Mathematical Biosciences 273 (2016) 45–56

Contents lists available at ScienceDirect

Mathematical Biosciences

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

Robustness analysis of elementary flux modes generated by column generation

Hildur Æsa Oddsdóttir[a,*], Erika Hagrot[b], Véronique Chotteau[b], Anders Forsgren[a]

[a] Department of Mathematics, Optimization and Systems Theory, KTH Royal Institute of Technology, Stockholm SE-100 44, Sweden
[b] Division of Industrial Biotechnology/Bioprocess Design, KTH Royal Institute of Technology, Albanova Center, Stockholm SE-106 91, Sweden

ARTICLE INFO

Article history:
Received 7 August 2015
Revised 5 December 2015
Accepted 22 December 2015
Available online 31 December 2015

Keywords:
Metabolic network
Robust optimization
Least-squares
Elementary flux mode
Chinese hamster ovary cell

ABSTRACT

Elementary flux modes (EFMs) are vectors defined from a metabolic reaction network, giving the connections between substrates and products. EFMs-based metabolic flux analysis (MFA) estimates the flux over each EFM from external flux measurements through least-squares data fitting. The measurements used in the data fitting are subject to errors. A robust optimization problem includes information on errors and gives a way to examine the sensitivity of the solution of the EFMs-based MFA to these errors. In general, formulating a robust optimization problem may make the problem significantly harder. We show that in the case of the EFMs-based MFA, when the errors are only in measurements and bounded by an interval, the robust problem can be stated as a convex quadratic programming (QP) problem. We have previously shown how the data fitting problem may be solved in a column-generation framework. In this paper, we show how column generation may be applied also to the robust problem, thereby avoiding explicit enumeration of EFMs. Furthermore, the option to indicate intervals on metabolites that are not measured is introduced in this column generation framework. The robustness of the data is evaluated in a case-study, which indicates that the solutions of our non-robust problems are in fact near-optimal also when robustness is considered, implying that the errors in measurement do not have a large impact on the optimal solution. Furthermore, we showed that the addition of intervals on unmeasured metabolites resulted in a change in the optimal solution.

1. Introduction

A metabolic reaction network is represented by the stoichiometric matrix A, which together with the flux vector (v) gives the overall change in concentration of each metabolite (C). The rows of the stoichiometric matrix (A) refer to either external metabolites (A_x) or internal metabolites (A_i). The flux space is given by a set of vectors v that satisfy the pseudo-steady state assumption and flow direction assumption on the internal metabolites,

$$\left\{ v : \begin{bmatrix} A_i \\ -A_i \\ -I_j \end{bmatrix} v \leq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \ j \in J_{\text{irrev}} \right\}, \tag{1}$$

where I_j is a reduced identity matrix with ones only when $j \in J_{\text{irrev}}$ and J_{irrev} is the set of irreversible reactions. When all reactions in the network are irreversible, the set defined by (1) is a cone where any ray can be written as a non-negative linear combination of the extreme rays [1, Part I.4 Theorem 4.8].

EFMs contain information on how extracellular metabolites are connected by detailing which reactions are required for their uptake or production [2]. They are vectors in the flux space, each EFM includes only a minimal set of reactions and is nondecomposable [3]. Further, any vector in the flux space can be denoted as a non-negative linear combination of the EFMs [4,5],

$$v = \sum_{l=1}^{L} w_l e_l = Ew, \qquad \gamma \geq 0, \tag{2}$$

where e denotes a single EFM and the matrix E contains the EFMs as columns. In this sense the EFMs generate the flux space and are related to the definition of extreme rays in the cone (1) with only irreversible reactions. In fact when a metabolic network only has irreversible reactions the EFMs and the extreme rays of the cone (1) are equal [6]. We assume, without loss of generality, that the metabolic network has only irreversible reactions, *i.e.*, $v_j \geq 0 \, \forall j$. When the network includes reversible reactions finding all the EFMs is equivalent to finding all the extreme rays of a cone

* Corresponding author.
E-mail address: haodd@kth.se (H. Oddsdóttir).

http://dx.doi.org/10.1016/j.mbs.2015.12.009
0025-5564/

in an extended space where all reactions are irreversible [6,7]. For modest-sized networks enumeration of EFMs is possible and computer programs exist for that purpose, *e.g.,* Metatool [8]. However, with increased network size enumeration of EFMs becomes prohibitive [3]. Thus focus has shifted to identify only a subset of the EFMs [9–11].

EFMs-based MFA uses the decomposition of v given by (2) to create a macroscopic network (A_xE). The macroscopic fluxes (w) are then adjusted so that the flux in the network fit the cell specific external flux measurements (Q), *i.e.,*

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \frac{1}{2}\|Q - \mathcal{I}A_xEw\|_2^2 \\ \text{subject to} \quad & w \geq 0. \end{aligned} \tag{3}$$

The formulation given by (3) includes multiple repetitions of the same experiments, *i.e.,* if q_k are results from one repetition, k, then $Q^T = [q_1^T, \dots q_d^T]$, where d denotes the number of repetitions. $\mathcal{I}$ is a stacked identity matrix consisting of d identity matrices of size M_{ext} (number of external metabolites) or $\mathcal{I} = [I_{M_{\text{ext}}}, \dots, I_{M_{\text{ext}}}]^T$, where $I_{M_{\text{ext}}}$ is repeated d times.

EFMs-based MFA as given by (3) requires the whole set of EFMs (*i.e.,* systematic enumeration of all the EFMs) limiting the application to simplified networks. Methods that can solve the EFMs-based MFA problem without enumerating EFMs exist. One method identifies EFMs beforehand through a series of linear programming (LP) problems [12]. This method is based on the existence of a feasible flux vector v, an assumption we will examine in Appendix A. In our previous work we introduced a more integrated approach that enables identification of EFMs in conjunction with solving the EFMs-based MFA problem [13]. This method generates a subset of EFMs specifically relevant for the available experimental data. The approach was based on an optimization technique named column generation [14], in which large networks can be handled by relying on a master problem and a subproblem that are solved iteratively. The subproblem gives the master problem a new column every iteration until the solution of the subproblem indicates that the solution of the master problem is optimal to the full optimization problem. In this manner a sub-set of the EFMs relevant with respect to the extracellular flux measurements could be identified without enumeration of EFMs. In practice, only a small subset of EFMs are generated by the column generation approach. This was the case in [13] and we demonstrate here that it is also true for the robust formulation.

In this work we consider the uncertainty in the EFMs-based MFA model posed by the errors in measurements together with identifying EFMs via column generation. The experimental measurements used to calculate the fluxes in Q in the EFMs-based MFA problem (3) are prone to errors, which have been stated to reach at least 20% [15]. It is important to know how much the data uncertainty affects the solution of the optimization problem. Robust optimization can be used to consider the effects of data uncertainty on the solution of the optimization problem in a comprehensive way. For more information on robust optimization see [16] or [17]. The uncertainties could be considered by a stochastic formulation or a worst-case scenario formulation. With stochastic formulation the data uncertainty is described by probability distributions, whereas worst-case scenario formulation represents uncertainty by a set where the uncertain parameters lie. The worst-case scenario formulation has the advantage of a simpler problem formulation and that formulation of the probability distribution of the uncertainty is not needed. In this paper, we consider the worst-case scenario formulation and to simplify the exposition refer to this as robust optimization. A classical approach to assess uncertainty is sensitivity analysis where the sensitivity of the solution to errors in the data is quantified. Sensitivity analysis is a local approach, typically valid for small errors, whereas robust optimization can handle large uncertainties in a worst-case sense. Thus a robust solution provides immunization against the uncertainty present in the data. We therefore derive a robust formulation of the EFMs-based MFA, using the worst-case scenario formulation. The sensitivity of the solution with respect to measurement errors can be considered with this robust formulation. The error on each measurement is assumed bounded, and the aim is to minimize the objective function subject to the most unfavorable error in the uncertainty set. Previous work on robust least-squares mainly focuses on errors in both the measurements and the model, in general those formulations are difficult to solve [18].

The robust optimization problem can be used to evaluate the results of the EFMs-based MFA with respect to how sensitive it is to errors. That is by comparing the robust solution to the non-robust solution, where the non-robust solution is the original optimization problem, which does not include the robust formulation. If the solutions are similar, then the non-robust problem is nearly optimal to the robust problem, that is almost self robust. If they are not similar, the non-robust solution is sensitive to the errors, which is not good for the validity of the solution, especially since the errors may be rather substantial.

The purpose of this paper is to formulate the robust problem of the EFMs-based MFA, to solve the robust problem with column generation, and to apply the robust problem to case studies. In particular the aim is to see if the solution to EFMs-based MFA is sensitive to measurement errors, especially with respect to the EFMs used in the solution. The paper is outlined as follows. In Section 2 we present a robust formulation of the EFMs-based MFA, where column generation can also be applied, along with a version in which intervals for unmeasured metabolites are included. In Section 3 we present some results comparing the solutions of the robust problem to the EFM-based MFA for two case-studies.

2. Robust formulation of the EFMs-based MFA

This section contains the main results of this work, here we present an optimization problem that can be used to analyze the model's sensitivity to errors. The optimization problem is a robust formulation of the EFMs-based MFA problem, where errors in Q are taken more directly into consideration. For this purpose we make use of a technique named robust optimization [16,17]. The robust optimization problem is to minimize the residual when the errors in data give a worst-case scenario outcome, *i.e.,* the errors in data are such that the residual is maximized.

Inherent in least-squares is the assumption that the errors are bounded by the two-norm, *i.e.,* $\|\Delta Q\| \leq \beta$. In fact, when the errors are assumed bounded by the two norm, the least squares problem gives the same solution as its robust variant. However, in this work we assume that the errors in Q are bounded by an interval, a more realistic assumption with respect to the types of errors in the model. This assumption of interval errors might cause the solution to change. The interval is such that $Q_{\text{real}} = Q + \Delta Q$ where $\Delta Q_i = [\Delta q_1^T, \dots \Delta q_d^T]^T$ and $|\Delta q_{ki}| \leq \theta_{ki}|q_{ki}|$, k refers to a specific repetition and i to the metabolite. In order to simplify notation θ is stacked in the same way as Q and ΔQ, the subindex s then refers to a specific element in those vectors. Note that in general the percentage of error on each metabolite is the same for all repetitions, *i.e.,*$\theta_{k_1 i} = \theta_{k_2 i}$ for all k_1 and k_2. The robust problem is then given by

$$\underset{w \geq 0}{\text{minimize}} \ \underset{|\Delta Q_s| \leq \theta_s |Q_s|}{\text{maximize}} \quad \frac{1}{2}\|\mathcal{I}A_xEw - Q + \Delta Q\|. \tag{4}$$

Formulating robust problems as an optimization problems that can be solved is not always possible [18]. However, in Appendix B we show how problem (4) can equivalently be formulated as a