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

In Commemoration of our Teacher and Friend
Professor Alexander Rubinov

Erik Kropat\textsuperscript{1}, Gerhard-Wilhelm Weber\textsuperscript{2} and Jan-J. Rückmann\textsuperscript{2}

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

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

\textsuperscript{3}School of Mathematics
University of Birmingham, United Kingdom
E-mail: J.Ruckmann@bham.ac.uk

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


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

1 Introduction

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

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

Gene-environment networks are generally built on measurement data from microarray experiments and environmental measurements which are always effected by errors and random noise. Error intervals can capture the deviations from specific data values by imposing individual bounds on each variable. Various regression models have been developed and studied with the help of generalized Chebychev approximation and, equivalently to that approximation, semi-infinite and even generalized semi-infinite optimization [45, 46, 47, 48, 49, 50, 54, 55, 57]. However, error intervals referring to single variables do not reflect correlations of the multivariate data within specific clusters of genetic and environmental items. A natural extension is a representation of uncertainty in terms of ellipsoidal sets which are directly related to multivariate Gaussian distributions and the corresponding covariance matrices. Ellipsoids have proved to be very flexible with respect to correlations
Clustering and classification allow to identify certain clusters of data commonly exerting influence on other groups of genes and/or environmental factors. The uncertain expression values of these clusters are represented by ellipsoidal states. The multiple interactions of genetic and environmental items are determined by (affine-) linear coupling rules which define a time-discrete regulatory model. The future states and predictions of the linear model can be calculated by ellipsoidal calculus and explicit representations of these predictions in terms of centers and shape matrices can be determined by an iterative procedure. Various regression models are introduced for an estimation of the unknown parameters of the regulatory model, which heavily depend on the cluster structure. These models compare the actual measurements and the (ellipsoidal) predictions of the linear model; however, in a set-theoretic sense. Beside a notion of distance between centers of ellipsoids, nonnegative criteria functions associated with the shape matrix of the ellipsoid are used in the objective function of the regression model. The trace and the determinant are examples of such measures and they establish the basis for the definition of regression models for parameter estimation of the regulatory system. Semi-definite programming as well as conic programming and interior point methods can then be applied for solution.

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

2 Ellipsoidal Calculus

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

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

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

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

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

The eigenvectors of $\Sigma$ point in the directions of principal semiaxes of $\mathcal{E}$. The lengths of the semiaxes of the ellipsoid $\mathcal{E}(c, \Sigma)$ are given by $\sqrt{\lambda_i}$, where $\lambda_i$ are the eigenvalues of $\Sigma$ for $i = 1, \ldots, p$. The volume of the ellipsoid $\mathcal{E}(c, \Sigma)$ is given by

$$\text{vol} \mathcal{E}(c, \Sigma) = V_p \sqrt{\det(\Sigma)},$$

where $V_p$ is the volume of the unit ball in $\mathbb{R}^p$, i.e.,

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

2.2 Affine Transformations

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

2.3 Sums of $K$ Ellipsoids

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

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

and

$$P = \left( \sum_{k=1}^K \sqrt{\text{Tr} \Sigma_k} \right) \left( \sum_{k=1}^K \frac{\Sigma_k}{\sqrt{\text{Tr} \Sigma_k}} \right).$$
2.4 Intersection of Ellipsoids

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

\[
E_\lambda(c_0, \Sigma_0) := \{ x \in \mathbb{R}^p \mid \lambda(x - c_1)^T \Sigma_1^{-1}(x - c_1) + (1 - \lambda)(x - c_2)^T \Sigma_2^{-1}(x - c_2) \leq 1 \},
\]

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

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

and

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

The ellipsoid \( E_\lambda(c_0, \Sigma_0) \) is given by the center

\[
e_0 = \mathcal{X}^{-1}(\lambda \Sigma_1^{-1} c_1 + (1 - \lambda) \Sigma_2^{-1} c_2)
\]

and shape matrix

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

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

\[
\tau(\det \mathcal{X}) \text{Tr}(\text{co}(\mathcal{X})(\Sigma_1^{-1} - \Sigma_2^{-1})) - p(\det \mathcal{X})^2 \\
\times (2e_0^T \Sigma_1^{-1} c_1 - 2e_0^T \Sigma_2^{-1} c_2 + c_0^T (\Sigma_2^{-1} - \Sigma_1^{-1}) c_0 - c_1^T \Sigma_1^{-1} c_1 + c_2^T \Sigma_2^{-1} c_2) = 0.
\]

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

\[
\tau(\det \mathcal{X})^2 \text{Tr}(\mathcal{X}^{-1}(\Sigma_1^{-1} - \Sigma_2^{-1})) - p(\det \mathcal{X})^2 \\
\times (2e_0^T \Sigma_1^{-1} c_1 - 2e_0^T \Sigma_2^{-1} c_2 + c_0^T (\Sigma_2^{-1} - \Sigma_1^{-1}) c_0 - c_1^T \Sigma_1^{-1} c_1 + c_2^T \Sigma_2^{-1} c_2) = 0.
\]

We note that it is also possible to define an inner ellipsoidal approximation. The method of finding the internal ellipsoidal approximation of the intersection of two ellipsoids is described in [44].
3 Gene-Environment Systems under Ellipsoidal Uncertainty

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

3.1 Time-Discrete Gene-Environment Systems

We consider time-discrete gene-environment regulatory systems with \( n \) genes and \( m \) environmental factors. Often, certain groups of genes and environmental items can be identified which exert a more or less regulating influence on other groups of data items. For this reason, the set of genes is divided in \( R \) disjoint clusters \( C_r \subset \{1, \ldots, n\} \), \( r = 1, \ldots, R \) in a preprocessing step of clustering. Similarly, the set of all environmental items is divided in \( S \) disjoint clusters \( D_s \subset \{1, \ldots, m\} \), \( s = 1, \ldots, S \). Here, motivated by the application in the genetic context, we focus on non-overlapping clusters and assume a strict sub-division of the variables, so that the relations 
\[
C_r \cap C_{r_2} = \emptyset \quad \text{for all} \quad r_1 \neq r_2
\]
and
\[
D_s \cap D_{s_2} = \emptyset \quad \text{for all} \quad s_1 \neq s_2
\]
are fulfilled. We note that our method is also applicable for overlapping clusters. This particular case will be addressed in a forthcoming paper [25].

There exist many clustering techniques in the fields of unsupervised learning and data mining and we do not focus on a specific method. One reasonable approach consists in nearest neighbor methods [21] in combination with the discriminant adaptive nearest neighbor metric that allow an identification of ellipsoidal neighborhoods [20]. Also clustering methods based on minimal spanning trees [8] or methods from nonsmooth optimization [7] are suited for the ellipsoidal approach.

Each cluster corresponds to a functionally related group of genes or environmental factors and the uncertain states of these cluster are represented in terms of ellipsoids 
\[
X_r = \mathcal{E}(\mu_r, \Sigma_r) \subset \mathbb{R}^{\mid C_r \mid}, \quad E_s = \mathcal{E}(\rho_s, \Pi_s) \subset \mathbb{R}^{\mid D_s \mid}.
\]

The dynamics and interactions between the various clusters of genetic and environmental items are given by the linear model
\[
\begin{align*}
X_j^{(\kappa+1)} &= \xi_{j0} + \left( \bigoplus_{r=1}^{R} A_{jr}^{GG} X_r^{(\kappa)} \bigoplus_{s=1}^{S} A_{js}^{EG} E_s^{(\kappa)} \right) \\
E_i^{(\kappa+1)} &= \zeta_{i0} + \left( \bigoplus_{r=1}^{R} A_{ir}^{GE} X_r^{(\kappa)} \bigoplus_{s=1}^{S} A_{is}^{EE} E_s^{(\kappa)} \right)
\end{align*}
\]
with \( \kappa \geq 0 \) and \( j = 1, 2, \ldots, R \), \( i = 1, 2, \ldots, S \).

In this time-discrete system, four types of cluster interactions and regulating effects are involved: (GG) gene \( \rightarrow \) gene, (EG) environment \( \rightarrow \) gene, (GE) gene \( \rightarrow \) environment, (EE) environment \( \rightarrow \) environment. The system (EC) is defined by (affine) linear coupling rules, what implies that all future states of genetic and environmental clusters are ellipsoids themselves. In particular, the sums 
\[
\bigoplus_{r=1}^{R} A_{jr}^{GG} X_r^{(\kappa)}
\]
and \( \bigoplus_{s=1}^{S} \mathcal{A}_{js}^{EG} \mathcal{E}_s^{(\kappa)} \) describe the cumulative effects of all genetic and environmental clusters exerted on the elements of cluster \( C_j \) in a set theoretic or ellipsoidal sense. In the same way, the (ellipsoidal) sums \( \bigoplus_{r=1}^{R} \mathcal{A}_{ir}^{GE} \mathcal{X}_r^{(\kappa)} \) and \( \bigoplus_{s=1}^{S} \mathcal{A}_{Es}^{EE} \mathcal{E}_s^{(\kappa)} \), 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) interaction matrices \( \mathcal{A}_{jr}^{GG} \subset \mathbb{R}^{\mid C_j \mid \times \mid C_r \mid}, \mathcal{A}_{js}^{EG} \subset \mathbb{R}^{\mid C_j \mid \times \mid D_s \mid}, \mathcal{A}_{ir}^{GE} \subset \mathbb{R}^{\mid D_i \mid \times \mid C_r \mid}, \) and \( \mathcal{A}_{Es}^{EE} \subset \mathbb{R}^{\mid D_i \mid \times \mid D_s \mid}. \) These matrices are in turn sub-matrices of the general interaction matrices \( \mathcal{A}^{GG} \in \mathbb{R}^{n \times n}, \mathcal{A}^{EG} \in \mathbb{R}^{n \times m}, \mathcal{A}^{GE} \in \mathbb{R}^{m \times n}, \mathcal{A}^{EE} \in \mathbb{R}^{m \times m}. \)

Since we assume disjoint clusters, the aforementioned sub-matrices are distinct building blocks of the general interactions matrices where the elements of the corresponding clusters define the position within the general interactions matrices. The intercepts \( \xi_0 \in \mathbb{R}^{\mid C_j \mid} \) and \( \zeta_0 \in \mathbb{R}^{\mid D_i \mid} \) are disjoint components of the vectors \( \xi_0 = (\xi_{10}, \ldots, \xi_{R0})^{T} \in \mathbb{R}^{n} \) and \( \zeta_0 = (\zeta_{10}, \ldots, \zeta_{S0})^{T} \in \mathbb{R}^{m}, \) respectively. We note that the initial values of the linear system \( (EC) \) can be defined by the first genetic and environmental measurements, i.e., \( X_j^{(0)} = \overline{X}_j^{(0)} \) and \( E_i^{(0)} = \overline{E}_i^{(0)}. \)

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

3.2 Algorithm

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

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

with center

\[
\mu_j^{(\kappa+1)} = \xi_{j0} + \sum_{r=1}^{R} \mathcal{A}_{jr}^{GG} \mu_r^{(\kappa)} + \sum_{s=1}^{S} \mathcal{A}_{js}^{EG} \rho_s^{(\kappa)}
\]
and configuration matrix

\[
\Sigma_j^{(\kappa+1)} = \left( \frac{\sqrt{\text{Tr} \, G_j^{(\kappa)}} + \sqrt{\text{Tr} \, H_j^{(\kappa)}}}{\sqrt{\text{Tr} \, G_j^{(\kappa)}}} \right) \cdot \left( \frac{\sqrt{\text{Tr} \, G_j^{(\kappa)}} + \frac{\sqrt{\text{Tr} \, H_j^{(\kappa)}}}{\sqrt{\text{Tr} \, H_j^{(\kappa)}}}}{\sqrt{\text{Tr} \, H_j^{(\kappa)}}} \right),
\]

where

\[
G_j^{(\kappa)} = \left( \sum_{r=1}^{R} \sqrt{\text{Tr} \left( A_{jr}^{GG} \Sigma_r^{(\kappa)} (A_{jr}^{GG})^T \right)} \right) \cdot \left( \sum_{r=1}^{R} \frac{A_{jr}^{GG} \Sigma_r^{(\kappa)} (A_{jr}^{GG})^T}{\sqrt{\text{Tr} \left( A_{jr}^{GG} \Sigma_r^{(\kappa)} (A_{jr}^{GG})^T \right)}} \right),
\]

\[
H_j^{(\kappa)} = \left( \sum_{s=1}^{S} \sqrt{\text{Tr} \left( A_{js}^{EG} \Pi_s^{(\kappa)} (A_{js}^{EG})^T \right)} \right) \cdot \left( \sum_{s=1}^{S} \frac{A_{js}^{EG} \Pi_s^{(\kappa)} (A_{js}^{EG})^T}{\sqrt{\text{Tr} \left( A_{js}^{EG} \Pi_s^{(\kappa)} (A_{js}^{EG})^T \right)}} \right).
\]

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

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

with center

\[
\rho_i^{(\kappa+1)} = \zeta_0 + \sum_{r=1}^{R} A_{ir}^{GE} \mu_r^{(\kappa)} + \sum_{s=1}^{S} A_{is}^{EE} \rho_s^{(\kappa)}
\]

and configuration matrix

\[
\Pi_i^{(\kappa+1)} = \left( \frac{\sqrt{\text{Tr} \, M_i^{(\kappa)}} + \sqrt{\text{Tr} \, N_i^{(\kappa)}}}{\sqrt{\text{Tr} \, M_i^{(\kappa)}}} \right) \cdot \left( \frac{\sqrt{\text{Tr} \, M_i^{(\kappa)}} + \frac{\sqrt{\text{Tr} \, N_i^{(\kappa)}}}{\sqrt{\text{Tr} \, N_i^{(\kappa)}}}}{\sqrt{\text{Tr} \, N_i^{(\kappa)}}} \right),
\]

where

\[
M_i^{(\kappa)} = \left( \sum_{r=1}^{R} \sqrt{\text{Tr} \left( A_{ir}^{GE} \Sigma_r^{(\kappa)} (A_{ir}^{GE})^T \right)} \right) \cdot \left( \sum_{r=1}^{R} \frac{A_{ir}^{GE} \Sigma_r^{(\kappa)} (A_{ir}^{GE})^T}{\sqrt{\text{Tr} \left( A_{ir}^{GE} \Sigma_r^{(\kappa)} (A_{ir}^{GE})^T \right)}} \right),
\]

\[
N_i^{(\kappa)} = \left( \sum_{s=1}^{S} \sqrt{\text{Tr} \left( A_{is}^{EE} \Pi_s^{(\kappa)} (A_{is}^{EE})^T \right)} \right) \cdot \left( \sum_{s=1}^{S} \frac{A_{is}^{EE} \Pi_s^{(\kappa)} (A_{is}^{EE})^T}{\sqrt{\text{Tr} \left( A_{is}^{EE} \Pi_s^{(\kappa)} (A_{is}^{EE})^T \right)}} \right).
\]
4 Regression Analysis Under Ellipsoidal Uncertainty

4.1 The Regression Problem

The unknown system parameters of model (EC) are given by the entries of the interaction matrices $A_{j\gamma}^{GG}$, $A_{j\alpha}^{EG}$, $A_{i\alpha}^{GE}$, and $A_{i\beta}^{EE}$ and the intercepts $\zeta_{ij0}$ and $\zeta_{ij0}^s$. For an estimation of these parameters and, thus, for regression analysis, we have to compare the predictions of the linear model (EC) and the observations from genetic and environmental measurements. The (ellipsoidal) measurements

$$\mathbf{X}_r^{(k)} = \mathcal{E}(\mathbf{p}_r^{(k)}, \mathbf{B}_r^{(k)}) \subset \mathbb{R}^{[C_r]}, \quad \mathbf{E}_s^{(k)} = \mathcal{E}(\mathbf{p}_s^{(k)}, \mathbf{B}_s^{(k)}) \subset \mathbb{R}^{[D_s]}$$

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

$$\hat{\mathbf{X}}_j^{(k+1)} = \mathcal{E}(\hat{\mathbf{p}}_j^{(k+1)}, \hat{\Sigma}_j^{(k+1)}) := \xi_{j0} + \left( R \bigoplus_{r=1}^R A_{j\gamma}^{GG} \mathbf{X}_r^{(k)} \right) + \left( S \bigoplus_{s=1}^S A_{j\alpha}^{EG} \mathbf{E}_s^{(k)} \right),$$

$$\hat{\mathbf{E}}_i^{(k+1)} = \mathcal{E}(\hat{\mathbf{p}}_i^{(k+1)}, \hat{\Pi}_i^{(k+1)}) := \zeta_{i0} + \left( R \bigoplus_{r=1}^R A_{i\alpha}^{GE} \mathbf{X}_r^{(k)} \right) + \left( S \bigoplus_{s=1}^S A_{i\beta}^{EE} \mathbf{E}_s^{(k)} \right),$$

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

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

$$\Delta \mathbf{X}_r^{(k)} := \hat{\mathbf{X}}_r^{(k)} \cap \mathbf{X}_r^{(k)} \quad \text{and} \quad \Delta \mathbf{E}_s^{(k)} := \hat{\mathbf{E}}_s^{(k)} \cap \mathbf{E}_s^{(k)}$$

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

$$\begin{equation}
(R) \quad \text{Maximize} \quad \sum_{k=1}^{T} \left\{ \sum_{r=1}^{R} \left\| \Delta \mathbf{X}_r^{(k)} \right\|_s - \left\| \hat{\mathbf{p}}_r^{(k)} - \mathbf{p}_r^{(k)} \right\|_2^2 \right. \right.
+ \left. \sum_{s=1}^{S} \left\| \Delta \mathbf{E}_s^{(k)} \right\|_s - \left\| \hat{\mathbf{p}}_s^{(k)} - \mathbf{p}_s^{(k)} \right\|_2^2 \right\}.
\end{equation}$$

Here, $\left\| \cdot \right\|_s$ denotes a measure that reflects the geometrical size of the intersections (fusions) and we assume that $\left\| \Delta \mathbf{X}_r^{(k)} \right\|_s = 0$, if $\Delta \mathbf{X}_r^{(k)} = \emptyset$ and $\left\| \Delta \mathbf{E}_s^{(k)} \right\|_s = 0$, if $\Delta \mathbf{E}_s^{(k)} = \emptyset$. There exist various measures related to the shape of the intersections, e.g., the volume (which corresponds to the ellipsoid matrix determinant) or the sum of squares of semiaxes (which corresponds to the trace of the configuration matrix).
These examples lead to specific formulations of the regression problem \((R)\) and they depend on the shape matrices of the fusions \(\Delta X^{(\kappa)}_r\) and \(\Delta E^{(\kappa)}_s\). For further details on geometrical (ellipsoidal) measures and the related regression problems we refer to [25].

For a deeper analysis of the regression problem \((R)\), explicit representations of the fusions \(\Delta X^{(\kappa)}_r\) and \(\Delta E^{(\kappa)}_s\) are required. The fusion \(\Delta X^{(\kappa)}_r = \hat{X}^{(\kappa)}_r \cap \bar{X}^{(\kappa)}_r\) is an ellipsoid \(E(\Delta \mu^{(\kappa)}_r, \Delta \Sigma^{(\kappa)}_r)\) with center

\[
\Delta \mu^{(\kappa)}_r = \left[\lambda \hat{\Sigma}^{(\kappa)}_r \right]^{-1} \left[\lambda \hat{\Sigma}^{(\kappa)}_r \right]^{-1} \hat{\mu}^{(\kappa)}_r + (1 - \lambda) \left[\Sigma^{(\kappa)}_r \right]^{-1} \mu^{(\kappa)}_r,
\]

and shape matrix

\[
\Delta \Sigma^{(\kappa)}_r = \xi^{(\kappa)}_r \left[\lambda \hat{\Sigma}^{(\kappa)}_r \right]^{-1},
\]

where

\[
\hat{X}^{(\kappa)}_r := \lambda \hat{\Sigma}^{(\kappa)}_r \left[\lambda \hat{\Sigma}^{(\kappa)}_r \right]^{-1} + (1 - \lambda) \left[\Sigma^{(\kappa)}_r \right]^{-1}
\]

and

\[
\xi^{(\kappa)}_r := 1 - \lambda (1 - \lambda) (\bar{\mu}^{(\kappa)}_r - \hat{\mu}^{(\kappa)}_r)^T \left[\lambda \hat{\Sigma}^{(\kappa)}_r \right]^{-1} \left[\lambda \hat{\Sigma}^{(\kappa)}_r \right]^{-1} (\bar{\mu}^{(\kappa)}_r - \hat{\mu}^{(\kappa)}_r).
\]
The parameter $\lambda$ is the only root in $(0, 1)$ of the following polynomial of degree $2\lvert C_r \rvert - 1$:

$$\xi_r^{(s)} (\det \mathcal{X}_r^{(s)})^{-2} \text{Tr} \left( [\mathcal{X}_r^{(s)}]^{-1} \left( \mathcal{S}_r^{(s)} [\mathcal{X}_r^{(s)}]^{-1} - [\mathcal{S}_r^{(s)}]^{-1} \right) \right) - \lvert C_r \rvert (\det \mathcal{X}_r^{(s)})^{-2} \times \left( 2 [\Delta \rho_r^{(s)}] T [\mathcal{S}_r^{(s)}]^{-1} \tilde{\rho}_r^{(s)} - 2 [\Delta \rho_r^{(s)}] T [\mathcal{S}_r^{(s)}]^{-1} \hat{\rho}_r^{(s)} \right)

[\Delta \rho_r^{(s)}] T \left( [\mathcal{S}_r^{(s)}]^{-1} \mathcal{S}_r^{(s)} - [\mathcal{S}_r^{(s)}]^{-1} \right) [\Delta \rho_r^{(s)}] T [\mathcal{S}_r^{(s)}]^{-1} \hat{\rho}_r^{(s)} + [\rho_r^{(s)}] T [\mathcal{S}_r^{(s)}]^{-1} \rho_r^{(s)} = 0.

Similarly, the fusion $\Delta E_s^{(s)} = \tilde{E}_s^{(s)} \cap \hat{E}_s^{(s)}$ is an ellipsoid $\mathcal{E}(\Delta \rho_s^{(s)}, \Delta \Pi_s^{(s)})$ with center

$$\Delta \rho_s^{(s)} = [\eta_s^{(s)}]^{-1} \left( \lambda [\hat{\Pi}_s^{(s)}]^{-1} \tilde{\rho}_s^{(s)} + (1 - \lambda) [\Pi_s^{(s)}]^{-1} \hat{\rho}_s^{(s)} \right)$$

and shape matrix

$$\Delta \Pi_s^{(s)} = \eta_s^{(s)} [\eta_s^{(s)}]^{-1},$$

where

$$\eta_s^{(s)} := 1 - \lambda (1 - \lambda) (\hat{\rho}_s^{(s)} - \tilde{\rho}_s^{(s)}) T [\Pi_s^{(s)}]^{-1} [\eta_s^{(s)}]^{-1} [\Pi_s^{(s)}]^{-1} (\tilde{\rho}_s^{(s)} - \hat{\rho}_s^{(s)}).$$

The parameter $\lambda$ is the only root in $(0, 1)$ of the following polynomial of degree $2\lvert D_s \rvert - 1$:

$$\eta_s^{(s)} (\det \mathcal{Y}_s^{(s)})^{-2} \text{Tr} \left( [\mathcal{Y}_s^{(s)}]^{-1} \left( \mathcal{H}_s^{(s)} [\mathcal{Y}_s^{(s)}]^{-1} - [\mathcal{H}_s^{(s)}]^{-1} \right) \right) - \lvert D_s \rvert (\det \mathcal{Y}_s^{(s)})^{-2} \times \left( 2 [\Delta \rho_s^{(s)}] T [\mathcal{H}_s^{(s)}]^{-1} \tilde{\rho}_s^{(s)} - 2 [\Delta \rho_s^{(s)}] T [\mathcal{H}_s^{(s)}]^{-1} \hat{\rho}_s^{(s)} \right)

[\Delta \rho_s^{(s)}] T \left( [\mathcal{H}_s^{(s)}]^{-1} \mathcal{H}_s^{(s)} - [\mathcal{H}_s^{(s)}]^{-1} \right) [\Delta \rho_s^{(s)}] T [\mathcal{H}_s^{(s)}]^{-1} \hat{\rho}_s^{(s)} + [\rho_s^{(s)}] T [\mathcal{H}_s^{(s)}]^{-1} \rho_s^{(s)} = 0.

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

(a) the trace of $Q$,

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

where $\lambda_i$ are the eigenvalues of $Q$ (i.e., Tr $Q$ is equal to the sum of the squares of the semiaxes),
(b) the determinant of $Q$,

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

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

$$
\text{vol } E(0, Q) = \pi^{\frac{p}{2}} (\det Q)^{\frac{1}{2}} (\Gamma\left(\frac{p}{2} + 1\right))^{-1}
$$

of the ellipsoid, where $\Gamma$ stands for the Gamma-function.

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

### 4.2 The Trace Criterion

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

$$
(R_{Tr}) \quad \text{Maximize } \sum_{\kappa=1}^{T} \left\{ \sum_{r=1}^{R} \left[ \text{Tr} (\Delta \Sigma_r^{(\kappa)}) - \sum_{j=1}^{[C_r]} (\hat{\mu}_{r,j}^{(\kappa)} - \mu_{r,j}^{(\kappa)})^2 \right] \\
+ \sum_{s=1}^{S} \left[ \text{Tr} (\Delta \Pi_s^{(\kappa)}) - \sum_{i=1}^{[D_s]} (\hat{\rho}_{s,i}^{(\kappa)} - \rho_{s,i}^{(\kappa)})^2 \right] \right\}.
$$

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

$$
(R'_{Tr}) \quad \text{Maximize } \sum_{\kappa=1}^{T} \left\{ \sum_{r=1}^{R} \sum_{j=1}^{[C_r]} \lambda_{r,j}^{(\kappa)} - (\hat{\mu}_{r,j}^{(\kappa)} - \mu_{r,j}^{(\kappa)})^2 \\
+ \sum_{s=1}^{S} \sum_{i=1}^{[D_s]} \Lambda_{s,i}^{(\kappa)} - (\hat{\rho}_{s,i}^{(\kappa)} - \rho_{s,i}^{(\kappa)})^2 \right\},
$$

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

### 4.3 The Determinant Criterion

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

$$
(R_{Det}) \quad \text{Maximize } \sum_{\kappa=1}^{T} \left\{ \sum_{r=1}^{R} \left[ \det (\Delta \Sigma_r^{(\kappa)}) - \sum_{j=1}^{[C_r]} (\hat{\mu}_{r,j}^{(\kappa)} - \mu_{r,j}^{(\kappa)})^2 \right] \\
+ \sum_{s=1}^{S} \left[ \det (\Delta \Pi_s^{(\kappa)}) - \sum_{i=1}^{[D_s]} (\hat{\rho}_{s,i}^{(\kappa)} - \rho_{s,i}^{(\kappa)})^2 \right] \right\}.
$$
Equivalent formulations of \((R_{Det})\) can be given in terms of the eigenvalues of the configuration matrices

\[ (R'_{Det}) \text{ Maximize } \sum_{\kappa=1}^{T} \left\{ \sum_{r=1}^{R} \left[ \left( \prod_{j=1}^{|C_r|} \lambda_{r,j}^{(\kappa)} \right) - \sum_{j=1}^{|C_r|} (\mu_{r,j}^{(\kappa)} - \bar{\mu}_{r,j}^{(\kappa)})^2 \right] \right. \]

\[ \left. + \sum_{s=1}^{S} \left[ \left( \prod_{i=1}^{|D_s|} \Lambda_{s,i}^{(\kappa)} \right) - \sum_{i=1}^{|D_s|} (\rho_{s,i}^{(\kappa)} - \bar{\rho}_{s,i}^{(\kappa)})^2 \right] \right\}, \]

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

\[ (R''_{Det}) \text{ Maximize } \sum_{\kappa=1}^{T} \left\{ \sum_{r=1}^{R} \left[ \pi \frac{1}{2} \Gamma \left( \frac{|C_r|}{2} \right) \text{vol} (\Delta X_r^{(\kappa)}) \right]^2 \right. \]

\[ \left. - \sum_{j=1}^{|C_r|} (\mu_{r,j}^{(\kappa)} - \bar{\mu}_{r,j}^{(\kappa)})^2 \right) \]

\[ \left. + \sum_{s=1}^{S} \left[ \pi \frac{1}{2} \Gamma \left( \frac{|D_s|}{2} \right) \text{vol} (\Delta E_s^{(\kappa)}) \right]^2 \right) \]

\[ - \sum_{i=1}^{|D_s|} (\rho_{s,i}^{(\kappa)} - \bar{\rho}_{s,i}^{(\kappa)})^2 \right\}. \]

5 Optimization Methods

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

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

is positive semidefinite representable \([11], p. 81\). For this reason, we introduce the regression model
\[(\tilde{R}_{\text{Det}}) \quad \text{Maximize} \quad \sum_{k=1}^{T} \left\{ - \sum_{r=1}^{R} \left[ \det^p \left( \Delta \Sigma_r^{(k)} \right) + \sum_{j=1}^{[C_r]} (\tilde{\mu}_{r,j}^{(k)} - \tilde{\mu}_{r,j}^{(k)})^2 \right] \\
\quad - \sum_{s=1}^{S} \left[ \det^q \left( \Delta \Pi_s^{(k)} \right) + \sum_{i=1}^{[D_s]} (\tilde{\rho}_{s,i}^{(k)} - \tilde{\rho}_{s,i}^{(k)})^2 \right] \right\}, \]

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

\[(\tilde{R}'_{\text{Det}}) \quad \text{Maximize} \quad \sum_{k=1}^{T} \left\{ - \sum_{r=1}^{R} \left[ \left( \prod_{j=1}^{[C_r]} \lambda_{r,j}^{(k)} \right)^p + \sum_{j=1}^{[C_r]} (\tilde{\mu}_{r,j}^{(k)} - \tilde{\mu}_{r,j}^{(k)})^2 \right] \\
\quad - \sum_{s=1}^{S} \left[ \left( \prod_{i=1}^{[D_s]} \lambda_{s,i}^{(k)} \right)^q + \sum_{i=1}^{[D_s]} (\tilde{\rho}_{s,i}^{(k)} - \tilde{\rho}_{s,i}^{(k)})^2 \right] \right\}, \]

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

\[(\tilde{R}''_{\text{Det}}) \quad \text{Maximize} \quad \sum_{k=1}^{T} \left\{ \sum_{r=1}^{R} \left[ \left( \prod_{j=1}^{[C_r]} \lambda_{r,j}^{(k)} \right)^p \Gamma \left( \frac{[C_r]}{2} + 1 \right) \right] \right\}^{2p} \\
\quad - \sum_{j=1}^{[C_r]} (\tilde{\mu}_{r,j}^{(k)} - \tilde{\mu}_{r,j}^{(k)})^2 \\
\quad + \sum_{s=1}^{S} \left[ \left( \prod_{i=1}^{[D_s]} \lambda_{s,i}^{(k)} \right)^q \Gamma \left( \frac{[D_s]}{2} + 1 \right) \right]^{2q} \\
\quad - \sum_{i=1}^{[D_s]} (\tilde{\rho}_{s,i}^{(k)} - \tilde{\rho}_{s,i}^{(k)})^2 \right\}. \]

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

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

of the symmetric \(n \times n\)-matrix \(M\) is positive semidefinite representable ([111], p. 83). Here, \(M \succ 0\) means that \(M\) is positive semidefinite. Now, with two positive rationals \(p, q\) we obtain a further regression model

\[(R''_{\text{Det}}) \quad \text{Maximize} \quad \sum_{k=1}^{T} \left\{ \sum_{r=1}^{R} \left[ \det^{-p} \left( \Delta \Sigma_r^{(k)} \right) + \sum_{j=1}^{[C_r]} (\tilde{\mu}_{r,j}^{(k)} - \tilde{\mu}_{r,j}^{(k)})^2 \right] \\
\quad + \sum_{s=1}^{S} \left[ \det^{-q} \left( \Delta \Pi_s^{(k)} \right) + \sum_{i=1}^{[D_s]} (\tilde{\rho}_{s,i}^{(k)} - \tilde{\rho}_{s,i}^{(k)})^2 \right] \right\}. \]
In the regression model \((R'_{\tau,\tau})\), sums of all eigenvalues of the configuration matrices \(\Delta\Sigma^{(c)}\) and \(\Delta\Pi^{(c)}\) are considered, which can also be regarded as positive semidefinite representable functions ([11], p. 80). In general, interior point methods can be applied which have a moderate complexity [29, 30, 31, 33]. Alternatively, for regression problems that consider sums of eigenvalues in the objective function, associated bilevel problems can be introduced which could be solved by gradient methods. In fact, in [32] structural frontiers of conic programming are discussed with other optimization methods compared, and future applications in machine learning and data mining prepared. However, we would like to underline that in the areas regression and classification of statistical learning (cf. e.g., [21, 37]), our optimization based methods provided and further promise very good and competitive results [15, 22, 41, 52, 58].

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

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

### 6 Conclusion

In this paper, we analyzed a time-discrete regression model for gene-environment regulatory networks under ellipsoidal uncertainty. The ellipsoidal approach - complemented by powerful optimization methods of semidefinite programming and the efficiency of interior point methods - offers a new perspective for the analysis of regulatory systems in systems biology and life sciences, and it could even establish an Ellipsoidal OR. The representation of the dynamic states of the underlying system in terms of ellipsoids was motivated by our and our colleagues’ studies on gene-environment networks and eco-finance networks, where errors and uncertainty are modeled by intervals [58, 59, 60]. Here, we further extended this approach by replacing the intervals by more general uncertainty sets. Ellipsoids are more flexible than intervals and parallelepipeds what also refers to stochastic dependencies between the various genes and environmental factors. Moreover, ellipsoids are directly related to covariance matrices and they provide good approximations of convex sets. In particular, models based on Gaussian random noise refer to the ellipsoidal approach. However, Gaussian random distributions are often used as simplifications,
and in many applications non-Gaussian probability distributions have to be applied. In future works, we will further extend our models based on ellipsoidal calculus and we will turn to a more set-theoretic representation of uncertainty based on semi-algebraic sets [9, 12]. We will combine this new perception with refined optimization methods, and by this we will offer a further avenue for the analysis of regulatory systems, particularly with regard to applications and real-world data. Furthermore, we refined collaborative game theory under interval uncertainty [1, 2, 3, 4] by our ellipsoidal calculus [5], herewith including correlations, as being expressed by sub-coalitions found by clustering. In terms if, e.g., joint implementation, collaborative game theory serves for a new interpretation of the TEM model, which is some dynamical model related with the Kyoto Protocol and a part of our gene-environment networks [35, 48, 49]. With this study, we aim scientific advances and the improvement of living conditions on earth.

References


Regression Analysis for Clusters in Gene-Environment Networks


Regression Analysis for Clusters in Gene-Environment Networks


