)}$) describing that the ith gene (or environmental factor) becomes changed by the jth gene (or environmental factor or the cumulative environmental item) in the step from time level k to $k+1$ (Defterli et al. (2009)).

2.3. Matrix arithmetics applied

Consider the following canonical form of matrix partitioning (cf. Uğur and Weber (2007); Weber et al. (2008c); Taştan et al. (2006, 2005)) for the time-continuous model given by equation (5)

$$\mathbb{M}(\mathbb{E}) = \begin{pmatrix} M(E) & \check{M}(E) \\ 0 & 0 \end{pmatrix}, \tag{12}$$

where $M(E)$ and $\check{M}(E)$ are the matrices having dimensions $n \times n$ and $n \times m$, respectively. The matrix $\mathbb{M}(\mathbb{E})$ is $(n+m) \times (n+m)$ and $\mathbb{E} = (E^T, \check{E}^T)^T$ and $\mathbb{T} = (T^T, \check{T}^T)^T$ are $(n+m)$-vectors. The structure of the gene and gene-environment network, so that the relations between the genes and the environment are represented by these matrices where $\mathbb{M}^{(k)}$ will be the basis of the networks. Since the product of two such canonical matrices is again canonical (see Uğur and Weber (2007); Weber et al. (2008c); Taştan (2005); Taştan et al. (2006); Weber et al. (2008b, 2009b); Taştan et al. (2005)), after simplifications Defterli et al. (2009) formulate that

$$\begin{aligned} \mathbb{M}^{(k)} &= \mathbb{I} + \frac{h_k}{4}\begin{pmatrix} M(E^{(k)}) & \check{M}(E^{(k)}) \\ 0 & 0 \end{pmatrix} + \frac{3h_k}{4}\begin{pmatrix} A & \tilde{A} \\ 0 & 0 \end{pmatrix} \\ &\quad + \frac{h_k^2}{2}\begin{pmatrix} B & \tilde{B} \\ 0 & 0 \end{pmatrix} \frac{h_k^3}{6}\begin{pmatrix} C & \tilde{C} \\ 0 & 0 \end{pmatrix}, \end{aligned} \tag{13}$$

where

$$\begin{aligned} A &:= M(E^{(k)} + \frac{2h_k}{3}(M(T^{(k)})T^{(k)} + \check{M}(T^{(k)})\check{T}^{(k)})), \\ \widetilde{A} &:= \check{M}(E^{(k)} + \frac{2h_k}{3}(M(T^{(k)})T^{(k)} + \check{M}(T^{(k)})\check{T}^{(k)})), \\ B &:= M(E^{(k)} + \frac{2h_k}{3}(M(T^{(k)})T^{(k)} + \check{M}(T^{(k)})\check{T}^{(k)}))M(T^{(k)}), \\ \widetilde{B} &:= M(E^{(k)} + \frac{2h_k}{3}(M(T^{(k)})T^{(k)} + \check{M}(T^{(k)})\check{T}^{(k)}))\check{M}(T^{(k)}), \\ C &:= M(E^{(k)} + \frac{2h_k}{3}(M(T^{(k)})T^{(k)} + \check{M}(T^{(k)})\check{T}^{(k)}))M(T^{(k)})M(E^{(k)}), \\ \widetilde{C} &:= M(E^{(k)} + \frac{2h_k}{3}(M(T^{(k)})T^{(k)} + \check{M}(T^{(k)})\check{T}^{(k)}))M(T^{(k)})\check{M}(E^{(k)}), \end{aligned}$$

and $T^{(k)} := E^{(k)} + \frac{h_k}{3}M(E^{(k)})E^{(k)}$, $\check{T}^{(k)} := \check{E}^{(k)} + \frac{h_k}{3}\check{M}(E^{(k)})\check{E}^{(k)}$, and $\mathbb{I} = I_d$ ($(d \times d)$-unit matrix) with $d = n + m$.

Finally, the canonical block form of $\mathbb{M}^{(k)}$ is:

$$\begin{pmatrix} \widehat{M(E^{(k)})} & \widehat{\dot{M}(E^{(k)})} \\ 0 & I_m \end{pmatrix}. \tag{14}$$

2.4. The optimization problem

In the general form of the gene-environment network given in Section 2 which is described by the dynamical equation $\dot{\mathbb{E}} = \mathbb{M}(\mathbb{E})\mathbb{E}$, the entries of matrix $\mathbb{M}(\mathbb{E})$, which can be polynomial, trigonometric, exponential, but otherwise logarithmic, hyperbolic, spline, etc., contain some parameters to be evaluated (Aster et al. (2004); Hastie et al. (2001)). These entries stand for the changes in the genetic or environmental concentration rates that we suppose by any kind of a priori information, observation or assumption (Gebert et al. (2004a)).

Two different levels of the problem concerning the parametrized entries of the matrices can just be distinguished: *optimization* and *stability analysis*, both of them constituting bilevel problems (Uğur and Weber (2007); Weber et al. (2008c); Gebert et al. (2004c, 2006); Weber et al. (2008b, 2009b)). The first step is the *optimization problem* of approximation with respect to squared errors:

$$\min_{y} \sum_{\kappa=0}^{l-1} \left\| \mathbb{M}_y(\bar{\mathbb{E}}^{(\kappa)})\bar{\mathbb{E}}^{(\kappa)} - \dot{\bar{\mathbb{E}}}^{(\kappa)} \right\|_2^2, \tag{15}$$

where y is the vector of a subset of all the parameters and the $\dot{\bar{\mathbb{E}}}^{(\kappa)}$ are the difference quotients based on the κth experimental data $\bar{\mathbb{E}}^{(\kappa)}$ with interval lengths $\bar{h}_\kappa := \bar{t}_{\kappa+1} - \bar{t}_\kappa$ between neighbouring samplings (Uğur and Weber (2007); Gebert et al. (2004a, 2007)).

The second step is the *stability of the dynamics* investigated with respect to the remaining parameters. For this a combinatorial algorithm based on polyhedra sequences is used to specify the regions of stability (Uğur and Weber (2007); Gebert et al. (2004a, 2006, 2007)).

Real-world gene-environment networks are huge, such that we have to rarefy them, for practical reasons, by reducing the number of arcs (Uğur and Weber (2007); Weber et al. (2008c)). The upper bounds on the outdegrees of nodes are introduced firstly; later on, these constraints are undergoing a relaxation. Therefore, the Boolean matrices $X = (\chi_{ij})_{i,j=1,\dots,n}$, $\Xi = (\xi_{i\ell})_{\substack{i=1,\dots,n \\ \ell=1,\dots,m}}$ and the vector $Z = (\zeta_i)_{i=1,\dots,n}$ are introduced in (Uğur and Weber (2007); Weber et al. (2008b,c, 2009b)), where the values of their entries are 1 if gene j regulates gene i, environmental item ℓ regulates gene i and the environment cumulatively regulates gene i, and 0 otherwise. Hence, $\sum_{i=1}^{n} \chi_{ij}$, $\sum_{i=1}^{n} \xi_{i\ell}$ and $\sum_{i=1}^{n} \zeta_i$ are the numbers of genes regulated by gene j (its *outdegree*), by environmental item ℓ or by the cumulative

environment, respectively. Our network rarefication by bounding the outdegrees obeys the principles of least squares (or maximum likelihood). We also imply any helpful a priori knowledge into the problem, especially, about degradation rates which is important for the connectedness of the network. A lower bound $\delta_{i,\min}$ on the degradation of gene i is often known or requests about the feasibility of special genetic or metabolic processes are given (Weber et al. (2008c); Gebert et al. (2007)). Now, the problem of parameter estimation takes the form of a *mixed-integer least squares approximation problem* (Uğur and Weber (2007); Weber et al. (2008b,c, 2009b)) as follows:

$$\min_{(m_{ij}),(\chi_{ij}),(\xi_{i\ell}),(\zeta_i)} \quad \sum_{\kappa=0}^{l-1} \left\| \mathbb{M}_y(\bar{\mathbb{E}}^{(\kappa)})\bar{\mathbb{E}}^{(\kappa)} - \dot{\bar{\mathbb{E}}}^{(\kappa)} \right\|_2^2,$$

subject to

$$\begin{aligned}
\sum_{i=1}^{n} \chi_{ij} &\le \alpha_j \qquad (j = 1, 2, \ldots, n), \\
\sum_{i=1}^{n} \xi_{i\ell} &\le \beta_\ell \qquad (\ell = 1, 2, \ldots, m), \\
\sum_{i=1}^{n} \zeta_i &\le \gamma, \\
m_{ii} &\ge \delta_{i,\min} \qquad (i = 1, 2, \ldots, n).
\end{aligned} \tag{16}$$

The loss of some edges emanating at a few genes which are considered to play a very important role in regulation, i.e., to have very high outdegrees, could strongly restrict the connectivity of the network. Such a loss can be the result of perturbations caused by the environment and affecting the problem in (16) with its rigid (exclusive) binary constraints.

2.5. The mixed-integer problem

Defterli et al. (2009) studied the model

$$\dot{E} = M(E)E, \tag{17}$$

as an example for the mixed-integer least squares approximation problem together with the newly iteration formula. Here, M is a $(n \times n)$-constant matrix with entries m_{ij} representing the amount of effect which the expression level of gene j has on the change of expression of gene i, where $G = \{1, \ldots, n\}$ is the set of genes (but

environmental factors could be included here, too). The matrix M is defined as *genetic regulation network* (Gebert et al. (2004a)). In order to compute a gene network based on gene expression data, Defterli et al. (2009) solved the following *mixed-integer nonlinear optimization problem* (MINLP):

$$\min_{M=(m_{ij})} \quad \sum_{k=1}^{l} \left\| M\bar{E}^{(k)} - \dot{\bar{E}}^{(k)} \right\|_2^2. \tag{18}$$

The above problem has a high degree of freedom so that it is needed to restrict the solution space according to the underlying biological motivation (Gebert et al. (2004a, 2006, 2007)). Otherwise, a very big amount of expression data is necessary to solve the minimization problem in (18). If the two genes i and j in the network do not interact at all, then $m_{ij} = 0$; normally, the values m_{ij} are nonnegative since no gene consumes another one. A constant vector $\lambda \in \mathbb{R}^n$ describes the lower bound for the amount of decrease of the transcript concentration between two time steps. Therefore, for $i, j \in G$ we have

$$m_{ij} \geq \begin{cases} -\lambda(i) & , \ i = j, \\ 0 & , \ i \neq j. \end{cases} \tag{19}$$

There is a need to limit the maximum outdegree and indegree of each node (see Gebert et al. (2004a, 2006, 2007) for biological motivation). Then, we bound the indegree of each gene i by a given parameter $deg_{max,i} \in \mathbb{Z}_+$ and introduce binary variables $y_{ij} \in \{0, 1\}$ as:

$$y_{ij} = \begin{cases} 0 & , \text{ if } m_{ij} = 0, \\ 1 & , \text{ if } m_{ij} \neq 0. \end{cases} \tag{20}$$

Reformulating (20) gives the nonlinear constraints:

$$(1 - y_{ij}) \cdot m_{ij} = 0, \qquad \forall\, i, j \in G. \tag{21}$$

The number of nonzero entries per row of the matrix $M = (m_{ij})_{1 \leq i,j \leq n}$ can be limited by the degree number, $deg_{max,i}$ as in the following:

$$\sum_{j \in G} y_{ij} \leq deg_{max,i} \qquad \forall\, i \in G. \tag{22}$$

The aim now is to solve the MINLP

$$\min (18), \ \text{subject to } \{(19), (21), (22)\}. \tag{23}$$

2.5.1. Numerical results

Defterli et al. (2009) numerically solved the problem in (23) for the model described by (17). According to the data in the paper of Gebert et al. (2004a), Table 1 presents four different genes and their expression levels at four different times.

time / genes	A	B	C	D	
1	255	250	0	255	$= \bar{E}_1^T$
2	255	200	50	0	$= \bar{E}_2^T$
3	255	180	70	255	$= \bar{E}_3^T$
4	255	170	80	0	$= \bar{E}_4^T$

Table 1. Expression scores of the genes A, B, C and D at four time points

An equally-spaced time discretization is used where $h_k = t_{k+1} - t_k := 1$ for all $k = 1, \ldots, l-1$. Applying the 3rd-order Heun's method to approximate the $\dot{\bar{E}}_t$ according to the data in Table 1, it is obtained that

$$\begin{array}{l} \dot{\bar{E}}_1^T = [0 \quad -50 \quad 50 \quad -255], \\ \dot{\bar{E}}_2^T = [0 \quad -20 \quad 20 \quad 255], \\ \dot{\bar{E}}_3^T = [0 \quad -20 \quad 20 \quad -255]. \end{array} \tag{24}$$

Then, by introducing the following constraints to the mixed-integer problem in (23)

$$\lambda(i) = 2, \qquad i = 1, \ldots, 4, \qquad deg_{max,i} = 2, \tag{25}$$

and solving the problem with the necessary software (see Defterli et al. (2009) for the details), the following matrix M is calculated

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.26 & -0.46 & 0 & 0 \\ 0.19 & 0 & -0.46 & 0 \\ 1 & 0 & 0 & -2 \end{pmatrix}, \tag{26}$$

where the corresponding objective function value is 92.31. Next, derive the 3rd-order Heun's time discretization formula for our model in (17) as

$$\begin{aligned} E_{k+1} &= E_k + h_k M E_k + \frac{h_k^2}{2} M^2 E_k + \frac{h_k^3}{6} M^3 E_k \\ &= (I + h_k M + \frac{h_k^2}{2} M^2 + \frac{h_k^3}{6} M^3) E_k \\ &= \mathbf{M} E_k \end{aligned} \tag{27}$$

and by using the matrix M in (26) with the iteration formula in (27), the approximate values of gene expressions are obtained (see Defterli et al. (2009) for details).

According to the generated time-discrete series of gene-expressions, by using the 3rd-order Heun's discretization scheme, it can be concluded that the structural behavior of the obtained results is almost the same (constant first column, decreasing second column and increasing third column) with the given data in above Table 1. The values of the last column of the generated gene-expressions, we reach a more smooth behaviour instead of an alternating behaviour as in Table 1. So, the generated time series results are convergent and we reach the stable values after 23 iterations which shows the speed and performance of the newly applied 3rd-order Heun's discretization scheme.

In addition to the studied new technique, further discretization methods will be applied and corresponding derivations will be obtained in our future studies. As a next step, comparisons of these methods will be done in terms of efficiency, errors, rate of convergence and stability.

3. Gene-Environment Networks Under Uncertainty

3.1. Introduction

Errors, uncertainties and *measurement ambiguities* are among the characteristic features of technology, even to hightech. Especially, for a large amount of genes the expression levels can be monitored comfortably by *DNA-microarray experiments* (Carbayo et al. (2000)). However, despite the emerging technological progress, it is still influenced by imprecision, ambivalence and uncertainty (Uğur and Weber (2007); Weber et al. (2008b, 2009b)). For this reason, we imply these errors into our model. We are facing measurement and reliability problems, as for the environmental levels and concentrations, such that we represent them in error terms (Li et al. (2004)). In Uğur and Weber (2007); Weber et al. (2008b, 2009b), we presented various kinds of errors by *intervals*. As a next step, Kropat et al. (2009b) introduced and analyzed the time-discrete target-environment regulatory systems (TE-systems) under ellipsoidal uncertainty.

During the previous decade, we observed the fast growing generation of biological data, mainly by the release of whole genome sequences and gene chip data (Weber et al. (2008a, 2009a)). Employing high-throughput technologies needs the development of refined computational methods which could enlighten significant biological facts from that huge amount of generated data. While the data are produced by biologists, computer scientists aim at demonstrating that inference of a logical genetic network is possible only with sets of gene expression data, and mathematicians are working on the question how much data at least are necessary

to construct an optimized network. The standards for everyone to speak in the same academical language were posed for microarray experiments in December 2001 (Brazma et al. (2001)). The *Minimum Information About a Microarray Experiment* (*MIAME*) gives a help to present, use, record and exchange data in a standardized way. Additionally to efforts on a consensus, certain suggestions were made to diminish the uncertainties in results driven by microarray experiments, as collecting biological samples, increasing the number of replicates or laying more attention to validation (Allison et al. (2006)). The technique of high-throughput data generator microarrays at the same time measures the activities of some thousands of genes, and it is verified that this is a helpful method to analyze levels of gene expression. In order not to forget the related variances and the impact on the reliability of network analysis, we have take into account the errors related with the DNA chip experiments.

A core issue of gene-environment networks and of any approach to analyze them consists in their *complexity*. Therefore, we put upper bounds into the evaluation of the parametric constellation. Hence, we force the number of edges of the network to diminish and let the process of parameter estimation become a mixed continuous-discrete program. For both the modeling deficiencies and algorithmic reasons, we relax the inequality constraints, turning them to become continuous to depend on the environmental items and on uncertainties located in *intervals*; then the problem becomes a one from *semi-infinite programming* (*SIP*). In case where the uncertainty is represented by clusters of genetic or environmental items and by related *ellipsoids*, we will however arrive at *semidefinite programming* (*SDP*). In the interval-valued case, when allowing dependence of the domain of combined external effects on the unknown environmental parameters, we even obtain a *generalized semi-infinite programming* (*GSIP*) problem. By this we allow a regulation of the network's edge density in a more flexible way and can more confidently permit existence and tractability of genetic-metabolic and environmental processes. In Uğur and Weber (2007); Weber and Tezel (2007); Weber et al. (2008c,b, 2009b, 2008a, 2009a) we associated the combinatorial mathematics of networks with GSIP and SDP, by this proposing a new scientific approach into computational biology and ecology. While GSIP is a wide problem class with many motivations, results, challenges and uses (Rückmann and Gómez (2006); Stein (2003); Weber (2003)), SDP is a well established methodology for both continuous and discrete optimization with a great variety of applications (Vandenberghe and Boyd (1996); Alizadeh (1995); Wolkowicz et al. (2000)). In the computational fields of biology and environment, a reasonable *modeling*, *forecasting* and *process optimization* are very essential for a deep understanding of the related *processes*, the *optimization of cell metabolism* and their applications in medicine, health care, food production, in industry and energy supply.

The additional core items of the *environmental*, how they cause catalyzing and multiplicative effects, are becoming more and more important, especially, as we are living in an epoch of globalization, high mobility and rapid information exchange, and of multicausalities in all kinds of systems, communities and societies. This paper surveys about this situation and about recent contribution to its challenges by Operational Research.

3.2. Preliminaries on interval calculus and interval games

In this section, some useful results from the theory of cooperative interval games and some preliminaries from interval calculus are given (cf. Alparslan Gök et al. (2009a)).

In the model of cooperative interval games, the bounds for payoffs of coalitions are known with certainty. Formally, a cooperative interval game in coalitional form is an ordered pair $< N, w >$, where $N = \{1, 2, \ldots, n\}$ is the set of players, and $w : 2^N \to I(\mathbb{R})$ is the characteristic function such that $w(\emptyset) = [0, 0]$, where $I(\mathbb{R})$ is the set of all nonempty, compact intervals in $\mathbb{R}$ (Alparslan Gök et al. (2009b)). For each $S \in 2^N$, the worth set (or worth interval) $w(S)$ of the coalition S in the interval game $< N, w >$ is of the form $[\underline{w}(S), \overline{w}(S)]$. By IG^N we denote the family of all interval games with player set N.

Let $I, J \in I(\mathbb{R})$ with $I = \left[\underline{I}, \overline{I}\right]$, $J = \left[\underline{J}, \overline{J}\right]$, $|I| = \overline{I} - \underline{I}$ and $\alpha \in \mathbb{R}_+$. Then,

(i) $I + J = \left[\underline{I} + \underline{J}, \overline{I} + \overline{J}\right]$;

(ii) $\alpha I = \left[\alpha \underline{I}, \alpha \overline{I}\right]$.

By (i) and (ii) we see that $I(\mathbb{R})$ has a cone structure.

For further use we denote by $I(\mathbb{R})^N$ the set of all n-dimensional vectors whose components are elements in $I(\mathbb{R})$ and n-tuples of intervals are denoted by $I = (I_1, \ldots, I_n)$ where $I_i \in I(\mathbb{R})$ for each $i \in N = \{1, 2, \ldots, n\}$. We recall that I is weakly better than J, which we denote by $I \succcurlyeq J$, if and only if $\underline{I} \geq \underline{J}$ and $\overline{I} \geq \overline{J}$. For $w_1, w_2 \in IG^N$ and $\lambda \in \mathbb{R}_+$ we define $< N, w_1 + w_2 >$ and $< N, \lambda w >$ by $(w_1 + w_2)(S) = w_1(S) + w_2(S)$ and $(\lambda w)(S) = \lambda \cdot w(S)$ for each $S \in 2^N$. So, we conclude that IG^N endowed with "$\succcurlyeq$" is a partially ordered set and has a cone structure with respect to addition and multiplication with non-negative scalars described above.

Now, we recall that the interval imputation set $\mathcal{I}(w)$ of the interval game w, is defined by

$$\mathcal{I}(w) = \left\{ I \in I(\mathbb{R})^N | \sum_{i \in N} I_i = w(N), I_i \succcurlyeq w(i) \right\}$$

for each $i \in N$, and the interval core $\mathcal{C}(w)$ of the interval game w is defined by

$$\mathcal{C}(w) = \left\{ I \in \mathcal{I}(w) | \sum_{i \in S} I_i \succcurlyeq w(S) \right\}$$

for each $S \in 2^N \setminus \{\emptyset\}$.

3.3. Networks under interval uncertainty

As explained in Section 2, gene-environment networks were primarily modeled by time-continuous systems of autonomous ordinary differential equations (ODEs) in the form $\dot{\mathbb{X}} = \mathbb{F}(\mathbb{X})$, where the d-vector $\mathbb{X} = (\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_d)^T$ comprises the positive concentration levels of proteins (or mRNAs, or small components) and certain levels of the environmental factors. As the nonlinear function $\mathbb{F}$ is determined by unknown parameters we have to deal with identification based on noise-prone data vectors $\bar{\mathbb{X}}$ obtained from microarray and environmental measurements. For this reason we have $\mathbb{X}_i = \bar{\mathbb{X}}_i \pm \mathrm{Err}_i \quad (i = 1, 2, \ldots, d)$, where $\mathrm{Err}_i > 0$ denotes the maximal error to be made at the measurements of the gene- or environmental expression level $\mathbb{X}_i$ (Weber and Uğur (2007)). This measurement error leads us to assume that the state $\mathbb{X}_i$ has to lie in the interval $[\bar{\mathbb{X}}_i - \mathrm{Err}_i, \bar{\mathbb{X}}_i + \mathrm{Err}_i]$ and, hence, the state vector $\mathbb{X} = (\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_d)^T$ has to be in the parallelpipe $\prod_{i=1}^{d} [\bar{\mathbb{X}}_i - \mathrm{Err}_i, \bar{\mathbb{X}}_i + \mathrm{Err}_i]$. Here, we can speak of confidence intervals and a confidence parallelpipe. Those parallelpipes and intervals usually come from a perspective where functional dependencies among any two of the errors made in the measurements of the gene-expression environmental levels $\mathbb{X}_i$ are not taken into account explicitly. Moreover, they are usually smaller than the ellipsoides and their orthogonal projections into the 2-dimensional Cartesian planes, respectively Aster et al. (2004). 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. According to his/her experience and wish for confidence (trust region), the modeler can enforce a certain size of the parallelpipe by additional constraints on the interval limits, which are the variables in our parameter estimation. We underline that a direct modeling with ellipsoides and corresponding parameters is possible, too. Our work is a pioneering one, demonstrating a basic approach with the help of intervals. The size of the intervals and, by this, the amount of error in real networks, is an outcome of our parameter estimation which we do by optimization theory based on real data given. For more notions and details of interval algebra and comparison, including binary fuzzy operator and membership values, we refer to Colloins and Hu (2005); Fiedler et al. (2006); Weber and Uğur (2007).

3.4. Improved modeling by GSIP extension

Another extension of model (4) was introduced and analyzed through generalized semi-infinite optimization in Weber et al. (2008c). By splitting the shift vector $C(E)$ of (4) into the sum $\mathrm{W}(E)\check{E} + \mathrm{V}(X)$ we arrive at the decomposition

$$\dot{E} \;=\; A(E)E + \mathrm{W}(E)\check{E} + \mathrm{V}(E), \tag{28}$$

where the vector $\check{E}(t) \;=\; \big(\check{E}_1(t), \check{E}_2(t), \ldots, \check{E}_m(t)\big)^T$ of intervals consists of the levels of the m environmental factors which may have an impact on the gene-expression levels and on their variation. The *single effects* of the factors $\check{E}_\ell$ on the gene-expression data E_i can be implied into the system by the weight matrix $\mathrm{W} = (\mathrm{w}_{i\ell})_{\substack{i=1,\ldots,n \\ \ell=1,\ldots,m}}$, what matched the n genes and the m environmental factors individually. Additionally, the column vector $\mathrm{V}(E) = (\mathrm{v}_i)_{i=1,\ldots,n}$ gene by gene comprises the cumulative effects of all the environmental items i (please visit Weber et al. (2009c)).

To identify the unknown parameters of the regulatory system we have to minimize the quadratic error between the difference quotients $\bar{\dot{E}}^{(\kappa_\alpha)}$ and the right-hand side of the differential equations ($\alpha = 0, 1, \ldots, l^* - 1$):

$$\min_{(a^*_{ij}),(\mathrm{W}^*_{i\ell}),(\mathrm{V}^*_i)} \sum_{\alpha=0}^{l^*-1} \Big\| A^* \bar{E}^{(\kappa_\alpha)} + \mathrm{W}^* \bar{\check{E}}^{(\kappa_\alpha)} + \mathrm{V}^* - \bar{\dot{E}}^{(\kappa_\alpha)} \Big\|_\infty^2, \tag{29}$$

where $\|\cdot\|_\infty$ stands for the Chebychev-norm. As in Subsection 2.4, network rarefication can be achieved by diminishing the number of arcs (Uğur and Weber (2007); Weber et al. (2008c)). Here, upper bounds on the outdegrees of nodes are introduced firstly. Then, these constraints are further weakened by a continuous optimization way of model improvement and preparation. Firstly, we introduce the Boolean matrices and vectors, $\chi = (\chi_{ij})_{i,j=1,\ldots,n}$, $\Xi = (\xi_{i\ell})_{\substack{i=1,\ldots,n \\ \ell=1,\ldots,m}}$ and $Z = (\zeta_i)_{i=1,\ldots,n}$, representing by the values 1 and 0 whether or not gene j regulates gene i, environmental item ℓ regulates gene i and the environment cumulatively regulates gene i. The *outdegrees* $\sum_{i=1}^n \chi_{ij}$, $\sum_{i=1}^n \xi_{i\ell}$ and $\sum_{i=1}^n \zeta_i$ count the numbers of genes regulated by gene j, by environmental item ℓ or by the cumulative environment, respectively. This gives to our parameter estimation task the form of a (generalized) *mixed-integer Chebychev approximation problem*:

$$\min_{(a^*_{ij}),(\mathrm{w}^*_{i\ell}),(\mathrm{v}^*_i),(\chi_{ij}),(\xi_{i\ell}),(\zeta_i)} \sum_{\alpha=0}^{l^*-1} \left\| A^* \bar{E}^{(\kappa_\alpha)} + \mathrm{W}^* \bar{\bar{E}}^{(\kappa_\alpha)} + \mathrm{V}^* - \dot{\bar{E}}^{(\kappa_\alpha)} \right\|_\infty^2,$$

$$\text{subject to } \left\{ \begin{array}{rcll} \sum_{i=1}^n \chi_{ij} & \leq & \alpha_j & (j = 1, 2, \ldots, n), \\ \sum_{i=1}^n \xi_{i\ell} & \leq & \beta_\ell & (\ell = 1, 2, \ldots, m), \\ \sum_{i=1}^n \zeta_i & \leq & \gamma, & \\ a_{ii} & \geq & \delta_{i,\min} & (i = 1, 2, \ldots, n). \end{array} \right. \tag{30}$$

That *mixed-integer Chebychev approximation problem* (30) reveals binary constraints which are rigid (discontinuous) in nature but can lead to a high sensitivity of the problem and, hence, network, whenever edges are *knocked out*. To relax the effects of these constraints, we turn the binary variables χ_{ij}, $\xi_{i\ell}$ and ζ_i into real variables $p_{ij}, q_{i\ell}, r_i \in [0, 1]$ which linearly depend on the elements of a_{ij}, $\mathrm{w}_{i\ell}$ and v_i and suppose some reasonable box constraints. Herewith, the values $\sum_{j=1}^n p_{ij}(a^*_{ij})$, $\sum_{i=1}^m q_{i\ell}(\mathrm{w}^*_{i\ell})$ and $\sum_{i=1}^m r_i(\mathrm{v}^*_i)$ serve as interval-valued approximations of the numbers of genes regulated by gene j, environmental item ℓ and cumulative environment, respectively. This yields us a *continuous optimization problem* (Uğur and Weber (2007); Weber et al. (2008b,c, 2009b)). The *set of combined environmental effects* is introduced as the convex hull of all the vectors $\mathrm{w}^*_{i\ell} e_{a(i-1)+\ell}$ and $\mathrm{v}^*_i e_{mn+i}$:

$$\begin{aligned} Y(\mathrm{V}^*, \mathrm{W}^*) &:= \mathrm{conv}\Big(\left\{ \mathrm{w}^*_{i\ell} e_{m(i-1)+\ell} \,\middle|\, i = 1, 2, \ldots, n;\ \ell = 1, 2, \ldots, m \right\} \\ &\qquad \cup \left\{ \mathrm{v}^*_i \sigma_{i,m+1} e_{mn+i} \,\middle|\, i = 1, 2, \ldots, n \right\} \Big) \\ &= \Bigg\{ \sum_{\substack{i=1,\ldots,n,\\ \ell=1,\ldots,m}} \sigma_{i\ell} \mathrm{w}^*_{i\ell} e_{m(i-1)+\ell} + \sum_{i=1,\ldots,n} \sigma_{i,m+1} \mathrm{v}^*_i e_{mn+i} \,\Bigg| \\ &\qquad \sum_{\substack{i=1,\ldots,n\\ \tau=1,\ldots,m+1}} \sigma_{i\tau} = 1,\ \ \sigma_{i\tau} \geq 0\ (i = 1, 2, \ldots, n;\ \tau = 1, 2, \ldots, m+1) \Bigg\}, \end{aligned}$$

where e_η stands for the ηth $((m+1)n)$-dimensional unit vector $(0, \ldots, 1, \ldots, 0)^T$. We may write $Y(\mathrm{V}^*, \mathrm{W}^*)$ as a parallelpipe (Weber et al. (2009a)) in the product form of

$$Y(\mathrm{V}^*, \mathrm{W}^*) = \prod_{\substack{i=1,\ldots,n\\ \ell=1,\ldots,m}} [0, \mathrm{w}^*_{i\ell}] \times \prod_{i=1,\ldots,n} [0, \mathrm{v}^*_i].$$

Then, we obtain our (generalized) *relaxed Chebychev approximation problem*:

$$\min_{(a^*_{ij}),(\mathrm{w}^*_{i\ell}),(\mathrm{v}^*_i)} \sum_{\alpha=0}^{l^*-1} \left\| A^* \bar{E}^{(\kappa_\alpha)} + \mathrm{W}^* \bar{\bar{E}}^{(\kappa_\alpha)} + \mathrm{V}^* - \dot{\bar{E}}^{(\kappa_\alpha)} \right\|_\infty^2, \tag{31}$$

subject to

$$\begin{array}{rcccll}
\sum_{i=1}^{n} p_{ij}(a_{ij}^*, y) & \leq & \alpha_j(y) & & & (y \in Y(\mathrm{V}^*, \mathrm{W}^*)), \\
\sum_{i=1}^{m} q_{i\ell}(\mathrm{w}_{i\ell}^*, y) & \leq & \beta_\ell(y) & & & (y \in Y(\mathrm{V}^*, \mathrm{W}^*)), \\
\sum_{i=1}^{m} r_i(\mathrm{v}_i^*, y) & \leq & \gamma(y) & & & (y \in Y(\mathrm{V}^*, \mathrm{W}^*)), \\
\delta_{i,\min} & \leq & a_{ii}^* & & & (i = 1, 2, \ldots, n), \\
\underline{a}_{ij}^* & \leq & a_{ij}^* & \leq & \overline{a}_{ij}^* & (i, j = 1, 2, \ldots, n), \\
\underline{\mathrm{w}}_{i\ell}^* & \leq & \mathrm{w}_{i\ell}^* & \leq & \overline{\mathrm{w}}_{i\ell}^* & (i = 1, 2, \ldots, n;\ \ell = 1, 2, \ldots, m), \\
\underline{\mathrm{v}}_i^* & \leq & \mathrm{v}_i^* & \leq & \overline{\mathrm{v}}_i^* & (i = 1, 2, \ldots, n).
\end{array}$$

We compare the values $\underline{a}_{ii}^*$ and $\delta_{i,\min}$ and choose the largest of both as a single lower bound instead, given that $\delta_{i,\min} < \overline{a}_{ii}^*$. With the objective function of our generalized Chebychev approximation, this uniform understanding of the "$\leq$" conditions introduces the SIP character of $(\mathcal{RCP})$. Moreover, through the additional coupling of the set $Y(\mathrm{V}^*, \mathrm{W}^*)$ of inequality constraints with the state variable of $(\mathrm{V}^*, \mathrm{W}^*)$, $(\mathcal{RCP})$ becomes a GSIP problem even. Inside of the objective function, the κth Chebychev norm terms $\|\cdot\|_\infty$ are nonsmooth max-type functions. Now, by a standard technique, $(\mathcal{RCP})$ becomes modeled smoothly: For each max-type function, additionally to the unknowns of $(\mathcal{RCP})$ we introduce a new coordinate τ_κ, as a uniform bound for the squared Euclidean norms of the elements in the Chebychev norms. So, we minimize the sum of the bounds, and as further inequalities we add these bounding conditions. We denote them such that the Euclidean norms of all the elements inside the Chebychev norms are to stay below their bounds uniformly. In this way applying GSIP for our gene-environment network problem $(\mathcal{RCP})$, we arrive at the following general program representation:

$$\mathcal{P}_{\mathcal{GSI}}(f, h, g, u, v) \qquad \left\{ \begin{array}{c} \text{minimize } f(x) \text{ on } M_{\mathcal{GSI}}[h, g], \text{ where} \\ M_{\mathcal{GSI}}[h, g] := \big\{ x \in \mathbb{R}^d \big|\, h_i(x) = 0\ (i \in I), \\ g^j(x, y) \geq 0\ (y \in Y^j(x),\ j \in J) \big\}, \end{array} \right\} \quad (\mathcal{A}_1)$$

where $|I|, |J| < \infty$, and the sets $Y^j = Y^j(x)$ are given in the form of *finitely constrained* $(\mathcal{F})$ feasible sets (Rückmann and Gómez (2006); Stein (2003); Weber (2003)):

$$\left. \begin{array}{rcc} Y^j(x) & = & M_{\mathcal{F}}[u^j(x, \cdot), v^j(x, \cdot)] \\ & := & \big\{ y \in \mathbb{R}^q \big|\, u_k(x, y) = 0\ (k \in K^j),\ v_\ell(x, y) \geq 0\ (\ell \in L^j) \big\}, \end{array} \right\} \quad (\mathcal{A}_2)$$

with sets $|K^j|, |L^j| < \infty$. The program $(\mathcal{A}_1)$-$(\mathcal{A}_2)$ permits equality constraints on both the upper (x-) level and lower (y-) level in order to imply additional metabolic restrictions, reactions or balance equations, etc. (Uğur and Weber (2007); Weber et al. (2008b,c)). In $(\mathcal{RCP})$, the outdegree inequalities may be of class C^2, too. The bounds imposed guarantee compactness of the feasible set $M_{\mathcal{GSI}}[h, g]$ projectively

in the sense of the original $2(n^2 + mn + n)$ unknowns, with intervals encoded by endpoint tuples. The noncompactness of the epigraph that would remain in the presence of the "height" variables could be overcome as explained in (Weber (2003)). The sets $Y^j(x)$ are compact and fulfill the *Linear Independence Constraint Qualification* (*LICQ*), an appropriate choice of the overall box constraints provided. Please visit (Stein (2003); Uğur and Weber (2007); Weber (2003); Weber et al. (2008c, 2009b)) for more explanations of details, characterizations of structural stability and generalizations of GSIP.

3.5. Preliminaries on ellipsoidal calculus

In this section, some preliminaries from ellipsoidal calculus are given based on Kropat et al. (2009b). The states of target and environmental variables of our model will be represented in terms of ellipsoids. Now, we introduce 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.

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 family of ellipsoids is closed with respect to affine transformations. Given an ellipsoid $\mathcal{E}(c, \Sigma) \subset \mathbb{R}^p$, a matrix $A \in \mathbb{R}^{m \times p}$ and a 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 a projection.

Given two non-degenerate ellipsoids $\mathcal{E}_1 = \mathcal{E}(c_1, \Sigma_1)$ and $\mathcal{E}_2 = \mathcal{E}(c_2, \Sigma_2)$, their geometric (Minkowski) 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. The range of values of $\mathcal{E}_1 + \mathcal{E}_1$ is contained in the ellipsoid

$$\mathcal{E}_1 \oplus \mathcal{E}_1 := \mathcal{E}(c_1 + c_2, \Sigma(s))$$

for all $s > 0$, where

$$\Sigma(s) = (1 + s^{-1})\Sigma_1 + (1 + s)\Sigma_2.$$

For a minimal and unique external ellipsoidal approximation an additional condition has to be fulfilled. The value of s is commonly chosen to minimize either the trace or the determinant of $\Sigma(s)$.

3.6. Networks under ellipsoidal uncertainty

Target-Environment regulatory systems (TE-regulatory systems), addressed in Kropat et al. (2009b), appear in various applications. They consist of two distinct groups of data which, although strongly related, display a completely different behaviour. The first group contains the targets: these are the most important variables of the system; they depend on an additional group of so-called environmental items. These regulatory systems occur in different applications, such as in gene-environment networks, i.e., modeling and prediction of gene-expression and environmental patterns, investigated to determine the complex interactions between genes, cells, tissues and their environmental determinants. In those cases, the target variables are the genetic expression levels, whereas the environmental items are toxins, transcription factors, radiation, etc. (Akhmet et al. (2005); Chen et al. (1999); Ergenç and Weber (2004); Gebert et al. (2004a,b,c, 2006, 2007); Hoon et al. (2003); Özöğür et al. (2005); Taştan (2005); Taştan et al. (2006, 2005); Uğur et al. (2009); Weber et al. (2008a, 2009a, 2008c,b); Weber and Tezel (2007); Weber et al. (2009b); Yılmaz (2004); Yılmaz et al. (2005)).

Usually, TE-models are constructed according to the measurements containing the effects of random noise and uncertainty, and we note that for the data corrupted by random noise the probability function is mostly assumed to be Gaussian (Kropat et al. (2009b)). This hypothesis has computational advantages but it is not sufficient, since in the case of real-world data we have to imply non-Gaussian or nonwhite noise. To resolve these obstacles, set-theoretic approaches can be developed where bounds are posed on the uncertain variables. In Kropat et al. (2009b), we focused on ellipsoids which have proved to be suitable for data affected with noise. Ellipsoids are very flexible with respect to correlations of the data, while intervals and parallelpipes usually come from a perspective where stochastic dependencies among any two of the errors in the gene or environmental measurements are not explicitly considered (Aster et al. (2004)). Furthermore, these sets are generally smaller than both the ellipsoids and their orthogonal projections into the 2-dimensional Cartesian planes (Aster et al. (2004)). In fact, these confidence ellipsoids are obtained with respect to stochastic dependencies of the error variables, which are the case in reality of microarray experiments and environmental studies. Reversely, each ellipsoid can be inscribed into an parallelpipe large enough that, in addition, could be located and directed in space around its eigenaxes suitably.

We now consider a time-discrete gene-environment regulatory system with n

genes and m environmental factors. Often, functionally related groups of genes and environmental items can be identified which exert a regulating influence on other groups of data items. In a preprocessing step of clustering, the set of genes is divided in R disjoint or overlapping clusters $C_r \subset \{1,\dots,n\}$ $(r = 1,\dots,R)$ and, similarly, the set of all environmental items is divided in S clusters $D_s \subset \{1,\dots,m\}$ $(s = 1,\dots,S)$. The uncertain states of these cluster are represented in terms of ellipsoids

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

with centers $\mu_r \in \mathbb{R}^{|C_r|}$, $\rho_s \in \mathbb{R}^{|D_s|}$ and shape matrices $\Sigma_r \in \mathbb{R}^{|C_r|\times|C_r|}$, $\Pi_s \in \mathbb{R}^{|D_s|\times|D_s|}$. The interactions between the clusters of genetic and environmental items are given by the linear model (Kropat et al. (2009a))

$$\begin{aligned} X_j^{(\kappa+1)} &= \xi_{j0} + \Big(\bigoplus_{r=1}^{R} \mathcal{A}_{jr}^{GG} X_r^{(\kappa)}\Big) + \Big(\bigoplus_{s=1}^{S} \mathcal{A}_{js}^{EG} E_s^{(\kappa)}\Big) \quad (j = 1,2,\dots,R) \\ E_i^{(\kappa+1)} &= \zeta_{i0} + \Big(\bigoplus_{r=1}^{R} \mathcal{A}_{ir}^{GE} X_r^{(\kappa)}\Big) + \Big(\bigoplus_{s=1}^{S} \mathcal{A}_{is}^{EE} E_s^{(\kappa)}\Big) \quad (i = 1,2,\dots,S) \end{aligned} \tag{32}$$

for $\kappa \geq 0$. The system (32) is based on ellipsoidal calculus (Kurzhanski and Vályi (1997); Kurzhanski and Varaiya (2008)) and (affine) linear coupling rules determine the (ellipsoidal) future states of genetic and environmental clusters. The sums $\bigoplus_{r=1}^{R} \mathcal{A}_{jr}^{GG} X_r^{(\kappa)}$ and $\bigoplus_{s=1}^{S} \mathcal{A}_{js}^{EG} E_s^{(\kappa)}$ of ellipsoids describe the *cumulative effects* of all genetic and environmental clusters exerted on the elements of cluster C_j in a set theoretic 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 $\mathcal{A}^{GG} \in \mathbb{R}^{n\times n}$, $\mathcal{A}^{EG} \in \mathbb{R}^{n\times m}$, $\mathcal{A}^{GE} \in \mathbb{R}^{m\times n}$, $\mathcal{A}^{EE} \in \mathbb{R}^{m\times m}$. The intercepts $\xi_{j0} \in \mathbb{R}^{|C_j|}$ and $\zeta_{i0} \in \mathbb{R}^{|D_i|}$ are components of the vectors $\xi_0 = (\xi_{10},\dots,\xi_{R0})^T \in \mathbb{R}^n$ and $\zeta_0 = (\zeta_{10},\dots,\zeta_{S0})^T \in \mathbb{R}^m$, respectively.

For a regression analysis we have to compare (ellipsoidal) measurements

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

$(r = 1, 2, \ldots, R,\ s = 1, 2, \ldots, S,\ \kappa = 0, 1, \ldots, T)$ taken at sampling times $t_0 < t_1 < \ldots < t_T$ and the first T predictions of model (32) given by the ellipsoids

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

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

$(j = 1, 2, \ldots, R,\ i = 1, 2, \ldots, S,\ \kappa = 0, 1, \ldots, T-1)$. The main idea of ellipsoidal regression analysis is to maximize the overlap of the predictions and measurement values (both ellipsoids). For this reason, we introduce the ellipsoids

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

$(r = 1, 2, \ldots, R,\ s = 1, 2, \ldots,\ \kappa = 1, \ldots, T)$, where $\cap$ denotes the fusion of ellipsoids (Ros et al. (2002)). In addition, the centers of the ellipsoids are adjusted, so that their distance becomes minimized. This leads us to the following regression problem:

$$\begin{aligned} \text{maximize} \quad & \sum_{\kappa=1}^{T} \Bigg\{ \sum_{r=1}^{R} \left\| \Delta X_r^{(\kappa)} \right\|_* - \left\| \widehat{\mu}_r^{(\kappa)} - \overline{\mu}_r^{(\kappa)} \right\|_2^2 \\ & \qquad + \sum_{s=1}^{S} \left\| \Delta E_s^{(\kappa)} \right\|_* - \left\| \widehat{\rho}_s^{(\kappa)} - \overline{\rho}_s^{(\kappa)} \right\|_2^2 \Bigg\}, \end{aligned} \tag{33}$$

where $\|\cdot\|_*$ denotes a measure that reflects the geometrical size of the intersections (fusions). There exist various measures related to the shape of the intersections, e.g., the *volume* (which corresponds to the ellipsoid matrix determinant), the *sum of squares of semiaxes* (which corresponds to the trace of the configuration matrix), the *length of the largest semiaxes* (which corresponds to the eigenvalues of the configuration matrix). Each of these measures leads to a specific regression problem and their solvability by semi-definite programming and interiour point methods is discussed in (Kropat et al. (2009a,b)).

4. Examples and Selected Related Topics

In this section, we turn to applications and further extensions of our theory on gene-environment networks under uncertainty. We demonstrate their usefulness for modern OR-applications in the energy sector and discuss their relation to the so-called Technology-Emissions Means model in CO_2-emissions control as an example of an eco-finance network (Kropat et al. (2008); Özceylan et al. (2009)). In

particular, we show how methods from game theory and the concepts of cooperative interval games and their extension to ellipsoidal games can become included. We note that these approaches can be further generalized and extended to an "uncertainty" concept of martingale theory for the financial sector as a core element of eco-finance networks (Stojaković (2009)).

4.1. A model from cooperative game theory

A *cooperative game under ellipsoidal uncertainty* is an ordered pair $< N, \tilde{w} >$, where $N = \{1, 2, \ldots, n\}$ is the set of players, and $\tilde{w} : 2^N \to \Xi$ is the characteristic function which assigns to each coalition $S \in 2^N$ an ellipsoid such that $\tilde{w}(\emptyset) = 0$, where Ξ is the family of all ellipsoids (Alparslan Gök and Weber (2009)).

Next, we give a motivating example from cooperative interval games and study its ellipsoidal extension.

EXAMPLE 4.1. (Interval glove game; Alparslan Gök (2009)) *Let $N = \{1, 2, 3\}$ consist of two disjoint subsets L and R of persons. The members of L possess each one left-hand glove, the members of R one right-hand glove. A single glove is worth nothing, a right-left pair of gloves is worth between 10 and 20 Euros. In case $L = \{1, 2\}$, this situation can be modeled as a three-person interval game, where the coalitions formed by players 1 and 3, players 2 and 3, and the grand coalition obtain the worth $[10, 20]$. The worth gained in other cases is $[0, 0]$, i.e., $w(1, 3) = w(2, 3) = w(1, 2, 3) = [10, 20]$ and $w(S) = [0, 0]$, otherwise. The interval core is $\mathcal{C}(w) = \{([0, 0], [0, 0], [10, 20])\}$.*

The following example illustrates the extension of an interval game to an ellipsoidal game.

EXAMPLE 4.2. (Ellipsoid glove game; Alparslan Gök and Weber (2009)) *Consider the game introduced in Example 4.1. In case $L = \{1, 2\}$, this situation can be modeled as a three-person ellipsoidal game, where the coalitions formed by players 1 and 3, players 2 and 3, and the grand coalition obtain the worth $\mathcal{E}(c, \Sigma)$. The worth gained in other cases is 0, i.e., $\tilde{w}(1, 3) = \tilde{w}(2, 3) = \tilde{w}(1, 2, 3) = \mathcal{E}(c, \Sigma)$ and $\tilde{w}(S) = 0$, otherwise.*

This example of the ellipsoid glove game, elaborating the interval glove game, gives rise to pose the general question of how the developed theory of collaborative interval games can be extended towards the collaborative ellipsoid case or, in reverse, how and to which extent ellipsoid games can be traced back to interval games.

4.2. The TEM-model and ellipsoidal cooperative games

The Technology-Emissions-Means Model (in short: TEM model) which was prepared at the occasion of the Kyoto protocol, allows a simulation of the cooperative economic behaviour of countries or enterprises with the aim of a reduction of greenhouse gas emissions, especially, of CO_2-emissions (Pickl (1998, 2001, 2002)).

Let us consider two complementary countries in technological sense, which do not have enough money to reduce their greenhouse gas emissions. In a sense we can consider these countries like one producing left-hand and the other right-hand gloves. These countries have to cooperate to reduce their emissions which shows the necessity of constructing a cooperative game.

Because the TEM model integrates the simulation of the technical and financial parameters and describes the economical interactions between several players (countries, companies) which intend to minimize their emissions or, in that process of emission reduction, to optimize their payoff function, by means of cooperative game theory, our ellipsoidal cooperative game model could help to these countries.

So the model of the ellipsoid glove game, elaborating the interval glove game given in this paper, gives rise to the importance and strict relationship of cooperative game theory with biology and ecology.

Further details on the TEM-model and its interval-valued reformulation are given in the next section.

4.3. TEM model, games and interval valued model reformulation

As mentioned above, the TEM model integrates both the simulation of the technical and financial parameters. It is treated as a time-discrete control problem (Lozovanu et al. (2004)). It describes the economical interactions between several actors (countries, or companies, etc., in general: players in a *game*) which intend to minimize their emissions, E_i, caused by technologies, T_i, using financial means, E_i, respectively. The index i stands for the ith player ($i = 1, 2, \ldots, N$). The players are linked by technical cooperations and the market, which expresses itself in the nonlinear time discrete dynamics of the TEM model (Pickl (2001, 2002)). For a transparent representation of the relationship between financial means and reduced emissions in a JI program, not mixing the indices, we rename the discrete times t_k by k and write them as arguments rather than as indices:

$$\Delta \mathsf{E}_i(k) = \sum_{j=1}^{N} \mathsf{em}_{ij}(k) \mathsf{M}_j(k),$$

$$\Delta \mathsf{M}_i(k) = -\lambda_i \mathsf{M}_i(k)(\overline{\mathsf{M}}_i - \mathsf{M}_i(k))(\mathsf{E}_i(k) + \varphi_i \Delta \mathsf{E}_i(k)).$$

Here, $\Delta \mathsf{E}_i(k) := \mathsf{E}_i(k+1) - \mathsf{E}_i(k)$ and $\Delta \mathsf{M}_i(k) := \mathsf{M}_i(k+1) - \mathsf{M}_i(k)$. Both differences can be interpreted as difference quotients $\dot{\mathsf{E}}_i^{(k)}$ and $\dot{\mathsf{M}}_i^{(k)}$, referring to a constant step length $h_k \equiv 1$. Furthermore, $\overline{\mathsf{M}}_i$ stands for the upper bounds for the financial investigations. The first equation describes the time-dependent behaviour of the emissions reduced so far by each player. These levels E_i $(i = 1, 2, \ldots, N)$ are influenced by financial investigations M_j $(j = 1, 2, \ldots, N)$ which are restricted by the second equation. We understand E_i as the reduced emissions of actor i in % and M_i as the financial means of actor i. The parameters φ_i are called *memory parameters*. Thus, the multiplication of $\Delta \mathsf{E}_i$ with φ_i can be regarded as a *memory effect*; this expression stands for the influence of earlier investments. The first part of the second equation resembles a logistic difference equation, where the proportional factor λ_i can be seen as a *growth parameter*. Each coefficient em_{ij} describes the effect on the emissions of the ith actor if the jth actor invests one unit of money for his technologies, e.g., devices of filters in energy production of consumption. This also shows how effective technology cooperations are, what is the kernel of the JI program (Pickl (1998)).

In the first equation, the level of the reduced emissions at the kth discrete time point depends upon the last value plus a market effect. This effect expresses itself in the additive terms which might be negative or positive. In general, $\mathsf{E}_i > 0$ implies that the actors have reached yet the demanded value $\mathsf{E}_i = 0$ (normalized *Kyoto-Level*). A value $\mathsf{E}_i < 0$ means that the emissions are less than the requirements of the treaty. The second equation reveals that for such a situation the financial means increase, whereas $\mathsf{E}_i > 0$ leads to a reduction: $\mathsf{M}_i(k+1) = \mathsf{M}_i(k) - \lambda_i \mathsf{M}_i(k)(\overline{\mathsf{M}}_i - \mathsf{M}_i(k))(\mathsf{E}_i(k) + \varphi_i \Delta \mathsf{E}_i(k))$.
The TEM model is a mathematical model which supports the development of a management tool in the creation of a JI program, which intends to strengthen technical cooperations in order to fulfill Kyoto Protocol. The different parameters em_{ij} stands for the technical relationships between the actors. Their economical interpretation can lead to case studies, in which the range of relevant data can be gained. These data sets might be a good basis for the iterative solution and game theoretic approach (Pickl (1998)).

The TEM model and its controlled version are time-discrete systems. Aiming at the time-discrete dynamics discussed, it can firstly be structured in this way: $(\mathsf{E}^T, \mathsf{M}^T)^{T^{(k+1)}} = M^{(k)}((\mathsf{E}^T, \mathsf{M}^T)^{T^{(k)}})(\mathsf{E}^T, \mathsf{M}^T)^{T^{(k)}}$. Having added the control parameter, we obtain:

$$\begin{pmatrix} \mathsf{E} \\ \mathsf{M} \end{pmatrix}^{(k+1)} = M^{(k)}\left(\begin{pmatrix} \mathsf{E} \\ \mathsf{M} \end{pmatrix}^{(k)}\right)\begin{pmatrix} \mathsf{E} \\ \mathsf{M} \end{pmatrix}^{(k)} + \begin{pmatrix} 0 \\ u^{(k)} \end{pmatrix},$$

which we compactly write as

$$(\mathcal{DE}) \qquad \mathbb{E}^{(k+1)} = \mathbb{M}^{(k)}\mathbb{E}^{(k)},$$

such that, now, the matrices $\mathbb{M}^{(k)}$ incorporate the control variables. In this extended space notation, the variable $\mathbb{E}$ and entire dynamics $(\mathcal{DE})$ could be enriched by further environmental and, in particular, genetical items and relations. The shift vector $(0^T, (u^{(k)})^T)^T$ can be regarded as parametric and as a realization of $\mathrm{V}(E, \check{E}^{\vee})$ then, our stability theory could be employed. According to how those matrices are adjusted, we arrive at different behaviours of stability or instability of $(\mathcal{DE})$, in the sense of dynamical systems or of parameter estimation. As a dual alternative to that feedback-like realization by the vector $\mathrm{V}(E, \check{E}^{\vee})$ which becomes incorporated into the matrix $\mathbb{M}^{(k)}$, the control vectors $u^{(k)}$ could also become integrated into $\mathbb{E}^{(k)}$. The time-dependent parameters $\mathsf{em}_{ij}^{(k)}$ can be treated in similar ways as the controls (cf. Weber et al. (2008a)).

5. Conclusion

In the first part of the paper, we contribute to an improved modeling of gene-environment networks, including their rarefication which may be regarded as a regularization, and to the numerical solution of their dynamics. By this, we supported to a better future prediction of evolution and behaviour of these kind of networks in time, with important consequences in health care, environmental protection, finance and in the field of education.

Combining our new numerical methods with those concepts of correlation, uncertainty and robustness will be our promising works to make modeling and prediction of such networks more accurate and stable.

We regard our survey paper as an invitation to the international community of researchers and practitioners to make use of the rich methodologies and toolboxes of modern OR. Indeed, OR provides the analytical tools of optimization, systems theory and decision making to investigate great present challenges, but also to create great chances, and to offer solutions. In this paper, we paid a special attention to the aspect of tacking uncertainty and of involving games, in fact, collaboration, into the exchange platforms to prepare a future for our peoples, fostered by OR.

Acknowledgement The authors cordially thank the coauthors of all the works on which this survey article bases, and they express their gratitude to EJOR for the invitation to this paper project. This work is partially supported by the Scientific and Technical Research Council of Turkey. A part of this paper contributes to the

Ph.D. thesis of Ozlem Defterli at the Department of Mathematics in Middle East Technical University, Ankara, Turkey.

References

Achterberg T. Constraint integer programming. *PhD. Thesis*. Technische Universität Berlin: Berlin; 2007.

Ahuja RK, Magnanti TL, Orlin JB. Network Flow: Theory, Algorithms and Applications. Prentice Hall: N.J.; 1993.

Akhmet MU, Gebert J, Öktem H, Pickl SW, Weber, G-W. An improved algorithm for analytical modeling and anticipation of gene expression patterns. Journal of Computational Technologies 2005; 10; 4; 3-20.

Alizadeh, F. Interior-point methods in semidefinite programming with applications to combinatorial optimization. SIAM Journal on Optimization 1995; 5; 1; 13-51.

Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. Nature Reviews Genetics 2006; 7; 55-65.

Alparslan Gök SZ. Cooperative Interval Games. *PhD. Dissertation*. Middle East Technical University. Ankara, Turkey; 2009.

Alparslan Gök SZ, Branzei R, Tijs S. Convex interval games. Journal of Applied Mathematics and Decision Sciences 2009; vol. 2009; Article ID 342089; 14 pages. DOI: 10.1115/2009/342089.

Alparslan Gök SZ, Miquel S, Tijs S. Cooperation under interval uncertainty. Mathematical Methods of Operations Research 2009; 69; 1; 99-109.

Alparslan Gök SZ, Weber G-W. Cooperative games under ellipsoidal uncertainty. Preprint No 148, Institute of Applied Mathematics, METU 2009 (to appear in the proceedings of PCO 2010, 3rd Global Conference on Power Control and Optimization, February 2-4, 2010, Gold Coast, Queensland, Australia).

Aster A, Borchers B, Thurber C. Parameter estimation and inverse problems. Academic Press: San Diego; 2004.

Berthold T, Heinz S, Vigerske S. Extending a CIP framework to solve MIQCPs. Konrad-Zuse-Zentrum fur Informationstechnik, Berlin. Technical Report ZIB-TR 09-23, 2009.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nature Genetics 2001; 29; 4; 3; 65-371.

Carbayo MS, Bornman W, Cardo CC. DNA Microchips: technical and practical considerations. Current Organic Chemistry 2000; 4; 9; 945-971.

Chen T, He HL, Church GM. Modeling gene expression with differential equations. In: Proceedings of Pacific Symposium on Biocomputing 1999; 29-40.

Collins WD, Hu C. Fuzzily determined interval matrix games. `http://www-bisc.cs.berkeley.edu/BISCSE2005/Abstracts_Proceeding/Friday/FM3/Chenyi_Hu.pdf`.

Córdoba Bueno M. Fundamentals and Practice of Financial Mathematics. Dykinson S.L.: Madrid, Spain; 2006.

Defterli O, Fügenschuh A, Weber G-W. New discretization and Ooptimization techniques with results in the dynamics of gene-environment networks. Preprint No 153, Institute of Applied Mathematics: METU; 2009 (to appear in the proceedings of PCO 2010, 3rd Global Conference on Power Control and Optimization, February 2-4, 2010, Gold Coast, Queensland, Australia).

DeRisi J, Iyer V, Brown P. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 1997; 278; 680-686.

Dress A, Karasözen B, Stadler PF, Weber G-W (guest editors). Special issue: Networks in Computational Biology. Discrete Applied Mathematics 2009; 157; 10.

Dubois DM, Kalisz E. Precision and stability of Euler, Runge-Kutta and incursive algorithm for the harmonic oscillator. International Journal of Computing Anticipatory Systems 2004; 14; 21-36.

Ergenç T, Weber G-W. Modeling and prediction of gene-expression patterns reconsidered with Runge-Kutta discretization. Special issue at the occasion of 70th birthday of Prof. Dr. Karl Roesner, TU Darmstadt. Journal of Computational Technologies 2004; 9; 6; 40-48.

Fiedler M, Nedoma J, Ramik J, Rohn J, Zimmermann K. Linear optimization problems with inexact data. Springer Verlag: Berlin; 2006.

Gebert J, Lätsch M, Pickl SW, Weber G-W, Wünschiers R. Genetic networks and anticipation of gene expression patterns. In: Computing Anticipatory Systems: CASYS(92)03 - Sixth International Conference. AIP Conference Proceedings. vol.718. 2004. 474-485.

Gebert J, Lätsch M, Quek EMP, Weber G-W. Analyzing and optimizing genetic network structure via path-finding. Journal of Computational Technologies 2004; 9; 3; 3-12.

Gebert J, Lätsch M, Pickl SW, Weber G-W, Wünschiers R. An algorithm to analyze stability of gene-expression pattern. In: Anthony M, Boros E, Hammer PL, Kogan A (Guest Eds.), (special issue) Discrete Mathematics and Data Mining II, vol.154 no.7. Discrete Applied Mathematics; 2006. 1140-1156.

Gebert J, Öktem H, Pickl SW, Radde N, Weber G-W, Yılmaz FB. Inference of gene expression patterns by using a hybrid system formulation - an algorithmic approach to local state transition matrices. In: Lasker GE, Dubois DM (Eds), Anticipative and Predictive Models in Systems Science I, IIAS (International Institute for Advanced Studies) in Windsor: Ontario; 2004. 63-66.

Gebert J, Radde N, Weber G-W. Modelling gene regulatory networks with piecewise linear differential equations. In: special issue (feature cluster) Challenges of Continuous Optimization in Theory and Applications. European Journal of Operational Research 2007; 181; 3; 1148-1165.

Gökmen A, Kayaligil S, Weber G-W, Gökmen I, Ecevit M, Sürmeli A, Bali T, Ecevit Y, Gökmen H, DeTombe DJ. Balaban Valley Project: Improving the quality of life in rural area in Turkey. International Scientific Journal of Methods and Models of Complexity 2004; 7; 1. www.fss.uu.nl/ms/cvd/isj/index7-1.htm.

Guckenheimer J, Holmes P. Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer: Berlin; 1997.

Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning - Data Mining, Inference and Prediction. Springer Series in Statistics: New York; 2001.

Heath M, Scientific Computing: An Introductory Survey. McGraw-Hill: New York; 2002.

Hoon MD, Imoto S, Kobayashi K, Ogasawara N , Miyano S. Inferring gene regulatory networks from time-ordered gene expression data of Bacillus subtilis using differential equations. In: Proceedings of Pacific Symposium on Biocomputing. 2003. 17-28.

Huang S. Gene expression profiling, genetic networks and cellular states: an integrating concept for tumorigenesis and drug discovery. Journal of Molecular Medicine 1999; 77; 469-480.

Koch T. Rapid mathematical programming *PhD. Thesis*. Technical Report ZIB-TR 04-58. Technische Universität Berlin: Berlin; 2004.

Kropat E, Weber G-W, Akteke Öztürk B. Eco-Finance Networks under Uncertainty. In: Herskovits J, Canelas A, Cortes H, Aroztegui M (Eds), Proceedings of the International Conference on Engineering Optimization (ISBN 978857650156-5, CD), EngOpt 2008. Rio de Janeiro, Brazil; 2008.

Kropat E, Weber G-W, Rückmann JJ. Regression Analysis for Clusters in Gene-Environment Networks based on Ellipsoidal Calculus and Optimization. Preprint 157, Middle East Technical University, Ankara: Turkey; 2009 (submitted to Dynamics of Continuous, Discrete and Impulsive Systems).

Kropat E, Weber G-W, Sekhar PC. Regulatory networks under ellipsoidal uncertainty - optimization theory and dynamical systems. Preprint No. 138. Institute of Applied Mathematics: METU; 2009 (submitted to SIAM Journal on Optimization (SIOPT)).

Kurzhanski AB, Vályi I. Ellipsoidal Calculus for Estimation and Control. Birkhäuser; 1997.

Kurzhanski AA, Varaiya P. Ellipsoidal Toolbox Manual. EECS Department, University of California, Berkeley; 2008.

Kyoto (1997a), Kyoto Contract,
`http://www.unfccc.org/resource/convkp.html`.

Kyoto (1997b), Kyoto Protocol,
`http://en.wikipedia.org/wiki/Kyoto-Protocol`
(state: September 22, 2007).

Kyoto (1997c), Kyoto Protocol,
`http://vitalgraphics.grida.no/kyoto`.

Li YF, Venkatesh S, Li D. Modeling global emissions and residues of pesticided. Environmental Modeling and Assessment 2004; 9; 237-243.

Lozovanu D, Pickl S, Weber G-W. Optimization, monotonicity and the determination of Nash equilibria - An algorithmic analysis. (Invited paper) Computing Anticipatory Systems: CASYS'03 - Sixth International Conference, AIP Conference Proceedings 2004; 718; 351-361.

Özcan S, Yıldırım V, Kaya L, Becher D, Hecker M, Özcengiz G. Phanerochaete chrysoporium proteome and a large scale study of heavy metal response. In: HIBIT - Proceedings of International Symposium on Health Informatics and Bioinformatics, Turkey'05, Antalya: Turkey; 2005. 108-114.

Özceylan E, Paksoy T, del Rosario E, Kropat E, Weber G-W. A review on the state of the energy sector of Turkey from the perspective of Operational Research - an invitation by OR -, to appear in the proceedings of PCO 2010, 3rd Global Conference on Power Control and Optimization, February 2-4, 2010, Gold Coast, Queensland: Australia.

Özöğür S, Sağdıçoğlu Celep AG, Karasözen B, Yıldırım N, Weber G-W, Dynamical modelling of enzymatic reactions, simulation and parameter estimation with genetic algorithms. In: HIBIT - Proceedings of International Symposium on Health Informatics and Bioinformatics, Turkey'05, Antalya: Turkey; 2005. 78-84.

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

Pickl S. Convex games and feasible sets in control theory. Mathematical Methods of Operations Research 2001; 53; 1; 51-66.

Pickl S. An iterative solution to the nonlinear time-discrete TEM model - The occurence of chaos and a control theoretic algorithmic approach. AIP Conference Proceedings 2002; 627; 1; 196-205.

Pickl SW, Weber G-W. Optimization of a time-discrete nonlinear dynamical system from a problem of ecology - an analytical and numerical approach. Journal of Computational Technologies 2001; 6; 1; 43-52.

Pickl SW, Weber G-W. Optimal control of heating processes with special emphasis on earth warming. In: Chamoni P, Leisten R, Martin A, Minnemann J, Stadler

H (Eds), Operations Research Proceedings 2001. Springer Verlag: Heidelberg, New York; 2002. 247-254.

Ros L, Sabater A, Thomas F. An ellipsoidal calculus based on propagation and fusion. IEEE Transactions on Systems. Man and Cybernetics, Part B: Cybernetics 2002; 32; 4; 430-442.

Rückmann JJ, Gómez JA. On generalized semi-infinite programming. (invited paper), TOP 2006; 14; 1; 57-59.

Sakamoto E, Iba H. Inferring a system of differential equations for a gene regulatory network by using genetic programming. In: Proc. Congress on Evolutionary Computation 2001; 720-726.

Stein O. Bi-level Strategies in Semi-infinite Programming. Kluwer Academic Publishers: Boston; 2003.

Stojaković M. Imprecise probability and application in finance. 14th International Congress on Computational and Applied Mathematics (ICCAM 2009), Antalya, Turkey; 2009.

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

Taştan M, Ergenç T, Pickl SW, Weber G-W. 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. 67-75.

Taştan M, Pickl SW, Weber G-W. Mathematical modeling and stability analysis of gene-expression patterns in an extended space and with Runge-Kutta discretization. In: Proceedings of Operations Research. Bremen; 2006. 443-450.

Taylan P, Weber G-W, Beck A. New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology. In: (special issue - in honour of Prof. Dr. Alexander Rubinov), Burachik B, Yang X (Guest Eds), Optimization 2007; 56; 5-6; 1-24.

Uğur Ö, Pickl SW, Weber G-W, Wünschiers R. An algorithmic approach to analyze genetic networks and biological energy production: an introduction and contribution where OR meets biology. Optimization 2009; 58; 1; 1-22.

Uğur Ö, Weber G-W. Optimization and dynamics of gene-environment networks with intervals. (special issue - in honour of Prof. Dr. Alexander Rubinov) Journal of Industrial Management and Optimization(IJOR) 2007; 3; 2; 357-379.

Vandenberghe L, Boyd S. Semidefinite programming. SIAM Review 1996; 38; 1; 49-95.

Weber G-W. Charakterisierung struktureller Stabilität in der nichtlinearen Optimierung. In: Bock HH, Jongen HT, Plesken W (eds.) Aachener Beiträge zur Mathematik 5, Augustinus publishing house (now: Mainz publishing house) Aachen; 1992.

Weber, G-W. Generalized Semi-Infinite Optimization and Related Topics. Research and Exposition in Mathematics 29, Hofmannn KH, Wille R (Eds). Heldermann publishing house: Lemgo; 2003.

Weber G-W, Alparslan Gök SZ, Dikmen N. Environmental and life sciences: Gene-environment networks-optimization, games and control-a survey on recent achievements. In: D. DeTombe(Guest Ed), special issue of Journal of Organizational Transformation and Social Change, vol. 5, no. 3. 2008. 197-233.

Weber G-W, Alparslan-Gök SZ, Söyler B. A new mathematical approach in environmental and life sciences: gene-environment networks and their dynamics. Environmental Modeling & Assessment 2009; 14; 2; 267-288.

Weber G-W, Kropat E, Akteke-Öztürk B, Görgülü ZK. A Survey on OR and Mathematical Methods Applied on Gene-Environment Networks. Central European Journal of Operations Research 2009; 17; 3; 315-341. Special Issue on "Innovative Approaches for Decision Analysis in Energy, Health, and Life Sciences" at the occasion of EURO XXII 2007 (Prague, Czech Republic, July 8-11, 2007), Dlouhy M, Pickl S, Rauner M, Leopold-Wildburger U (guest eds.), DOI: 10.1007/s10100-009-0092-4.

Weber G-W, Taylan P, Alparslan Gök SZ, Özöğür S, Akteke Öztürk B. Optimization of gene-environment networks in the presence of errors and uncertainty with Chebychev approximation. TOP 2008; 16; 2; 284-318.

Weber G-W, Tezel A. On generalized semi-infinite optimization of genetic networks. TOP, the Operational Research Journal of SEIO (Spanish Statistics and Operations Research Society) 2007; 15; 1; 65-77.

Weber G-W, Tezel A, Taylan P, Soyler A, Çetin M. Mathematical contributions to dynamics and optimization of gene-environment networks. In: Pallaschke D,

Stein O (Guest Eds), special issue (in Celebration of Prof. Dr. Hubertus Th. Jongen's 60th Birthday). Optimization 2008; 57; 2; 353 - 377.

Weber G-W, Uğur Ö. Optimizing gene-environment networks: generalized semi-infinite programming approach with intervals. In: Proceedings of International Symposium on Health Informatics and Bioinformatics Turkey '07, HIBIT, Antalya, Turkey, April 30 - May 2 (2007)
`http://hibit.ii.metu.edu.tr/07/index.html`.

Weber G-W, Uğur Ö, Taylan P, Tezel A. On optimization, dynamics and uncertainty: a tutorial for gene-environment networks. In: (special issue) Networks in Computational Biology. Discrete Applied Mathematics 2009; 157; 10; 2494-2513.

Wolkowicz H, Saigal R, Vandenberghe L. Handbook of Semidefinite Programming: Theory, Algorithms, and Applications. International series in operations research and management science; 27. Kluwer Academic Publishers: Dordrecht, The Netherlands; 2000.

Wunderling R. Paralleler und objektorientierter Simplex-Algorithmus. *PhD. Thesis*. Technical Report ZIB-TR 96-09. Technische Universitat Berlin: Berlin; 1996.

Yılmaz FB. 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: Turkey; 2004.

Yılmaz FB, Öktem H, Weber G-W. Mathematical modeling and approximation of gene expression patterns and gene networks. In: Fleuren F, den Hertog D, Kort P (Eds), Operations Research Proceedings; 2005. 280-287.