FOUNDATIONS OF SEMIALGEBRAIC GENE-ENVIRONMENT NETWORKS

Erik Kropat

1Christerweg, 12, 83624 Otterfing, Germany

Gerhard-Wilhelm Weber and Erfan Babaee Tirkolaee

2Faculty of Engineering Management, Poznan University of Technology, Poznan, Poland
3Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey
4Department of Industrial Engineering, Mazandaran University of Science and Technology, Babol, Iran
5Department of Industrial and Systems Engineering, Istinye University, Istanbul, Turkey

(Communicated by Bruno Miguel Paz Mendes de Oliveira)

Abstract. Gene-environment network studies rely on data originating from different disciplines such as chemistry, biology, psychology or social sciences. Sophisticated regulatory models are required for a deeper investigation of the unknown and hidden functional relationships between genetic and environmental factors. At the same time, various kinds of uncertainty can arise and interfere with the system’s evolution. The aim of this study is to go beyond traditional stochastic approaches and to propose a novel framework of semi-algebraic gene-environment networks. Foundation is laid for future research, methodology and application. This approach is a natural extension of interconnected systems based on stochastic, polyhedral, ellipsoidal or fuzzy (linguistic) uncertainty. It allows for a reconstruction of the underlying network from uncertain (semialgebraic) data sets and for a prediction of the uncertain futures states of the system. In addition, aspects of network pruning for large regulatory systems in genome-wide studies are discussed leading to mixed-integer programming (MIP) and continuous programming.

1. Introduction. Gene-environment networks provide a mathematical framework for the analysis and investigation of regulatory networks in genetics, bioinformatics and medicine [21]. The identification of the underlying hidden mechanisms and pathways allows for a deeper understanding of the system under consideration. Typically, different types of environmental factors play a significant role in the internal organisation and development of complex interconnected systems. In particular, environmental variation can be caused by measurable chemical, physical and biological parameters as well as life events or behavioral patterns often expressed in linguistic terms. Data uncertainty associated with such measurements has to be addressed in a sophisticated mathematical model.

2020 Mathematics Subject Classification. Primary: 90C35, 93A30, 92D10, 14P10; Secondary: 62J02.

Key words and phrases. Gene-environment networks, semialgebraic regression, semialgebraic uncertainty, geometric model, forecasting, uncertainty modeling, mixed-integer programming, continuous programming.

* Corresponding author: Erik Kropat.
It is a major advantage of the mathematical network model that it allows for an investigation of the system by graph theoretic approaches and it supports research and analysis in various ways. Both genetic and environmental factors contribute to the etiology of many complex diseases. For this reason, an important application consists in the search for novel genes related to diseases [47]. Other studies put a particular focus on risk factors and pathways for the development of diseases such as depression [44, 45] or multiple sclerosis [8]. Such studies have a potential relevance for clinical application. For example, better prognostic models in pharmacogenetics may rely on genotype information and can lead to a more precise prognosis, targeted intervention and personalized treatment [23]. In particular, research can lead to new therapeutic approaches to cognitive and affective disorders in neuropharmacology [31] and precision medicine in case of Alzheimer’s disease and other kinds of dementia [10].

1.1. Gene-environment network models under uncertainty. In previous studies, we addressed various mathematical gene-environment network models for many different types of data uncertainty [53, 54, 55]. Metric data interval arithmetics as well as polyhedral and ellipsoidal uncertainty regions [17, 18]. In case of linguistic variables such as life events or behavioral traits fuzzy gene-environment networks can be applied [15, 16]. In addition, models based on stochastic differential equations [37] and spline regression approaches [25, 26, 27] have been discussed previously.

1.2. Semialgebraic uncertainty regions. In this study, we propose a new framework of semialgebraic gene-environment networks. Semialgebraic uncertainty regions for measurements and predictions provide a higher flexibility for the design of regulatory models. They aim at an improvement of accuracy and precision in the reconstruction of the network and the prediction of future states. In particular, this framework includes and further extends polyhedral and ellipsoidal regression approaches [17, 18, 55].

1.3. Genome-wide association studies. In recent years, genome-wide association studies aimed at an identification of the loci associated with complex diseases [3, 22, 34]. From a mathematical point of view, very large gene-environment networks arise. In order to reduce the overall complexity and to avoid the curse of dimensionality, methods for network pruning are required. We derive two distinct approaches for gene-environment networks with semialgebraic uncertainty sets. These methods eventually lead to regression approaches based on mixed-integer programming (MIP) and continuous programming.

1.4. Outline. The paper is organized as follows: In Section 2, the generic regulatory model for gene-environment interaction is introduced. Gene-environment networks for clusters of genes and environmental factors are presented in Section 3. Then, in Section 4, we review some basic facts about semialgebraic sets and semialgebraic functions. Section 5 turns to regulatory models based on semialgebraic uncertainty regions. The following Section 6 states some basic assumptions about the required semialgebraic measurements. Thereafter, Section 7 introduces the regression model for semialgebraic gene-environment networks.
Several forms of network pruning for very large datasets are discussed in Section 8. The corresponding regression problems can be tackled by strategies from MIP and continuous programming. Selected classes of optimization problems and associated algorithmic techniques are presented in Section 9. Finally, we conclude with an outlook on future directions of research along with areas of real-world application problems.

2. Gene-environment networks. Gene-environment regulatory networks distinguish two major groups of entities: a) the genes and b) the environmental factors. The environmental factors modulate the system of genes by external conditions acting as regulatory components, additional controls or disturbances of the genes and the regulatory system itself. The interconnection between the single genes and environmental factors is encoded in the adjacency matrix $A$. In addition, the functional relationship between genes and/or environmental factors has to be integrated into the modeling of the gene-environment network.

**Definition 2.1.** A gene-environment network $(\text{GEN}_\theta) = (G, E, A_{\text{entities}}, F(\theta))$ consists of

i) the genes $G = (G_1, \ldots, G_n)^T \in \mathbb{R}^n$ and the environmental factors $E = (E_1, \ldots, E_m)^T \in \mathbb{R}^m$,

ii) an adjacency matrix $A_{\text{entities}} \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m}$, indicating the (directed) connections between genes and/or environmental factors, where

$$A_{\text{entities}}(i, j) = A_{\text{entities}}(j, i) = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are connected}, \\
0, & \text{otherwise}, \end{cases}$$

in the case of undirected branches and

$$A_{\text{entities}}(i, j) = \begin{cases} 1, & \text{if } i \text{ is connected to } j \text{ by a directed edge}, \\
-1, & \text{if } j \text{ is connected to } i \text{ by a directed edge}, \\
0, & \text{otherwise}, \end{cases}$$

in the case of directed edges,

iii) a functional relationship

$$F(\theta, \cdot) = (F_1(\theta, \cdot), \ldots, F_{n+m}(\theta, \cdot))^T : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$$

between genes and/or environmental factors, where $F_i(\theta, \cdot) : \mathbb{R}^{n+m} \to \mathbb{R}$ for $i = 1, \ldots, n + m$. Here, $\theta \subset \Theta$ is the unknown parameter vector and $\Theta \subset \mathbb{R}^{\ell}$, $\ell \in \mathbb{N}$, denotes the parameter space.

Figure 1 shows the various interactions between genes and environmental factors of a target-environment network.

3. Gene-environment networks for clustered partitions. The regulatory network consists of $n$ genes $G = (G_1, \ldots, G_n)^T \in \mathbb{R}^n$ and $m$ environmental factors $E = (E_1, \ldots, E_m)^T \in \mathbb{R}^m$. In gene-environment models, functionally related groups of genes or environmental factors often take influence on the interactions between its components. For this reason, the set of genes, $G$, is divided into $R \in \mathbb{N}$ genetic clusters and the set of all environmental items, $E$, is split into $S \in \mathbb{N}$ environmental clusters. The pair $(\mathcal{C}, \mathcal{D})$ with $\mathcal{C} = \{C_1, \ldots, C_R\}$, $\mathcal{D} = \{D_1, \ldots, D_R\}$ and $C_r \subset \{1, \ldots, n\}$, $D_s \subset \{1, \ldots, m\}$ is called a cluster partition.
Here, \((\mathcal{C}, \mathcal{D})\) is a disjoint cluster partition if \(C_{r_1} \cap C_{r_2} = \emptyset\) for all \(r_1, r_2 \in \{1, \ldots, R\}\) with \(r_1 \neq r_2\) and \(D_{s_1} \cap D_{s_2} = \emptyset\) for all \(s_1, s_2 \in \{1, \ldots, S\}\) with \(s_1 \neq s_2\). Otherwise, the cluster partition \((\mathcal{C}, \mathcal{D})\) is called overlapping. The (crisp) states of the elements of each cluster are given by subsets of the vectors \(G\) and \(E\). For each \(r \in \{1, \ldots, R\}\) the \(|C_r|\)-subvector \(G_r\) of \(G\) given by the indices of \(C_r\) is a cluster vector of the genetic cluster \(C_r\). For each \(s \in \{1, \ldots, S\}\) the \(|D_s|\)-subvector \(E_s\) of \(E\) defined by the indices of \(D_s\) is a cluster vector of the environmental cluster \(D_s\). After these preparations, the notion of a gene-environment network with cluster partition \((\mathcal{C}, \mathcal{D})\) can be introduced.

**Definition 3.1.** A gene-environment network with cluster structure \((\mathcal{C}, \mathcal{D})\) is denoted by \((\text{GEN-}C_{\theta}) = (G, E, \mathcal{C}, \mathcal{D}, A_{\text{clusters}}, H(\theta_1), K(\theta_2))\) consists of

i) the vector \(G = (G_1, \ldots, G_n)^T \in \mathbb{R}^n\) of the \(n\) genes and the vector \(E = (E_1, \ldots, E_m)^T \in \mathbb{R}^m\) of the \(m\) environmental factors,

ii) the cluster partition \((\mathcal{C}, \mathcal{D})\) with the set of genetic clusters \(\mathcal{C} = \{C_1, \ldots, C_R\}\) and the set of environmental clusters \(\mathcal{D} = \{D_1, \ldots, D_S\}\),

iv) the adjacency matrix \(A_{\text{clusters}} \in \mathbb{R}^{R+S} \times \mathbb{R}^{R+S}\), indicating the (directed) connections between target clusters and/or environmental clusters

v) the parameter-dependent functional relationships:

\[
H(\theta_1, \cdot) = (H_1(\theta_1, \cdot), \ldots, H_R(\theta_1, \cdot))^T, \\
K(\theta_2, \cdot) = (K_1(\theta_2, \cdot), \ldots, K_S(\theta_2, \cdot))^T, 
\]

where

\[
H_r(\theta_1, \cdot) : \left( \bigtimes_{r=1}^R \mathbb{R}^{|C_r|} \right) \times \left( \bigtimes_{s=1}^S \mathbb{R}^{|D_s|} \right) \to \mathbb{R}^{|C_r|}, \\
K_s(\theta_2, \cdot) : \left( \bigtimes_{r=1}^R \mathbb{R}^{|C_r|} \right) \times \left( \bigtimes_{s=1}^S \mathbb{R}^{|D_s|} \right) \to \mathbb{R}^{|D_s|}
\]

are set-valued functions depending on the cluster partitions of genetic clusters and/or environmental clusters. Here, \((\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 \subset \mathbb{R}^{\ell_1} \times \mathbb{R}^{\ell_2}, \ell_1, \ell_2 \in \mathbb{N}\), denotes the unknown parameter vector.
Figure 2 summarizes the interactions between genetic clusters and/or environmental clusters.

**Figure 2.** Gene-environment networks with cluster partition: Interactions between genetic clusters (red), environmental clusters and genetic clusters (green) and between environmental clusters (blue).

4. **Semialgebraic uncertainty regions.** In this study, semialgebraic sets represent the uncertain states of each genetic and environmental cluster. We recall that a semi-algebraic set in $\mathbb{R}^p$, $p \in \mathbb{N}$, has the form

$$S = \bigcup_{i=1}^{s} \bigcap_{j=1}^{r_i} \{ x \in \mathbb{R}^p \mid f_{i,j} \ast_{i,j} 0 \},$$

where $f_{i,j} \in \mathbb{R}[X_1, \ldots, X_p]$ and $\ast_{i,j} \in \{>, <, =\}$ for $i = 1, \ldots, s$, and $j = 1, \ldots, r_i$. The set of semialgebraic sets in $\mathbb{R}^p$ is denoted by $\mathcal{S}^p$. The family of semialgebraic sets in $\mathbb{R}^p$ is closed with respect to finite intersections, finite unions and complements. For this reason, they allow us to introduce a semi-algebraic calculus on semialgebraic sets. The sign conditions $f(x) > 0$, $f(x) < 0$ or $f(x) = 0$ of a polynomial $f$ imply that a semialgebraic set $S \subset \mathbb{R}^p$ is obtained by a finite boolean combination of a finite number of polynomials (i.e., by disjunction, conjunction and negation; see [5]). Let $A \subset \mathbb{R}^p$ and $B \subset \mathbb{R}^q$ be semialgebraic sets, where $p, q \in \mathbb{N}$. A mapping $f : A \to B$ is said to be semialgebraic if its graph $\Gamma_f = \{(x, y) \in A \times B \mid y = f(x)\}$ is a semialgebraic subset of $\mathbb{R}^p \times \mathbb{R}^q$. In particular, each polynomial mapping $f : A \to B$ is semialgebraic. In addition, each regular rational mapping $g : A \to B$ (all its coordinates are rational fractions whose denominators do not vanish on $A$) is semialgebraic. Furthermore, the image and the inverse image of a semialgebraic set by a semialgebraic mapping are semialgebraic. For further details on semialgebraic structures we refer to [5, 7].

5. **Gene-environment networks under semialgebraic uncertainty.** The uncertain states of the gene-environment network are represented in terms of semialgebraic sets. Here, we assume that $(\mathcal{G}, \mathcal{D})$ denotes a static cluster partition with $S$ genetic clusters and $R$ environmental clusters that does not change over the time
interval \( \overline{T} \subset \mathbb{N} \), where \( T = \{0, 1, \ldots, T\} \) and \( \overline{T} = \{0, 1, \ldots, T, T + 1\} \) for a given \( T \in \mathbb{N} \). In the following, measurements are indicated by a bar and predictions are given by a hat (e.g., \( \overline{G}_i \) and \( \hat{G}_i \) denote a measurement and a prediction of gene \( i \), respectively). In addition, for each \( t \in \overline{T} \) we set
\[
\hat{G}^{(t)}(\mathcal{C}) = (\hat{G}_1^{(t)}, \ldots, \hat{G}_R^{(t)}), \\
\tilde{E}^{(t)}(\mathcal{S}) = (\tilde{E}_1^{(t)}, \ldots, \tilde{E}_S^{(t)}).
\]

Semialgebraic prediction models of degree \( D = 1 \) are defined as evolving semialgebraic network models that only depend on the previous state in time.

**Definition 5.1.** A parameter-dependent gene-environment network model of degree \( D = 1 \) under semialgebraic uncertainty is defined as:
\[
\begin{align*}
\hat{G}^{(t+1)}_r &= \mathbb{H}_r(\theta_1; \hat{G}^{(t)}(\mathcal{C}), \tilde{E}^{(t)}(\mathcal{S})) \quad (t \in T; \ r = 1, \ldots, R), \\
\tilde{E}^{(t+1)}_s &= \mathbb{K}_s(\theta_2; \hat{G}^{(t)}(\mathcal{C}), \tilde{E}^{(t)}(\mathcal{S})) \quad (t \in T; \ s = 1, \ldots, S),
\end{align*}
\]
with the initial conditions \( \hat{G}^{(0)}_r = \overline{G}^{(0)}_r \) (\( r = 1, \ldots, R \)) and \( \tilde{E}^{(0)}_s = \overline{E}^{(0)}_s \) (\( s = 1, \ldots, S \)). Here, the set \( \theta = (\theta_1, \theta_2) \in \Theta \), is the unknown parameter vector, where \( \Theta = (\Theta_1, \Theta_2) \subset \mathbb{R}^\ell, \ell \in \mathbb{N} \), denotes the parameter space. The parameter vector \( \theta = (\theta_1, \theta_2) \) determines the set-valued functions \( \mathbb{H}_r(\theta_1; \cdot, \cdot) \) and \( \mathbb{K}_s(\theta_2; \cdot, \cdot) \), where
\[
\begin{align*}
\mathbb{H}_r(\theta_1; \cdot, \cdot) : \Theta_1 \times \left( \bigtimes_{r=1}^R \mathfrak{S}^{[C_r]}, \bigtimes_{s=1}^S \mathfrak{S}^{[D_s]} \right) &\to \mathfrak{S}^{[C_r]}, \\
\mathbb{K}_s(\theta_2; \cdot, \cdot) : \Theta_2 \times \left( \bigtimes_{r=1}^R \mathfrak{S}^{[C_r]}, \bigtimes_{s=1}^S \mathfrak{S}^{[D_s]} \right) &\to \mathfrak{S}^{[D_s]},
\end{align*}
\]
are semialgebraic functions depending on the semialgebraic representations of genetic clusters and/or environmental clusters. The initial values \( \hat{G}^{(0)}_r \in \mathfrak{S}^{[C_r]}, \ r \in \{1, \ldots, R\}, \) and \( \tilde{E}^{(0)}_s \in \mathfrak{S}^{[D_s]}, \ s \in \{1, \ldots, S\}, \) are semialgebraic uncertainty sets.

Figure 3 illustrates the various interactions between genetic clusters and/or environmental clusters as well as the corresponding semialgebraic uncertainty sets.

The above definitions can be further extended to models that take \( T + 1 \) previous system’s states into account.

**Definition 5.2.** A parameter-dependent target-environment network model of degree \( D = T + 1 \geq 1 \) under semialgebraic uncertainty with \( T \in \mathbb{N}_0 \) is defined by
\[
\begin{align*}
\hat{G}^{(t+1)}_r &= \mathbb{H}_r(\theta_1; \hat{G}^{(t)}(\mathcal{C}), \hat{G}^{(t-T)}(\mathcal{C}), \tilde{E}^{(t)}(\mathcal{S}), \ldots, \tilde{E}^{(t-T)}(\mathcal{S})), \\
\tilde{E}^{(t+1)}_s &= \mathbb{K}_s(\theta_2; \hat{G}^{(t)}(\mathcal{C}), \hat{G}^{(t-T)}(\mathcal{C}), \tilde{E}^{(t)}(\mathcal{S}), \ldots, \tilde{E}^{(t-T)}(\mathcal{S}))
\end{align*}
\]
(SPM)
with the initial conditions
\[
\begin{align*}
\hat{G}^{(T)}_r &= \overline{G}^{(T)}_r, \hat{G}^{(T-1)}_r = \overline{G}^{(T-1)}_r, \ldots, \hat{G}^{(0)}_r = \overline{G}^{(0)}_r \quad (r = 1, \ldots, R), \\
\tilde{E}^{(T)}_s &= \overline{E}^{(T)}_s, \tilde{E}^{(T-1)}_s = \overline{E}^{(T-1)}_s, \ldots, \tilde{E}^{(0)}_s = \overline{E}^{(0)}_s \quad (s = 1, \ldots, S).
\end{align*}
\]
Here, the set $\theta = (\theta_1, \theta_2) \in \Theta$ is the unknown parameter vector, where $\Theta = (\Theta_1, \Theta_2) \subset \mathbb{R}^\ell$, $\ell \in \mathbb{N}$, denotes the parameter space. The parameter vector $\theta = (\theta_1, \theta_2)$ determines the set-valued functions $H_r(\theta_1; \cdot, \cdot)$ and $K_s(\theta_2; \cdot, \cdot)$, where

$$H_r(\theta_1; \cdot, \cdot) : \Theta_1 \times \left( \bigtimes_{r=1}^R \mathcal{S}^{[C_r]} \right) \times \left( \bigtimes_{s=1}^S \mathcal{S}^{[D_s]} \right) \to \mathcal{S}^{[C_r]}$$

and

$$K_s(\theta_2; \cdot, \cdot) : \Theta_2 \times \left( \bigtimes_{r=1}^R \mathcal{S}^{[C_r]} \right) \times \left( \bigtimes_{s=1}^S \mathcal{S}^{[D_s]} \right) \to \mathcal{S}^{[D_s]}$$

are semialgebraic functions depending on the semialgebraic representations of genetic clusters and/or environmental clusters. The initial values $G^{(t)}_r \in \mathcal{S}^{[C_r]}$ and $E^{(t)}_s \in \mathcal{S}^{[D_s]}$ are semialgebraic sets for $t \in T$.

6. Semialgebraic measurements and predictions. In the sequel, a regression approach is introduced for parameter identification of semialgebraic prediction models. The semialgebraic regression strategy compares semialgebraic measurements and parameter-dependent semialgebraic predictions obtained with model (SPM).

At a given time $\kappa \in \mathbb{N}$ the number of previous measurements required for the prediction depends on two factors:

a) the number of measurements, $K \in \mathbb{N}$, needed to determine the parameter vector $\theta \in \Theta$,

b) the number of previous states, $D \in \mathbb{N}_0$, used in the corresponding model to compute a prediction.

During the regression process, the measurements have to be compared to the semialgebraic predictions (see Figure 4). The optimal parameter vector $\theta \in \Theta$ maximizes the intersection of the semialgebraic measurements and the semialgebraic predictions.
Figure 4. Comparison of measurements and predictions. The computation of the parameter vector $\theta \in \Theta$ depends on $K$ direct comparisons of the intersection between semialgebraic measurements and predictions.

**Assumption 1.** Let $M \in \mathbb{N}$ be the number of measurements required for the identification of the system’s parameters. For a given degree $T + 1 \geq 1$ assume that there exist $M + T + 1$ measurements of genetic cluster’s states and environmental cluster’s states (both semialgebraic sets) at time $t \in \mathbb{N}$:

$$
\begin{align*}
\bigcap_{r=1}^{T} G_r(t-M-T), G_r(t-M+T+1), \ldots, G_r(t-M-1), G_r(t-M), G_r(t-M+1), \ldots, G_r(t-1), G_r(t), \quad (M1) \\
\bigcap_{s=1}^{M} E_s(t-M-T), E_s(t-M+T+1), \ldots, E_s(t-M-1), E_s(t-M), E_s(t-M+1), \ldots, E_s(t-1), E_s(t), \quad (M2)
\end{align*}
$$

where $r \in \{1, \ldots, R\}$, $s \in \{1, \ldots, S\}$.

7. **Semialgebraic regression for gene-environment networks.** For an estimation of parameters at time $t \in \mathbb{N}$, $t > M$, the **semialgebraic predictions** $\hat{G}_r^{(t-m)}$ and $\hat{E}_s^{(t-m)}$ of the states of genetic and environmental clusters are compared to the **semialgebraic data** $G_r^{(t-m)}$ and $E_s^{(t-m)}$ received from measurements genetic and environmental factors, where $r \in \{1, \ldots, R\}$, $s \in \{1, \ldots, S\}$, and $m \in \{0,1,\ldots,M\}$.

The **intersections of measurement values and predictions** (both semialgebraic sets) are denoted by

$$
\Delta G_r^{(t-m)} := \hat{G}_r^{(t-m)} \cap G_r^{(t-m)} \quad \text{and} \quad \Delta E_s^{(t-m)} := \hat{E}_s^{(t-m)} \cap E_s^{(t-m)}.
$$

The size of the semialgebraic intersections is measured by a so-called **criteria function**. Here, **regular criteria functions** are applied that are monotonous by increasing with respect to inclusion.

**Definition 7.1.** A function $\mathcal{N} : \mathcal{G}^p \to \mathbb{R}_{\geq 0}$ is a **regular criteria function** on the set $\mathcal{G}^p$ of all semialgebraic sets in $\mathbb{R}^p$, if $\mathcal{N}(S_1) \leq \mathcal{N}(S_2)$ for all $S_1, S_2 \in \mathcal{G}^p$ with $S_1 \subseteq S_2$.

Regular criteria functions are usually associated with a measure of the size of the intersection of a semialgebraic measurement and semialgebraic prediction. For example, the **volume** of the semialgebraic set $S$ is a regular criteria function defined by the lesbesgue measure of $S$, i.e., $\text{vol}(S) := \int_S \mathbb{1}(x) \, dx$, where $\mathbb{1}(x) = 1$ if $x \in S$, and $\mathbb{1}(x) = 0$ if $x \not\in S$. 
**Definition 7.2.** Consider a parameter-dependent semialgebraic prediction model on a gene-environment network. At time $t \in \mathbb{N}$, $t > M$, the $M + T + 1$ measurements \((M_1), (M_2)\) are available. For $m = 0, 1, \ldots, M$, the following sets and functions are introduced:

i) the intersections (fusions) $\Delta G^{(t-m)}_r$ and $\Delta E^{(t-m)}_s$ of genetic clusters’ states of environmental clusters’ states, where $r \in \{1, \ldots, R\}$, $s \in \{1, \ldots, S\}$,

ii) the regular criteria functions $N^{(t-m)}_r : |C_r| \to \mathbb{R}_{\geq 0}$ and $N^{(t-m)}_s : |D_s| \to \mathbb{R}_{\geq 0}$.

The semialgebraic regression problem for gene-environment networks based on regular criteria functions is defined by

$$\max_{\theta \in \Theta} \sum_{m=0}^{M} \left\{ \sum_{r=1}^{R} N^{(t-m)}_r \left( \Delta G^{(t-m)}_r \right) + \sum_{s=1}^{S} N^{(t-m)}_s \left( \Delta E^{(t-m)}_s \right) \right\}. \quad \text{(SRP)}$$

8. **Network pruning.** Modern high-throughput technologies allow for a simultaneous measurement of huge samples. For this reason, gene-environment network models can consist of a large number of genes and environmental factors, in particular in genome-wide studies. From a theoretical perspective, the number of interconnections can be extremely large with negative effects on the computational side. Unimportant edges can be deleted in order to avoid the curse of dimensionality. Network pruning has been known as an efficient solution for different real-world applications. For example, deep convolutional neural networks (CNNs) have recently become deeper to attain high efficiency in various applications. In spite of their remarkable success, it is not always practical to apply them to resource-constrained problems. An effective solution to deal with this problem is applying network pruning to decrease the computation cost of over-parameterized CNNs [9]. In the sequel, two different methods are introduced for network pruning. Since network nodes are often regulated by a few entities the number of ingoing and outgoing branches should be constrained. We propose two approaches: The first method imposes binary constraints and leads to a mixed-integer regression problem. In this way, the number of branches can be considerably reduced. Nevertheless, binary constraints are very strict and important branches may be deleted in case of inappropriate constraints. For this reason, the second method relies on continuous constraints that may be interpreted as probabilities, membership assignments or fuzzy values and can be solved by continuous programming.

8.1. **Mixed-integer regression problem.** The first approach towards network pruning depends on the indegree and the outdegree of clusters. The number of regulating connections of each cluster is encoded in the adjacency matrix $A$ of the underlying gene-environment network under uncertainty (GEN-$C_\theta$).

8.1.1. **Indegree of clusters.** Given two clusters $A, B$ of the gene-environment network under uncertainty (GEN-$C_\theta$) the notion $A \leftarrow B$ states that cluster $A$ is regulated by cluster $B$ and $A \not\leftarrow B$ means that cluster $A$ is not regulated by cluster
B. The Boolean matrices

\[
\chi_{ir}^{GG} = \begin{cases} 
1, & \text{if } C_i \leftarrow C_r, \\
0, & \text{if } C_i \nleftrightarrow C_r,
\end{cases}
\quad \chi_{js}^{EG} = \begin{cases} 
1, & \text{if } C_j \leftarrow D_s, \\
0, & \text{if } C_j \nleftrightarrow D_s,
\end{cases}
\]

\[
\chi_{ir}^{GE} = \begin{cases} 
1, & \text{if } D_i \leftarrow C_r, \\
0, & \text{if } D_i \nleftrightarrow C_r,
\end{cases}
\quad \chi_{js}^{EE} = \begin{cases} 
1, & \text{if } D_j \leftarrow D_s, \\
0, & \text{if } D_j \nleftrightarrow D_s,
\end{cases}
\]

indicate whether or not pairs of clusters in the gene-environment regulatory network are directly connected.

For a specific genetic cluster \(C_j\), the indegrees \(\deg(C_j)^{GG}\) and \(\deg(C_j)^{EG}\) count the number of genetic and environmental clusters that regulate cluster \(C_j\).

**Definition 8.1.** The indegree of the genetic cluster \(C_j\) in the regulatory network is defined with respect to all genetic and environmental clusters by

\[
\deg(C_j)^{GG} := \sum_{r=1}^{R} \chi_{ir}^{GG} \quad \text{and} \quad \deg(C_j)^{EG} := \sum_{s=1}^{S} \chi_{js}^{EG},
\]

where \(j \in \{1, \ldots, R\}\). The total indegree of the genetic cluster \(C_j\) is defined by

\[
\deg(C_j)^{GG} + \deg(C_j)^{EG}.
\]

In a similar way, the indegree of an environmental cluster \(D_i\) can be defined. Again, the indegrees \(\deg(D_i)^{GE}\) and \(\deg(D_i)^{EE}\) count the number of genetic and environmental clusters that regulate cluster \(D_i\).

**Definition 8.2.** For \(i \in \{1, \ldots, S\}\) the indegree of environmental cluster \(D_i\) with respect to the environmental clusters and the genetic clusters is given by

\[
\deg(D_i)^{GE} := \sum_{r=1}^{R} \chi_{ir}^{GE} \quad \text{and} \quad \deg(D_i)^{EE} := \sum_{s=1}^{S} \chi_{is}^{EE},
\]

where \(j \in \{1, \ldots, R\}\). The total indegree of cluster \(D_i\) is defined by

\[
\deg(D_i)^{GE} + \deg(D_i)^{EE}.
\]

1.1.2. Outdegree of clusters. In the same way as before, bounds on the outdegree (i.e., the number of outgoing branches) can be introduced. Firstly, binary values are defined to determine whether or not there is an outgoing connection:

\[
\zeta_{ir}^{GG} = \begin{cases} 
1, & \text{if } C_r \rightarrow C_i, \\
0, & \text{if } C_r \nleftrightarrow C_i, \n\end{cases}
\quad \zeta_{js}^{EG} = \begin{cases} 
1, & \text{if } D_s \rightarrow C_j, \\
0, & \text{if } D_s \nleftrightarrow C_j,
\end{cases}
\]

\[
\zeta_{ir}^{GE} = \begin{cases} 
1, & \text{if } C_r \rightarrow D_i, \\
0, & \text{if } C_r \nleftrightarrow D_i,
\end{cases}
\quad \zeta_{js}^{EE} = \begin{cases} 
1, & \text{if } D_s \rightarrow D_i, \\
0, & \text{if } D_s \nleftrightarrow D_i.
\end{cases}
\]

The outdegrees \(\deg(C_j)^{GG}_{out}\) and \(\deg(C_j)^{EG}_{out}\) count the number of genetic and environmental clusters regulated by cluster \(C_j\).

**Definition 8.3.** The outdegree of genetic cluster \(C_j\) with respect to all genetic and environmental clusters can be expressed as

\[
\deg(C_j)^{GG}_{out} := \sum_{r=1}^{R} \zeta_{ir}^{GG} \quad \text{and} \quad \deg(C_j)^{GE}_{out} := \sum_{s=1}^{S} \zeta_{js}^{GE},
\]

where \(j \in \{1, \ldots, R\}\). The total outdegree of genetic cluster \(C_j\) is given by

\[
\deg(C_j)^{GG}_{out} + \deg(C_j)^{GE}_{out}.
\]

The outdegrees \(\deg(D_i)^{GE}_{out}\) and \(\deg(D_i)^{EE}_{out}\) count the number of genetic and environmental clusters regulated by cluster \(D_i\).

**Definition 8.4.** The outdegree of environmental cluster \(D_i\) with respect to the environmental clusters and the genetic clusters is defined by

\[
\deg(D_i)^{GE}_{out} := \sum_{r=1}^{R} \zeta_{ir}^{GE} \quad \text{and} \quad \deg(D_i)^{EE}_{out} := \sum_{s=1}^{S} \zeta_{is}^{EE},
\]

where \(i \in \{1, \ldots, S\}\). The total outdegree of cluster \(D_i\) is

\[
\deg(D_i)^{GE}_{out} + \deg(D_i)^{EE}_{out}.
\]
8.1.3. Restrictions on ingoing/outgoing branches for each cluster. As mentioned before, upper bounds on the indegrees and the outdegrees of the nodes (clusters) are introduced with the aim of network pruning. These values have to be given by the practitioner and they can depend on any a priori information. If these additional constraints are integrated into the regression model, the following mixed-integer optimization problem for network pruning is achieved:

\[
\text{Maximize } \sum_{m=1}^{M} \left\{ \sum_{r=1}^{R} \left\| \Delta G_r^{(t-m)} \right\| + \sum_{s=1}^{S} \left\| \Delta E_s^{(t-m)} \right\| \right\} \quad (M1)
\]

subject to \( \deg(C_j)_{\text{in}} \leq \alpha_j^{GG} \), \( \deg(C_j)_{\text{out}} \leq \alpha_j^{EG} \),

\( \deg(C_j)_{\text{in}} \leq \beta_j^{GG} \), \( \deg(C_j)_{\text{out}} \leq \beta_j^{EG} \),

\( \deg(D_i)_{\text{in}} \leq \alpha_i^{EG} \), \( \deg(D_i)_{\text{out}} \leq \alpha_i^{EE} \),

\( \deg(D_i)_{\text{in}} \leq \beta_i^{GE} \), \( \deg(D_i)_{\text{out}} \leq \beta_i^{EE} \)

\( (j = 1, \ldots, R; i = 1, \ldots, S) \).

In model (M1), individual bounds on the indegrees and outdegrees of each genetic and environmental cluster are imposed in order to control the connectivity of the gene-environment network. This approach is an extension of the regression problems with bounds on the indegrees discussed in [17, 18]. Mixed-integer problems for an analysis of gene-environment networks based on interval arithmetic were presented in [50, 51, 53].

8.1.4. Restrictions on the total number of ingoing/outgoing branches. The number of connections of each cluster with other genetic and environmental clusters in model (M1) are considered separately. In a further step, these bounds can be combined to impose restrictions on the total number of all ingoing or outgoing branches of each cluster:

\[
\text{Maximize } \sum_{m=1}^{M} \left\{ \sum_{r=1}^{R} \left\| \Delta G_r^{(t-m)} \right\| + \sum_{s=1}^{S} \left\| \Delta E_s^{(t-m)} \right\| \right\} \quad (M2)
\]

subject to \( \deg(C_j)_{\text{in}} \leq \gamma_j \), \( \deg(C_j)_{\text{out}} \leq \epsilon_j \) \( (j = 1, \ldots, R) \),

\( \deg(D_i)_{\text{in}} \leq \delta_i \), \( \deg(D_i)_{\text{out}} \leq \varphi_i \) \( (i = 1, \ldots, S) \).

8.2. Continuous programming. The binary constraints in (M1) and (M2) are very strict and if the constraints are not chosen in an appropriate way, important branches of the regulatory network could be deleted. For this reason, continuous optimization is applied for a relaxation of (M1) by replacing the binary variables \( \chi_{jr}^{GG}, \chi_{jr}^{EG}, \chi_{ir}^{GE}, \chi_{is}^{EE} \in \{0,1\} \) with real variables \( \rho_{jr}^{GG}, \rho_{jr}^{EG}, \rho_{ir}^{GE}, \rho_{is}^{EE} \in [0,1] \). These variables can also be interpreted as probabilities or fuzzy membership values (see [35] for optimization models with probabilistic constraints). In addition, the variables should linearly depend on the corresponding subvectors of the parameter vector \( \theta \in \Theta \) denoted by \( \theta_{jr}^{GG}, \theta_{jr}^{EG}, \theta_{ir}^{GE}, \theta_{is}^{EE} \).
8.2.1. Indegree of clusters.

**Definition 8.5.** The real-valued indegree of genetic cluster $C_j$ in the regulatory network with respect to the genetic and environmental clusters is now defined by

$$\text{deg}(C_j)_{\text{in}}^{GG} := \sum_{r=1}^{R} \varphi_{jr}^{GG} (\theta_{jr}^{GG}) \quad \text{and} \quad \text{deg}(C_j)_{\text{in}}^{EG} := \sum_{s=1}^{S} \varphi_{js}^{EG} (\theta_{js}^{EG}).$$

Similarly, the real-valued indegree of cluster $D_i$ is adapted in the following way.

**Definition 8.6.** The real-valued indegree of cluster $D_i$ is given by

$$\text{deg}(D_i)_{\text{in}}^{GE} := \sum_{r=1}^{R} \varphi_{ir}^{GE} (\theta_{ir}^{GE}) \quad \text{and} \quad \text{deg}(D_i)_{\text{in}}^{EE} := \sum_{s=1}^{S} \varphi_{is}^{EE} (\theta_{is}^{EE}).$$

8.2.2. Outdegree of clusters. In the same way, the outdegrees of clusters can be adapted by replacing the binary variables $\varphi_{jr}^{GG}$, $\varphi_{js}^{EG}$, $\varphi_{ir}^{GE}$, and $\varphi_{is}^{EE}$ with real variables $\varphi_{jr}^{GG}$, $\varphi_{js}^{EG}$, $\varphi_{ir}^{GE}$, and $\varphi_{is}^{EE}$ in $[0,1]$ linearly depending on the corresponding elements of the parameter vector $\theta \in \Theta$ denoted by $\theta_{jr}^{GG}$, $\theta_{js}^{EG}$, $\theta_{ir}^{GE}$, and $\theta_{is}^{EE}$.

**Definition 8.7.** The real-valued outdegrees of genetic cluster $C_j$ with respect to the genetic and environmental clusters are defined by

$$\text{deg}(C_j)_{\text{out}}^{GG} := \sum_{r=1}^{R} \varphi_{jr}^{GG} (\theta_{jr}^{GG}) \quad \text{and} \quad \text{deg}(C_j)_{\text{out}}^{EG} := \sum_{s=1}^{S} \varphi_{js}^{EG} (\theta_{js}^{EG}).$$

In addition, the outdegree of environmental clusters can be adapted to comprise real values.

**Definition 8.8.** The real-valued outdegree of environmental cluster $D_i$ is given by

$$\text{deg}(D_i)_{\text{out}}^{GE} := \sum_{r=1}^{R} \varphi_{ir}^{GE} (\theta_{ir}^{GE}) \quad \text{and} \quad \text{deg}(D_i)_{\text{out}}^{EE} := \sum_{s=1}^{S} \varphi_{is}^{EE} (\theta_{is}^{EE}).$$

8.2.3. Restrictions on ingoing/outgoing branches for each cluster. By replacing the strict binary constraints of the mixed-integer problem (MI1) with the aforementioned ‘soft constraints’, the following continuous programming problem for network pruning is obtained:

Maximize \[ \sum_{k=1}^{T} \left\{ \sum_{r=1}^{R} \left\| \Delta G_{r}^{(k)} \right\| + \sum_{s=1}^{S} \left\| \Delta E_{s}^{(k)} \right\| \right\} \] \tag{C1}

subject to

\[ \sum_{r=1}^{R} \varphi_{jr}^{GG} (\theta_{jr}^{GG}) \leq \alpha_{j}^{GG}, \quad \sum_{s=1}^{S} \varphi_{js}^{EG} (\theta_{js}^{EG}) \leq \alpha_{j}^{EG}, \]
\[ \sum_{r=1}^{R} \varphi_{ir}^{GE} (\theta_{ir}^{GE}) \leq \alpha_{i}^{GE}, \quad \sum_{s=1}^{S} \varphi_{is}^{EE} (\theta_{is}^{EE}) \leq \alpha_{i}^{EE}, \]
\[ \sum_{r=1}^{R} \varphi_{jr}^{GG} (\theta_{jr}^{GG}) \leq \beta_{j}^{GG}, \quad \sum_{s=1}^{S} \varphi_{js}^{EG} (\theta_{js}^{EG}) \leq \beta_{j}^{EG}, \]
\[ \sum_{r=1}^{R} \varphi_{ir}^{GE} (\theta_{ir}^{GE}) \leq \beta_{i}^{GE}, \quad \sum_{s=1}^{S} \varphi_{is}^{EE} (\theta_{is}^{EE}) \leq \beta_{i}^{EE} \]
\[(j = 1, \ldots, R, i = 1, \ldots, S)\]
8.2.4. Restrictions on the total number of ingoing/outgoing branches. It is also possible to combine these real-valued restrictions on the total number of all ingoing or outgoing branches of each cluster and to obtain a relaxation of model (MI2) in terms of the following continuous programming problem:

\[ \text{Maximize} \sum_{\kappa=1}^{T} \left\{ \sum_{r=1}^{R} \| \Delta G_{\kappa}^{(r)} \| + \sum_{s=1}^{S} \| \Delta E_{\kappa}^{(s)} \| \right\} \tag{C2} \]

subject to

\[ \sum_{r=1}^{R} G_{jr}(\theta_{jr}) + \sum_{s=1}^{S} G_{js}(\theta_{js}) \leq \gamma_j \quad (j = 1, \ldots, R), \]
\[ \sum_{r=1}^{R} G_{ir}(\theta_{ir}) + \sum_{s=1}^{S} G_{is}(\theta_{is}) \leq \beta_i \quad (i = 1, \ldots, S), \]
\[ \sum_{r=1}^{R} E_{jr}(\theta_{jr}) + \sum_{s=1}^{S} E_{js}(\theta_{js}) \leq \gamma_j \quad (j = 1, \ldots, R), \]
\[ \sum_{r=1}^{R} E_{ir}(\theta_{ir}) + \sum_{s=1}^{S} E_{is}(\theta_{is}) \leq \delta_i \quad (i = 1, \ldots, S). \]

9. Selected classes of optimization problems and their algorithmic techniques. This paper is devoted to the foundations of semialgebraic gene-environment networks, i.e., to bases of the future development of algorithmic methods to assess those networks and to real-work applications. Already today we may mention that the following refined classes of techniques can be naturally suggested for our new class of networks under uncertainty, for their identification, optimization and extension: (i) Tchebychev Approximation [53], (ii) Semi-Infinite Optimization [29], (iii) Generalized Semi-Infinite Optimization [48, 49, 56], (iv) Bi-level and Multi-level Optimization [36], (v) Disjunctive Optimization [2], (vi) Robust Optimization [24, 39, 42], (vii) Conic Optimization [4], (viii) Optimal Control [1], and (ix) Stochastic Optimal Control [33].

Concerning classes of future real-world applications we would like to recommend emerging challenges of, for example, (a) Collaborative Games under Ellipsoidal Uncertainty or (per inner or outer approximations) Hypercube Uncertainty [11, 52], (b) Transportation (“Piano Mover’s” and many more) problems [32, 40], (c) Supply Chain and Inventory Management [12, 20, 30, 41, 43, 57] Production Planning [38], various kinds of (d) Design problems [46], (e) Artificial Intelligence and Machine Learning (e.g., “Infinite Kernel Learning”) [6, 13, 28], and (f) Finance, Actuarial Sciences and Pension Fund Systems [14, 19, 55].

10. Conclusion. In this study, we lay the foundations of the novel concept of semialgebraic gene-environment networks. In this approach, the uncertain states of genes and environmental factors are represented by semialgebraic sets. The functional relationships between the genes and/or environmental factors are given by semialgebraic functions. This approach further extends previous set-theoretic regression strategies for gene-environment networks under uncertainty based on intervals, polyhedral and ellipsoidal uncertainty and the corresponding continuous programming. Gene-environment studies often rely on a mixture of measurable chemical, physical and biological data as well as data on behavioral patterns and life events. The system states have to reflect different types of uncertainty such a
probabilities and fuzzy values. It is a major advantage of the proposed semialgebraic
gene-environment networks that the generic structure of semialgebraic uncertainty
sets allows for a simultaneous consideration of such data. In addition, it provides
a major computational advantage. Uncertain system states can be described by a
finite set of intersections and unions of underlying regions of uncertainty. In this
way, a semialgebraic calculus is introduced that builds the basis for the semialgebraic
regression model. Future studies will further explore graph-theoretic characteristics
of gene-environment networks under semialgebraic uncertainty. In addition, set-
theoretic notions of boundedness, stability and reachability can support the analysis
since they are often associated with a critical deviation in bio-systems. Furthermore,
the effects of exogenous disturbances have to be discussed in terms of robustness
of the gene-environment regulatory system. The combination of graph theoretic
network analysis and the behaviour of trajectories of the underlying regulatory
system offers a promising avenue for gene-environment studies with more complex
uncertainty regions in genetics, neuroscience and psychobiology.

REFERENCES

[1] M. U. Akhmet, M. Kirane, M. A. Tleubergenova and G. Weber, Control and optimal response
problems for quasilinear impulsive integrodifferential equations, European J. Oper. Res., 169
(2006), 1128–1147.
[2] B. Akteke-Öztürk, G.-W. Weber and G. Köksal, Optimization of generalized desirability
functions under model uncertainty, Optimization, 66 (2017), 2157–2169.
[3] A. Arnau-Soler, et al., Genome-wide by environment interaction studies of depressive symp-
toms and psychosocial stress in UK Biobank and Generation Scotland, Translational Psychi-
atry, 9 (2019), 13 pp.
[4] E. Ayyildiz, V. Purutçuoğlu and G. W. Weber, Loop-based conic multivariate adaptive re-
gression splines is a novel method for advanced construction of complex biological networks,
European J. Oper. Res., 270 (2018), 852–861.
[5] J. Bochnak, M. Coste and M.-F. Roy, Semi-algebraic Sets, A Series of Modern Surveys in
Mathematics, 36, Springer Berlin Heidelberg, Berlin, Heidelberg, 1998, 23–58.
[6] A. Çevik, G. Weber, B. M. Eyüboğlu and K. K. Onguz, The Alzheimer’s disease neuroimaging
initiative Voxel-MARS: A method for early detection of Alzheimer’s disease by classification
of structural brain MRI, Ann. Oper. Res., 258 (2017), 31–57.
[7] H. Delfs and M. Knebusch, Locally Semialgebraic Spaces, Lecture Notes in Mathematics, Vol.
1173, Springer-Verlag, Berlin, 1985.
[8] R. Dobson and G. Giovannoni, Multiple sclerosis - A review, European Journal of Neurology,
26 (2019), 27–40.
[9] X. Dong and Y. Yang, Network pruning via transformable architecture search, NeurIPS, 32
(2019), 760–771.
[10] A. Eid, I. Mhatre and J. R. Richardson, Gene-environment interactions in Alzheimer’s disease:
A potential path to precision medicine, Pharmacology & Therapeutics, 199 (2019), 173–187.
[11] S. Z. A. Gök and G.-W. Weber, On dominance core and stable sets for cooperative ellipsoidal
games, Optimization, 62 (2013), 1297–1308.
[12] A. Goli, H. K. Zare, R. Tavakkoli-Moghaddam and A. Sadeghieh, Multiobjective fuzzy math-
ematical model for a financially constrained closed-loop supply chain with labor employment,
Computational Intelligence, 36 (2020), 4–36.
[13] A. Goli, H. Khademi Zare, R. Tavakkoli-Moghaddam and A. Sadeghieh, Hybrid artificial
intelligence and robust optimization for a multi-objective product portfolio problem case
study: The dairy products industry, Computers and Industrial Engineering, 137 (2019),
106090.
[14] G. Kara, A. Özmen and G.-W. Weber, Stability advances in robust portfolio optimization
under parallelepiped uncertainty, CEJOR Cent. Eur. J. Oper. Res., 27 (2019), 241–261.
[15] E. Kropat, A. Özmen, G.-W. Weber, S. Meyer-Nieberg and O. Defterli, Fuzzy prediction
strategies for gene-environment networks– fuzzy regression analysis for two-modal regulatory
systems, RAIRO-Oper. Res., 50 (2016), 413–435.
[16] E. Kropat and G.-W. Weber, Fuzzy target-environment networks and fuzzy-regression approaches, *Numer. Algebra, Control Optim.*, 8 (2018), 135–155.

[17] E. Kropat, G.-W. Weber and C. Pedamallu, Regulatory networks under ellipsoidal uncertainty – data analysis and prediction by optimization theory and dynamical systems, in *Data Mining: Foundations and Intelligent Paradigms Vol. 2*, Intell. Syst. Ref. Libr., 24, Springer, Berlin, 2012, 27–56.

[18] E. Kropat, G.-W. Weber and J.-J. Rückmann, Regression analysis for clusters in gene-environment networks based on ellipsoidal calculus and optimization, *Dyn. Contin. Discrete Impuls. Syst. Ser. B Appl. Algorithms*, 17 (2010), 639–657.

[19] E. Küürüm, G.-W Weber and C. Iyigun, Early warning on stock market bubbles via methods of optimization, clustering and inverse problems, *Ann. Oper. Res.*, 260 (2018), 293–320.

[20] A. Mardani, D. Kannan, R. E. Hooker, S. Ozkul, M. Alrasheedi and E. B. Tirkolaee, Evaluation of green and sustainable supply chain management using structural equation modelling: A systematic review of the state of the art literature and recommendations for future research, *Journal of Cleaner Production*, 249 (2020), 119383.

[21] K. McAllister, et al., Current challenges and new opportunities for gene-environment interaction studies of complex diseases, *American Journal of Epidemiology*, 186 (2017), 753–761.

[22] M. C. Mills and C. Rahal, A scientometric review of genome-wide association studies, *Communications Biology*, 2 (2019), 11 pp.

[23] R. J. Musci, J. L. Augustinavicius and H. Volk, Gene-environment interactions in psychiatry: Recent evidence and clinical implications, *Current Psychiatry Reports*, 21 (2019), 81 pp.

[24] Ö. N. Onak, Y. S. Dogrusoz and G. Weber, Robustness of reduced order nonparametric model for inverse ECG solution against modeling and measurement noise, in *Computing in Cardiology, CinC 2018*, Maastricht, The Netherlands, September 23-26, 2018, 2018, 1–4.

[25] A. Özmen, E. Kropat and G.-W. Weber, Spline regression models for complex multi-modal regulatory networks, *Optim. Methods Softw.*, 29 (2014), 515–534.

[26] A. Özmen, E. Kropat and G.-W. Weber, Robust optimization in spline regression models for multi-model regulatory networks under polyhedral uncertainty, *Optimization*, 66 (2017), 2135–2155.

[27] A. Özmen, G.-W. Weber and E. Kropat, Robustification of conic generalized partial linear models under polyhedral uncertainty, *International IFNA-ANS Scientific Journal “Problems of Nonlinear Analysis in Engineering Systems”*, 38 (2012), 104–113.

[28] S. Özgüür-Akyüz and G.-W. Weber, On numerical optimization theory of infinite kernel learning, *J. Global Optim.*, 48 (2010), 215–239.

[29] S. Özgüür-Akyüz and G.-W. Weber, Infinite kernel learning via infinite and semi-infinite programming, *Optim. Methods Softw.*, 25 (2010), 937–970.

[30] T. Paksoy, E. Özceylan and G.-W. Weber, Profit oriented supply chain network optimization, *CEJOR Cent. Eur. J. Oper. Res.*, 21 (2013), 455–478.

[31] J. Rogers, T. Renoir and A. J. Hannan, Gene-environment interactions informing therapeutic approaches to cognitive and affective disorders, *Neuropharmacology*, 145 (2019), 37–48.

[32] S. K. Roy, G. Maity and G.-W Weber, Multi-objective two-stage grey transportation problem using utility function with goals, *CEJOR Cent. Eur. J. Oper. Res.*, 25 (2017), 417–439.

[33] E. Savku and G.-W Weber, A stochastic maximum principle for a Markov regime-switching jump-diffusion model with delay and an application to finance, *J. Optim. Theory Appl.*, 179 (2018), 696–721.

[34] A. J. Schork, et al., A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment, *Nature Neuroscience*, 22 (2019), 353–361.

[35] A. Shapiro, D. Dentcheva and A. Ruszczynski, *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[36] O. Stein, *Bi-Level Strategies in Semi-Infinite Programming*, Nonconvex Optimization and its Applications, 71, Kluwer Academic Publishers, Boston, MA, 2003.

[37] P. Taylan, G.-W. Weber and E. Kropat, Approximation of stochastic differential equations by additive models using splines and conic programming, *International Journal of Computing Anticipatory Systems*, 21 (2008), 341–352.

[38] E. B. Tirkolaee, A. Goli and G.-W. Weber, *Multi-objective Aggregate Production Planning Model Considering Overtime and Outsourcing Options Under Fuzzy Seasonal Demand*, Springer, Cham, 2019, 81–96.
[39] E. B. Tirkolaee, S. Hadian, G.-W. Weber and I. Mahdavi, A robust green traffic-based routing problem for perishable products distribution, *Computational Intelligence*, 36 (2020), 80–101.

[40] E. B. Tirkolaee, I. Mahdavi, M. M. Seyyed Esfahani and G.-W. Weber, A hybrid augmented ant colony optimization for the multi-trip capacitated arc routing problem under fuzzy demands for urban solid waste management, *Waste Management and Research*, 38 (2020), 156–172.

[41] E. B. Tirkolaee, I. Mahdavi, M. M. Seyyed Esfahani and G.-W. Weber, A novel hybrid method using fuzzy decision making and multi-objective programming for sustainable-reliable supplier selection in two-echelon supply chain design, *Journal of Cleaner Production*, 250 (2020), 119517.

[42] E. B. Tirkolaee, I. Mahdavi, M. M. Seyyed Esfahani and G.-W. Weber, A robust green location-allocation-inventory problem to design an urban waste management system under uncertainty, *Waste Management*, 102 (2020), 340–350.

[43] E. B. Tirkolaee, J. Mahmoodkhani, M. R. Bourani and R. Tavakkoli-Moghaddam, A self-learning particle swarm optimization for robust multi-echelon capacitated location-allocation-inventory problem, *Journal of Advanced Manufacturing Systems*, 18 (2019), 677–694.

[44] R. Uher, The implications of gene-environment interactions in depression: Will cause inform cure?, *Molecular Psychiatry*, 13 (2008), 1070–1078.

[45] S. Van der Auwera, et al., Genome-wide gene-environment interaction in depression: A systematic evaluation of candidate genes, *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 177 (2018), 40–49.

[46] A. Vasin, P. Kartunova and G.-W Weber, Models for capacity and electricity market design, CEJOR Cent. Eur. J. Oper. Res., 21 (2013), 651–661.

[47] A. Vasilis, P. Kartunova and G.-W Weber, Models for capacity and electricity market design, CEJOR Cent. Eur. J. Oper. Res., 21 (2013), 651–661.

[48] A. Vasin, P. Kartunova and G.-W Weber, Models for capacity and electricity market design, CEJOR Cent. Eur. J. Oper. Res., 21 (2013), 651–661.

[49] G.-W. Weber, Charakterisierung struktureller stabilität in der nichtlinearen optimierung, Aachen Beiträge zur Mathematik, 5, Augustinus publishing house (now: Mainz publishing house), Aachen, 1992.

[50] G.-W. Weber, Generalized Semi-infinite Optimization and Related Topics, Research and Exposition in Mathematics, 29, Heldermann Verlag, Lemgo, 2003.

[51] G.-W. Weber, S. Alparslan-Gök and N. Dikmen, Environmental and life sciences: Gene-environment networks optimization, games and control - a survey on recent achievements, *Journal of Organisational Transformation & Social Change*, 5 (2008), 197–233.

[52] G.-W. Weber, S. Alparslan-Gök and B. Söyler, A new mathematical approach in environmental and life sciences: Gene-environment networks and their dynamics, *Environmental Modeling & Assessment*, 14 (2009), 267–288.

[53] G.-W. Weber, R. Branzei and S. Alparslan-Gök, On cooperative ellipsoidal games, in 24th Mini EURO Conference-On Continuous Optimization and Information-Based Technologies in the Financial Sector, MEC EuroPT, 2010, 369–372.

[54] G.-W. Weber, E. Kropat, B. Akteke-Öztürk and Z.-K. Görgülü, A survey on OR and mathematical methods applied on gene-environment networks, CEJOR Cent. Eur. J. Oper. Res., 17 (2009), 315–341.

[55] G.-W. Weber, E. Kropat, A. Tezel and S. Belen, Optimization applied on regulatory and eco-finance networks – survey and new developments, *Pac. J. Optim.*, 6 (2010), 319–340.

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

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

[58] A. Yildizbas, A. Calik, T. Paksoy, R. Farahani and G.-W. Weber, Multi-level optimization of an automotive closed-loop supply chain network with interactive fuzzy programming approaches, *Technological and Economic Development of Economy*, 24 (2018), 1004–1028.

Received for publication December 2019.

*E-mail address*: erik.kropat@online.de  
*E-mail address*: gerhard.weber@put.poznan.pl  
*E-mail address*: e.babaee@ustmb.ac.ir