European Journal of Operational Research 270 (2018) 784–793

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

A graph-based algorithm for the multi-objective optimization of gene regulatory networks

Philippe Nghe[a,b,*], Bela M. Mulder[a], Sander J. Tans[a]

[a] FOM Institute AMOLF, Science Park 104, 1098XG Amsterdam, The Netherlands
[b] Laboratoire de Biochimie, UMR CNRS-ESPCI Chimie Biologie Innovation 8231, PSL Research University, ESPCI, 10 rue Vauquelin, Paris 75005, France

ARTICLE INFO

Article history:
Received 7 November 2016
Accepted 11 April 2018
Available online 21 April 2018

Keywords:
Multiple objective programming
Dynamic programming
Monotone systems
Gene regulatory networks
Systems biology

ABSTRACT

The evolution of gene regulatory networks in variable environments poses Multi-objective Optimization Problem (MOP), where the expression levels of genes must be tuned to meet the demands of each environment. When formalized in the context of monotone systems, this problem falls into a sub-class of linear MOPs. Here, the constraints are partial orders and the objectives consist of either the minimization or maximization of single variables, but their number can be very large. To efficiently and exhaustively find Pareto optimal solutions, we introduce a mapping between colored Hasse diagrams and polytopes associated with an ideal point. A dynamic program based on edge contractions yields an exact closed-form description of the Pareto optimal set, in polynomial time of the number of objectives relative to the number of faces of the Pareto front. We additionally discuss the special case of series-parallel graphs with monochromatic connected components of bounded size, for which the running time and the representation of solutions can in principle be linear in the number of objectives.

1. Introduction

Adaptation to changing environments *via* gene regulation is by nature a multi-objective problem, where the expression level of genes must be optimally set given multiple possible combinations of environmental signals (Poelwijk, de Vos, & Tans, 2011). The multi-valued output response to a collection of inputs is determined by the connectivity of the gene network (Alon, 2006). Biological evolution adds a second layer of complexity, whereby the strength of the connections can be altered by mutations, modulating the multi-valued response itself. There does not currently exist any generic approach as to predict bounds to how well this response can be optimized, given the connectivity of the network.

Here, we apply the notion of Pareto optimality to gene expression in several environments. Pareto optimality is a natural extension of the concept of maximum to multi-objective optimization problems. A solution is part of the Pareto optimal set, or Pareto front, if it is impossible to improve one objective without worsening another. Instead of imposing an aggregation of the different objectives into a scalar function, Pareto optimality keeps track of all potentially interesting solutions in the presence of trade-offs. The Pareto approach, originally introduced in economics (Pareto, 1971; Voorneveld, 2003), has proved useful in many engineering applications (Dhaenens, Julien, & El-Ghazali, 2010; Ehrgott & Gandibleux, 2000; Geilen, Basten, Theelen, & Otten, 2005; Zitzler, Thiele, Laumanns, Fonseca, & VGD, 2003), decision-making analysis (Yang & Catthoor, 2003), and recently, medicine (Cruz-Ramírez, Hervás-Martaínez, Fernández, Briceño, & de la, 2013) and biology (Schuetz, Zamboni, Zampieri, Heinemann, & Sauer, 2012; Shoval et al., 2012).

We approach the problem in the framework of monotone systems, which is widely used in control theory (Angeli & Sontag, 2003), and more specifically for modeling gene regulatory networks (Sontag, 2005). This formalism, as detailed in Section 2, leads us to define partial order constraints between the expression levels of a gene given the environmental inputs. Expression of a gene is considered to be either beneficial or detrimental in each given environment, meaning that expression of a gene must be either minimized or maximized given an input. The expression level of the gene in each environment corresponds to a dimension of the decision space. Given that one objective is associated with each environment, the objective space has the same dimensionality.

The partial order constraints that define the feasible set imply that our problem falls into the category of linear Multi-objective Optimization Problems (linear MOPs) (Greco, Figueira, & Ehrgott, 2005). This problem could in principle be tackled with existing strategies, such as the multi-objective simplex algorithm or

* Corresponding author at: Laboratoire de Biochimie, UMR CNRS-ESPCI Chimie Biologie Innovation 8231, PSL Research University, ESPCI, 10 rue Vauquelin, Paris 75005, France.
E-mail address: philippe.nghe@espci.fr (P. Nghe).

https://doi.org/10.1016/j.ejor.2018.04.020
0377-2217/

Benson's algorithm (Ehrgott, Löhne, & Shao, 2012; Löhne & Weissing, 2017). However, although such algorithms can cope with large sized decision spaces, in our case the number of optimization objectives equals the number of decision variables and can reach several dozen. The unusually high number of objectives imposed by our application is considered a particularly difficult problem in the general case, as the category of *many-objectives optimization problems* starts as soon as there are more than three or four objectives (Fleming & Purshouse, 2002; Jaimes & CAC, 2015).

The algorithm presented here exploits the specificities of our linear MOP, in which constraints are exclusively partial orders and objectives are either the maximization or minimization of coordinates of the decision variables. This algorithm provides an efficient, exact and exhaustive description of the Pareto front, even with such a large number of objectives.

Note that the resolution of the problem provided in the present work could apply beyond the framework of monotone systems, in cases where constraints are expressed in the form of partial orders from the start. This could be the case for task scheduling problems: some tasks must be realized in a certain temporal order relative to each other due to design constraints (Policella, Cesta, Oddi, & Smith, 2007) (e.g. the assembly of the different parts of a car), with some having to be carried out as soon as possible and others as late as possible due to externalities (e.g. supply constraints, processing unit occupancy).

This paper is structured as follows: in Section 2, we detail the problem of determining bounds to the evolutionary potential of gene regulatory networks and its formalization in terms of multi-objective optimization under partial order constraints. In Section 3, we provide a graph formulation of this problem, using Hasse diagrams (Skiena, 1990), which we color according to optimization objectives. In Section 4, we describe and prove a graph algorithm based on successive edge contractions with appropriate vertex coloring rules. In Section 5, we discuss the complexity of our algorithm as a function of the number of objectives N and of the number N_P of maximal convex subsets of the Pareto front, or maximal efficient faces as defined in the papers (Yu & Zeleny, 1975) or (Ecker, Hegner, & Kouada, 1980). This leads us to propose an improved version of the algorithm running in $O(P(N)\cdot N_P)$ time, where P is a polynomial. We discuss how a parameterized complexity approach (Alber, Fernau, & Niedermeier, 2004) can provide a combinatorial description of the Pareto front, with a complexity of $O(N)$ in the case of series-parallel partial orders with monochromatic connected components of bounded size. In Section 6, we provide an example of an explicit resolution and an exhaustive characterization of Pareto front sizes in the problem of 2D dimensional gradients of signals with stress patches.

2. Multi-objective optimization in regulatory networks

Organisms are typically confronted with a large variety of environmental signals that can themselves combine into an even larger diversity of spatio-temporal niches (Chait, Palmer, Yelin, & Kishony, 2016). Whether organisms are able to evolve an appropriate response to this diversity of environments depends fundamentally on the constraints imposed by their current regulatory responses. Despite the multi-objective nature of this problem, research on the evolution of gene regulatory networks has so far relied almost exclusively on explicit simulations of complex responses given a single objective, such as the number of maxima of a biological trait in a gradient (François & Hakim, 2004), but Pareto optimality has been considered only scarcely (Warmflash, Francois, & Siggia, 2012). One issue is that the diversity of environmental conditions is rarely known and the number of potential environments is very large. For example, an *Escherichia coli* bacterium harbours more than 400 regulatory genes responding to typically as many signals. Following an inference approach and testing all potential environments given a fixed gene regulatory architecture would require the consideration of combinations of these signals, the number of which is very large.

In cells, external signals are processed by signal molecules or gene products (e.g. transcription factors) modulating the production or the activity of others. These modulations follow various connectivity patterns comprising cascades and logical integration. While it is not yet possible to fully predict the response of an arbitrary gene network based only on its connectivity, simplified approaches allow classification of the behavior of networks, among which is the theory of monotone systems (Angeli & Sontag, 2003). A system is said to be monotone if the relation between any pair of input and output is monotone and this monotonicity is independent of the state of the rest of the system.

Gene networks are most often represented by signed graphs, where the sign of the arrow connecting two genes represents a monotone relation between the upstream and downstream gene. It is not guaranteed that the relation between any two arbitrary genes taken within a larger signed network is monotone. Nevertheless, it has been shown that gene networks are essentially monotone both in practice and in theory (Gjuvsland, Wang, Plahte, & Omholt, 2013; Sontag, 2005), in the sense that: (i) the whole network can be decomposed into a few large monotone components and (ii) the network can be made monotone by removing a small number of genes (a few among hundreds to thousands). It is important to note that the monotonicity described here is established only by the sign of the interactions. Thus, mutations changing the strength of interactions but conserving the signed regulatory architecture leave the monotonicity properties invariant.

In this work, we focus our attention on the expression level $g \in \mathbb{R}$ of a given gene in response to a vector of input signals $I \in \mathbb{R}^k$ (Fig. 1). We assume that the response $\mathcal{F}$ between the multiple inputs I and the single output g is monotone. A variable environment can be represented by a list of vectors of input signals $I_1, \ldots, I_N \in \mathbb{R}^k$ and their corresponding responses $g_1, \ldots, g_N \in \mathbb{R}$, indexed by environment. The resulting vector $G = (g_1, \ldots, g_N) \in \mathbb{R}^N$ is the vector of expression levels of a single gene in the N environments, $\mathbb{R}^N$ being the decision space.

The natural order on $\mathbb{R}$ induces a partial order between the elements $I_i \in \mathbb{R}^k$, which, given the monotonicity of $\mathcal{F}$, induces partial order constraints between the values $g_1, \ldots, g_N$. These partial order constraints in turn determine the feasible space for the vector $G \in \mathbb{R}^N$. Here we are interested in predicting the evolutionary potential of a fixed signed regulatory network topology. Under the latter condition, mutations alter the strength of the connections, leading to the modification of the response function from $\mathcal{F}$ to $\mathcal{F}'$, but the monotonicity of $\mathcal{F}$ and $\mathcal{F}'$ is the same. This way, *via* a transformation from $\mathcal{F}$ to $\mathcal{F}'$, mutations occurring during biological evolution move the system from G to G' in the feasible space constrained by the partial order.

We now introduce the optimization objectives: certain coordinates g_i must be maximized while others must be minimized, in accordance with the idea that genes can be considered either detrimental (cost of gene expression) or beneficial, depending on the environment. Searching for Pareto optimal solutions given the partial order constraints finally reveals the set of expression levels G that can be reached during the evolution of a fixed signed monotone regulatory network.

3. Notations and formulation of the problem

3.1. Pareto optimality

We consider a partially ordered set $(\Omega, \succeq)$ with the corresponding strict order $x \succ y \Leftrightarrow (x \succeq y \text{ and } x \neq y)$. We denote $Par(X)$ the strong