REGRESSION ANALYSIS FOR CLUSTERS IN GENE-ENVIRONMENT NETWORKS BASED ON ELLIPSOIDAL CALCULUS AND OPTIMIZATION

Erik Kropat[1], Gerhard-Wilhelm Weber[2] and Jan-J. Rückmann[3]

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

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

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

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

Keywords. Gene-environment networks, ellipsoidal OR, exact clustering, semidefinite programming, dynamical systems, computational statistics, uncertainty, identification, systems biology, computational biology.

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

1 Introduction

Genetic systems are often characterized by the presence of a high number of variables and parameters resulting in an extraordinary complexity of the underlying regulatory networks. The technological progress of the last decades and the development of high throughput technologies resulted in a generation of massive quantities of data that can be used to enlighten the hidden network structure. Along with these developments, the availability of extremely large data sets also challenged