Robust Flux Balance Analysis of Metabolic Networks

Michael M. Zavlanos and A. Agung Julius

Abstract—Metabolic networks describe the set of biochemical reactions and regulatory interactions of metabolism that govern the phenotypical properties of a cell. Analysis of such networks is critical not only to promote biological knowledge, but also in drug discovery, where it can be used to identify and knockout the targeted pathways. Flux Balance Analysis (FBA) has been widely used to study metabolic networks. This powerful technique employs the reaction stoichiometries and reversibility constraints along with experimental measurements of phenotypical properties of the cell, e.g., biomass composition or ATP synthesis, to compute the fluxes of metabolites that are best manifested in the cell. Although FBA has been shown to satisfactorily capture cell behavior, its performance could be significantly improved if measurement uncertainty is introduced in the models. In this paper we propose Robust Flux Balance Analysis (RFBA) to determine optimal fluxes of metabolites for all phenotypical measurements in a given uncertainty set. We derive a least squares bi-criterion approximation of the uncertain problem and, using the S-procedure and tools from matrix analysis, we show that this is equivalent to a semidefinite program that can be solved optimally using available techniques. We illustrate our approach on synthetic metabolic networks and discuss the effect of regularization on the final solutions. Due to its convex nature, our approach can be applied to genome-scale networks.

I. INTRODUCTION

Metabolic networks map the biochemical reactions in a living cell to the flow of various chemical substances in the cell, which are called metabolites. The metabolic network of an organism can be thought of as production lines in a large scale biochemical plant. It captures the totality of metabolic reactions in which chemical substances are consumed to produce metabolic products. Analysis of such networks is critical not only to promote biological knowledge, but also in drug discovery, where it can be used to identify and knockout the targeted pathways.

Metabolic Flux Balance Analysis (FBA) [1], [2] studies the feasible and optimal reaction fluxes through the network at steady state [3], subject to structural, reversibility, and flux capacity constraints [4], [5]. Structural constraints arise from the stoichiometry of the metabolic reactions. (Ir)reversibility constraints are thermodynamic in nature and capture the direction in which chemical substances flow within a reaction. Finally, flux capacity constraints can be derived from the availability of nutrients, the existence of a knockout, and biochemical data on the maximum throughput of enzymes. Given such constraints, the flux of chemical substrates through the network is limited to a feasible region defined by a convex polytope, and the objective of FBA is to determine a feasible set of fluxes that is best manifested in the biological system under consideration. The assumption commonly made is that the metabolic system exhibits a metabolic state that is optimal in terms of cellular growth [4], [6]. Cellular growth can be represented by accumulation of cellular biomass, which is composed of cellular metabolites. If the composition of the cellular biomass is known, cellular growth can be captured by an artificial biomass reaction involving metabolites at stoichiometries defined by their contribution in biomass composition. Therefore, calculation of the optimal growth rate and the corresponding metabolic fluxes can be posed as a Linear Program (LP) [1], [2].

To date, robustness analysis of metabolic networks has primarily focused on the response of the network to structural changes, such as gene knockouts or gene deletions, and has traditionally relied on “brute force” FBA applied to different knockout combinations [7], [8]. Minimal cutset algorithms for knockout experiment design were proposed in [9], [10] and were recently extended to arbitrarily large networks [11]. The related literature also includes mixed integer linear programming approaches [12], [13], as well as a convex relaxations that scale better with the network size [14].

In this paper, we switch gears and study robustness of FBA to perturbations in the biomass composition and the metabolic flux capacities. These quantities are typically subject to measurement uncertainty, which raises the need for new Robust FBA (RFBA) techniques that return flux distributions that are optimal for worst case measurements. An important technical challenge in introducing uncertainty in FBA is that this typically appears in the stoichiometric equality constraints of the original LP, in the form of uncertain stoichiometric coefficients of an artificial biomass reaction [15], [16]. For this, we propose a least squares approximation of the original uncertain LP that results in a bi-criterion optimization problem. Using the S-procedure and tools from matrix analysis, we show that this approximation is equivalent to a semidefinite optimization problem that can be solved optimally using available techniques [17]. We illustrate our approach on synthetic data and study the effect of regularization on the final solution.

This paper is organized as follows: In Section II we describe FBA for maximization of cellular growth. In Section III we introduce measurement uncertainty and develop RFBA based on a least squares approximation of the original uncertain LP. We illustrate our approach on synthetic data in Section IV, and study the effect of regularization.
II. Flux Balance Analysis (FBA)

A. Metabolic Network Modeling at Steady State

Consider a metabolic network with \( n \) metabolites and \( m \) reactions. The \( k \)-th reaction can be written as
\[
\alpha_{1,k} A_1 + \cdots + \alpha_{n,k} A_n \rightarrow \beta_{1,k} A_1 + \cdots + \beta_{n,k} A_n,
\]
where \( A_j \) denotes the \( i \)-th metabolite, and \( \alpha_{i,k}, \beta_{i,k} \) are non-negative integers that denote the stoichiometric coefficients of the \( k \)-th reaction. Obviously, if \( A_i \) is not involved as a reactant in the \( k \)-th reaction, then \( \alpha_{i,k} = 0 \). Similarly, if \( A_i \) is not involved as a product in the \( k \)-th reaction, then \( \beta_{i,k} = 0 \). In regular reactions we have
\[
\begin{align*}
\alpha_{i,k} & \neq 0, \quad (2a) \\
\beta_{i,k} & \neq 0, \quad (2b)
\end{align*}
\]
which means that there is always some reactant and product associated with the reaction. Here we assume that all reactions are irreversible. This is done without any loss of generality, since reversible reactions can be written as two opposite irreversible reactions.

In addition to the regular reactions, we also have uptake reactions. These are reactions that can be written as
\[
i \rightarrow A_i,
\]
and model the uptake of metabolite \( A_i \) from the environment. Uptake reactions can also be expressed as in (1), without the restriction of (2a).

If we denote the concentration of the \( i \)-th metabolite as \( x_i \) and the rate of the \( k \)-th reaction as \( \omega_k \), then we can show that \( x \) and \( \omega \) are related through
\[
\frac{dx}{dt} = (\beta - \alpha)\omega, \quad \omega \geq 0
\]
where \( \alpha \) and \( \beta \) are the \( n \times m \) matrices formed by the coefficients of (1), and the symbol \( \geq \) denotes element-wise inequality.

In microbes, the transient dynamics of the metabolic network are faster than both cellular growth rates and the dynamic changes in the organism’s environment. In analyzing the network, thus, it is assumed that it is in its steady-state. In steady-state, the rates \( dx/dt \) represent the accumulation of metabolites and must be element-wise nonnegative. This is because the cell can act as a perpetual sink, but not as a perpetual source (without any uptake). Thus, in steady-state condition, the following relations hold:
\[
\begin{align*}
(\beta - \alpha)\omega - \frac{dx}{dt} &= 0, \quad (5a) \\
\omega &\geq 0, \quad \frac{dx}{dt} \geq 0. \quad (5b)
\end{align*}
\]
We can rewrite (5) in a more compact form by introducing pseudo-reactions as sinks. These are reactions that can be written as
\[
A_i \rightarrow *,
\]
We associate a sink with every metabolite. Thus, there are \( n \) pseudo-reactions. Equation (5) can, therefore, be written compactly as [14]
\[
Su = 0, \quad v \geq 0,
\]
where
\[
S \triangleq \begin{bmatrix} \beta - \alpha & -I \end{bmatrix} \in \mathbb{Z}_{+}^{n \times (m+n)}, \quad v \triangleq \begin{bmatrix} \omega \frac{dx}{dt} \end{bmatrix} \in \mathbb{R}^{m+n}.
\]

Since, typically, the number of reactions is greater than the number of metabolites, i.e., \( S \) is a wide matrix, the system (7) may have multiple solutions corresponding to flux distributions representing different metabolic states. Therefore, the null space, or the set of all feasible flux distributions, represents the capabilities of the metabolic genotype. The transport fluxes represent environmental conditions that, along with the genotype, define the metabolic state. However, obtaining all possible metabolic states for any genotype-environment interaction depends on how well the genotype and environmental factors are characterized [18].

B. Maximization of Cellular Growth

The objective of Flux Balance Analysis (FBA) is to determine a feasible metabolic state that is best manifested in the biological system under consideration. The assumption commonly made is that the metabolic system exhibits a metabolic state that is optimal under some criteria. In the case of cell growth, the objective is biomass production, i.e., the rate at which metabolic compounds are converted into biomass constituents, such as nucleic acids, proteins and lipids. Biomass production can be mathematically represented by an artificial biomass reaction [15], [16]
\[
\sum_{i=1}^{n} b_i A_i \xrightarrow{\omega} \text{Biomass}
\]
that consumes precursor metabolites \( A_i \) at stoichiometries \( b_i \) that simulate biomass production. The biomass reaction is based on experimental measurements of the biomass components \( b_i \) contained in the vector \( b \) and is scaled so that the flux through it is equal to the exponential growth rate \( \mu = \ln(2)/T \) of the organism, where \( T > 0 \) is the doubling time. Reaction (9) introduces an additional column in the stoichiometric matrix, which becomes
\[
S_b \triangleq \begin{bmatrix} \beta - \alpha & -b & -I \end{bmatrix} \in \mathbb{R}_{+}^{n \times (n+m+1)},
\]
with corresponding flux vector
\[
v_b \triangleq \begin{bmatrix} \omega^T \omega_b \omega \frac{dx}{dt} \end{bmatrix} \in \mathbb{R}^{n+m+1},
\]
where \( \omega_b \) is the rate of the artificial biomass reaction (9). Therefore, we can define an optimization problem to determine the metabolic fluxes \( v_b \) that ensure desired cell growth, dictated by precursor requirements contained in \( b \), as
\[
\begin{align*}
\text{maximize} & \quad v_b^T v_b \\
\text{subject to} & \quad S_b v_b = 0 \\
& \quad 0 \leq v_b \leq v_{\text{max}}
\end{align*}
\]
where $e_b$ is a column vector with all entries equal to zero except for the $(m+1)$-st entry that is equal to one and corresponds to the position of $\omega_b$ in $v_b$ (c.f. (11)). In problem (12), we have also included flux capacities $v_{\text{max}}$, which in the case of the precursors correspond to their actual experimentally measured concentrations for given cell growth. If no such knowledge is available, the fluxes can be unconstrained.

III. ROBUST FLUX BALANCE ANALYSIS (RFBA)

The experimentally measured biomass composition vector $b \in \mathbb{R}^n_+$ and the flux capacities $v_{\text{max}} \in \mathbb{R}^{n+m+1}$ are typically subject to uncertainty. In this section we introduce measurement uncertainty in (12) and propose a reformulation of the FBA problem that is robust with respect to worst case parameter uncertainty.

Observe first that parameter uncertainty enters (12) in the equality constraints $S_tv_b = 0$, which poses technical difficulties in finding a unique flux distribution $v_b$ that satisfies these constraints for all possible evaluations of $b$ within an uncertainty set. Therefore, we approximate (12) by the least squares bi-criterion optimization problem

$$
\begin{align*}
& \text{minimize } \epsilon ||S_tv_b||^2 - (1-\epsilon)e_b^Tv_b, \\
& \text{subject to } 0 \leq v_b \leq v_{\text{max}},
\end{align*}
$$

with $S_b$ defined as in (10) for biomass composition vector $b_i$. Therefore, we can define the robust counterpart of problem (13) by

$$
\begin{align*}
& \text{minimize } r(S_b, v_b, \rho) - (1-\epsilon)e_b^Tv_b, \\
& \text{subject to } 0 \leq v_b \leq v_{\text{max}},
\end{align*}
$$

where

$$
r(S_b, v_b, \rho) = \max_{||x||_2 \leq \rho} ||S_b(x)v_b||_2
$$

denotes the worst case stoichiometric error. Without the presence of the objective $e_b^Tv_b$, problem (14) is also known as a Robust Least Squares problem [19], [20]. Let

$$
M(v_b) = [S_{b_1}, v_b \ldots S_{b_p}, v_b]
$$

and define the quantities

$$
F = M^T(v_b), \quad g = M^T(b_0)v_b, \quad h = ||S_bv_b||^2_2.
$$

Then,

$$
||S_b(x)v_b||^2_2 = ||S_{b_0}v_b + \sum_{i=1}^p \xi_iS_{b_i}v_b||^2_2
$$

where $\epsilon \in [0, 1]$ is a tuning (regularization) parameter [17] that regulates the relative contribution of the two objectives $||S_tv_b||^2_2$ and $-e_b^Tv_b$ in (13). In problem (13) we trade exact satisfaction of the stoichiometric equality constraints for maximization of biomass. In choosing $\epsilon$, we should ensure that equality violation is not too large, i.e., that the stoichiometric error $||S_tv_b||^2_2$ is small enough. We will study sensitivity of the solution of problem (13) to the tuning parameter $\epsilon$ in Section IV.

A. Uncertainty in the Biomass Composition

To model the uncertainty in the biomass composition vector $b \in \mathbb{R}^n_+$, assume that there are $p$ available measurements $\{b_i\}_{i=1}^p \in \mathbb{R}^n_+$ of the biomass composition and for every $\xi \in \mathbb{R}^p$ with $||\xi||_2 \leq \rho$ let

$$
b(\xi) = b_0 + \sum_{i=1}^p \xi_ib_i
$$

where $b_0 = \frac{1}{p}\sum_{i=1}^p b_i$ denotes a mean biomass composition vector ($b_0$ can also be taken the zero vector $0^n$). Then, the stoichiometric matrix $S_b$ becomes

$$
S_b(\xi) = S_{b_0} + \sum_{i=1}^p \xi_iS_{b_i},
$$

1Hereafter, $b_i$ will denote the $i$-th measurement of the biomass composition vector, rather than the stoichiometric coefficient of the $i$-th metabolite in the biomass reaction previously defined in (9). The stoichiometric coefficient of the $j$-th metabolite of the $i$-th measurement of the biomass composition vector will be denoted by $b_{ij}$.

$^2$We write $X \succeq 0$ if and only if the symmetric matrix $X \in \mathbb{S}^n$ belongs in the positive semidefinite cone, defined by $\mathbb{S}^+_n = \{ X \in \mathbb{S}^n | X \succeq 0 \}$. 

$$
\begin{align*}
& \text{minimize } \epsilon \lambda - (1-\epsilon)e_b^Tv_b, \\
& \text{subject to } 0 \leq v_b \leq v_{\text{max}},
\end{align*}
$$

$$
\begin{align*}
& \sum_{i=1}^p \rho^2 \geq \lambda + \sum_{i=1}^p \xi_i, \\
& \sum_{i=1}^p \frac{\rho^2}{\xi_i} \geq \frac{\lambda}{2},
\end{align*}
$$

for some $\tau \geq 0$. Therefore, problem (16) can be equivalently written as

$$
\begin{align*}
& \text{minimize } \epsilon \lambda - (1-\epsilon)e_b^Tv_b, \\
& \text{subject to } 0 \leq v_b \leq v_{\text{max}},
\end{align*}
$$

$$
\begin{align*}
& \lambda - \rho^2 \tau - h - g^T(I - F) \geq 0, \\
& 0 \leq v_b \leq v_{\text{max}}.
\end{align*}
$$

1Hereafter, $b_i$ will denote the $i$-th measurement of the biomass composition vector, rather than the stoichiometric coefficient of the $i$-th metabolite in the biomass reaction previously defined in (9). The stoichiometric coefficient of the $j$-th metabolite of the $i$-th measurement of the biomass composition vector will be denoted by $b_{ij}$. 

$^2$We write $X \succeq 0$ if and only if the symmetric matrix $X \in \mathbb{S}^n$ belongs in the positive semidefinite cone, defined by $\mathbb{S}^+_n = \{ X \in \mathbb{S}^n | X \succeq 0 \}$. 

$$
\begin{align*}
& \text{minimize } \epsilon \lambda - (1-\epsilon)e_b^Tv_b, \\
& \text{subject to } 0 \leq v_b \leq v_{\text{max}},
\end{align*}
$$

$$
\begin{align*}
& \lambda - \rho^2 \tau - h - g^T(I - F) \geq 0, \\
& 0 \leq v_b \leq v_{\text{max}}.
\end{align*}$$
Since
\[
\begin{bmatrix}
\lambda - \rho^2 \tau & 0 \\
0 & \tau I
\end{bmatrix}
- \begin{bmatrix}
(S_{b_0} v_b)^T \\
M^T(v_b)
\end{bmatrix}
\begin{bmatrix}
\tau I \\
I
\end{bmatrix}
\begin{bmatrix}
S_{b_0} v_b \\
M(v_b)
\end{bmatrix},
\]
we can apply Schur complements to problem (16) to obtain
\[
\begin{align*}
& \text{minimize} \quad \epsilon \lambda - (1 - \epsilon) \epsilon_i^T v_b \\
& \text{subject to} \quad \mathcal{F}(\lambda, \tau, v_b) \geq 0 \\
& \quad 0 \leq \rho^2 \tau \leq \lambda \\
& \quad 0 \leq v_b \leq v_{max}
\end{align*}
\]
where
\[
\mathcal{F}(\lambda, \tau, v_b) = \begin{bmatrix}
\lambda - \rho^2 \tau & 0 \\
0 & \tau I
\end{bmatrix}
- \begin{bmatrix}
(S_{b_0} v_b)^T \\
M^T(v_b)
\end{bmatrix}
\begin{bmatrix}
\tau I \\
I
\end{bmatrix}
\begin{bmatrix}
S_{b_0} v_b \\
M(v_b)
\end{bmatrix}.
\]
Problem (17) is a semidefinite program in variables \((\lambda, \tau, v_b)\) and can be solved using available techniques [17, 21].

B. Uncertainty in the Metabolic Flux Capacities

To introduce uncertainty in the flux capacities \(v_{max} \in \mathbb{R}^{n+m+1}_+\), assume that there are \(q\) available measurements \(\{v_{max}^{k}\}_{k=1}^{q} \in \mathbb{R}^{n+m+1}_+\), and for every \(\zeta \in \mathbb{R}^q\) with \(\|\zeta\|_2 \leq \eta\) and every flux \(i = 1, \ldots, n + m + 2\) let
\[
e_i^T v_{max}(\zeta) = e_i^T v_{max}^0 + \sum_{k=1}^{q} \zeta_k e_i^T v_{max}^k,
\]
where \(v_{max}^0 = \frac{1}{q} \sum_{k=1}^{q} v_{max}^k\) and \(e_i \in \mathbb{R}^{n+m+1}_+\) is a column vector with all entries equal to zero except for the \(i\)-th entry that is equal to one. The inner product of \(v_{max}\) with \(e_i\) corresponds to the capacity of the \(i\)-th flux. Therefore, (17) can be reformulated to account for uncertainty in the flux capacities as
\[
\begin{align*}
& \text{minimize} \quad \epsilon \lambda - (1 - \epsilon) \epsilon_i^T v_b \\
& \text{subject to} \quad \mathcal{F}(\lambda, \tau, v_b) \geq 0 \\
& \quad 0 \leq \rho^2 \tau \leq \lambda \\
& \quad 0 \leq e_i^T v_b \leq \inf_{\|\zeta\|_2 \leq \eta} \{e_i^T v_{max}(\zeta)\}
\end{align*}
\]
for all fluxes \(i = 1, \ldots, n + m + 2\). Let \(N = [v_{max}^1 \ldots v_{max}^q]\) and observe that
\[
\begin{align*}
\inf_{\|\zeta\|_2 \leq \eta} \{e_i^T v_{max}(\zeta)\} &= e_i^T v_{max}^0 + \inf_{\|\zeta\|_2 \leq \eta} \{e_i^T N \zeta\} \\
&= e_i^T v_{max}^0 + \eta \inf_{\|\zeta\|_2 \leq \eta} \{e_i^T N(\zeta/\eta)\} \\
&= e_i^T v_{max}^0 + \eta \inf_{\|\zeta\|_2 \leq \eta} \{e_i^T N(\zeta)\} \\
&= e_i^T v_{max}^0 + \eta \inf_{\|\zeta\|_2 \leq \eta} \{e_i^T N(\zeta)\} \\
&= e_i^T v_{max}^0 - \eta \sup_{\|\zeta\|_2 \leq \eta} \{e_i^T N(\zeta)\} \\
&= e_i^T v_{max}^0 - \eta \|N^T e_i\|_1,
\end{align*}
\]
where \(\|u\|_* = \sup \{u^T x \mid \|x\|_2 \leq 1\}\) denotes the dual norm of \(u\), which can be interpreted as the operator norm of \(z^T\) if it is considered a \(1 \times n\) matrix.\(^3\) Substituting in (19) we get
\[
\begin{align*}
& \text{minimize} \quad \epsilon \lambda - (1 - \epsilon) \epsilon_i^T v_b \\
& \text{subject to} \quad \mathcal{F}(\lambda, \tau, v_b) \geq 0 \\
& \quad 0 \leq \rho^2 \tau \leq \lambda \\
& \quad 0 \leq e_i^T v_b \leq \eta \text{vec}(\mathcal{N}_{e_i}(1, \ldots, n+2))
\end{align*}
\]
where \(\text{vec}(\{x_1, \ldots, x_k\})\) denotes a column vector with elements \(x_1, \ldots, x_k\). Problem (20) is a semidefinite program and can be solved using available techniques [17, 21].

IV. SIMULATION RESULTS

In this section, we illustrate problem (20) on synthetic metabolic networks and study the effect of regularization (tuning parameter \(\epsilon\)) on the final solution. In particular, consider a metabolic network consisting of \(n = 10\) metabolites labeled \(\{A_i\}_{i=1}^{10}\), and \(m = 5\) reversible reactions
\[
\begin{align*}
R_1 : \quad 2A_1 + A_2 &\rightarrow A_4, \\
R_2 : \quad A_3 + A_4 &\rightarrow A_5, \\
R_3 : \quad A_1 + 2A_5 &\rightarrow 2A_6 + A_7, \\
R_4 : \quad 2A_3 + A_6 + A_7 &\rightarrow 2A_8 + A_9, \\
R_5 : \quad 2A_4 + A_9 &\rightarrow A_{10}, \\
&\quad \quad \quad + A_8 \rightarrow A_9.
\end{align*}
\]
three irreversible uptake reactions
\[
\begin{align*}
R_6 : \quad * &\rightarrow A_1, \\
R_7 : \quad * &\rightarrow A_2, \\
R_8 : \quad * &\rightarrow A_3,
\end{align*}
\]
\(^3\)Note that \(\inf \{X\} = -\sup \{-X\}\) for any set \(X\), where \(-X = \{-x \mid x \in X\}\).
and an artificial biomass reaction

\[ R_9 : \quad b_{1,5}A_5 + b_{1,6}A_6 + b_{1,8}A_8 + b_{1,9}A_9 \rightarrow \text{Biomass} \]  

with \( b_i = [0 \ 0 \ 0 \ 0 \ b_{i,5} \ b_{i,6} \ 0 \ b_{i,8} \ b_{i,9} \ 0]^T \)

the \( i \)-th measurement of the biomass composition vector, for \( i = 1, \ldots, p \) with \( p = 10 \) (Fig. 1). We assume that \( b_i \) has mean \([0 \ 0 \ 0 \ 0 \ 3 \ 1 \ 0 \ 2 \ 1 \ 0]^T\) and that every one of its entries is subject to zero mean and 0.5 variance gaussian noise. Every biomass composition vector \( b_i \) is normalized so that \( b_{i,5} + b_{i,6} + b_{i,8} + b_{i,9} = 1 \). Let \( T = 1h \) be the doubling time of the organism, so that the growth rate is \( \mu = \ln(2)/T = 0.69b^{-1} \). Then, the stoichiometric matrix of the network under consideration is shown in Table I. Furthermore, we assume that all fluxes are unconstrained, except for the uptake fluxes of reactions \( R_6, R_7 \) and \( R_8 \) that are upper bounded by 0.1. For simplicity, we assume that these bounds are deterministic.

We evaluated the performance problem (20) for different values of the tuning parameter \( \epsilon \in [0, 1] \) and \( \rho = 1.5 \). Simulations were performed in MATLAB using the cvx toolbox for disciplined convex programming [21]. Fig. 2 illustrates as a function of the tuning parameter \( \epsilon \in [0, 1] \) (a) the biomass objective \( e^T_i v_{b,i} \), (b) the upper bound \( \lambda \) on the worst case stoichiometric error \( r(S_b, v_b, \rho) \) defined in (15), and (c) the stoichiometric errors \( \|S_b v_b\|_2 \) for all measured biomass compositions \( b_i \), with \( i = 1, \ldots, p \). Observe that the values of \( \|S_b v_b\|_2 \) are always upper bounded by \( \lambda \), since \( \rho = 1.5 \geq 1 \). In other words, the set \( \{\|S_b(\xi) v_b\|_2 \mid \|\xi\|_2 \leq \rho\} \) includes the errors \( \|S_b v_b\|_2 \) for all measured biomass compositions (15). This is not necessarily the case if \( \rho < 1 \).

For \( \epsilon = 0 \) the stoichiometric equality constraints are ineffective and, therefore, the resulting value of \( e^T_i v_{b,i} \) has no biological meaning. We are interested in regions of the plot where \( \lambda \) is minimum and \( e^T_i v_{b,i} \) is maximum. Note that \( \lambda \) can not become identically zero, since there does not exist a unique nontrivial flux vector \( v_b \) for which \( \|S_b v_b\|_2 = 0 \) for all biomass compositions \( b \in \{b_0 + \sum_{i=1}^{p} \xi_i b_i \mid \|\xi\|_2 \leq \rho\} \).

From Fig. 2 we see that \( \lambda \) is almost at its minimum for \( \epsilon \geq 0.85 \). Since \( e^T_i v_{b,i} \) decreases rapidly as \( \epsilon \) increases beyond 0.85, we choose \( \epsilon = 0.85 \) to obtain \( e^T_i v_{b,i} = 0.1593 \). The flux vector \( v_b \) obtained by the solution of problem (20) for \( \epsilon = 0.85 \) is shown in Table II. Note that for this value of \( \epsilon = 0.85 \) there is maximum uptake of metabolites \( A_1 \) and \( A_3 \) (fluxes \( \omega_6 \) and \( \omega_8 \)) equal to the upper bound 0.1, which is distributed among the five reactions to result in metabolites that maximize biomass. All steady state metabolite concentrations are effectively constant (\( dx_i/dt = 0 \)), except for the concentrations of \( A_2 \) and \( A_9 \) that slightly increase. This increase is due to the stoichiometries in the network that prevent these metabolites from being fully consumed to produce biomass. The values of the errors \( \|S_b v_b\|_2 \) for \( i = 1, \ldots, p \) are small, ranging between 0.01–0.02 in value.

**Remark 4.1 (\( \epsilon \to 1 \))**: Observe that as \( \epsilon \to 1 \), both \( e^T_i v_{b,i} \) \( \to 0 \) and \( \|S_b v_b\|_2 \) \( \to 0 \), for all measurements \( i = 1, \ldots, p \) (Fig. 2). This behavior is justified, since as \( \epsilon \to 1 \) the biomass objective \( e^T_i v_{b,i} \) becomes effectively inactive and the optimization problem (20) determines fluxes \( v_b \) to only minimize the worst case stoichiometric error \( r(S_b, v_b, \rho) \). Numerically, a sparse flux vector \( v_b \) as the one shown in Table II is not optimal as a minimizer of \( r(S_b, v_b, \rho) \). In fact, if \( \epsilon = 1 \), the flux vector \( v_b \) returned by (20) for the metabolic network of Fig. 1 is rather dense with

---

**TABLE I**

Stoichiometric matrix \([\beta - \alpha \mid -b]\) for the metabolic network shown in Fig. 1. The reactions \( R_k^+ \) indicate opposite directionality with respect to the reactions \( R_k \), and are introduced to model reversibility (21). Reactions \( R_k^+ \) through \( R_k^8 \) specify model uptake of metabolites from the environment (22), while reaction \( R_0 \) corresponds to the artificial biomass reaction (23).

| \( A \) | \( R_1 \) | \( R_2 \) | \( R_3 \) | \( R_4 \) | \( R_5 \) | \( R_6 \) | \( R_7 \) | \( R_8 \) | \( R_0 \) |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| \( A_1 \) | -2 | 0 | -1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| \( A_2 \) | -1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| \( A_3 \) | 0 | -1 | 0 | -2 | 0 | 0 | -1 | 0 | -2 | 0 |
| \( A_4 \) | 1 | -1 | 0 | 0 | -2 | -1 | 1 | 0 | 0 | 2 |
| \( A_5 \) | 0 | 1 | -2 | 0 | 0 | 0 | -1 | 2 | 0 | 0 |
| \( A_6 \) | 0 | 0 | 2 | -1 | 0 | 0 | 0 | -2 | 1 | 0 |
| \( A_7 \) | 0 | 0 | 1 | -1 | 0 | 0 | 0 | -1 | 1 | 0 |
| \( A_8 \) | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | -2 | 0 |
| \( A_9 \) | 0 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | -1 | 1 |
| \( A_{10} \) | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | -1 |
very little biomass is produced (Table III). As expected, all resources are consumed to produce metabolites $A_1, \ldots, A_{10}$ and obtain a dense flux vector $v_9$. The stoichiometric errors $\|S_o v_9\|_2^P$ in this case are of the order of $10^{-7}$, as expected (but still nonzero).

V. Conclusions

Metabolic Flux Balance Analysis (FBA) is a powerful optimization-based technique that studies the feasible and optimal reaction fluxes through the network at steady state, subject to structural, reversibility, and flux capacity constraints. Among the large number of possible flux distributions, FBA determines the one that is best manifested in the system under consideration. The assumption commonly made is that the metabolic systems exhibit a metabolic state that is optimal in terms of cellular growth, which is typically represented by accumulation of cellular biomass.

In this paper, we proposed Robust Flux Balance Analysis (RFBA) to account for uncertainty in the biomass composition and flux capacities. We showed that flux distributions that are robust to worst case parameter uncertainty can be obtained by the solution of a bi-criterion semidefinite program, which can be solved to optimality using available techniques and scales well to large networks due its convex nature. We illustrated our approach on synthetic data and studied the effect of regularization on the final solution. Future work involves application of our method to real experimental data and comparison with existing techniques.

REFERENCES


TABLE II

<table>
<thead>
<tr>
<th>Flux</th>
<th>Value</th>
<th>Flux</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1 = \omega(R_1)$</td>
<td>0.0448</td>
<td>$dx_1/dt$</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\omega_2 = \omega(R_2)$</td>
<td>0.0701</td>
<td>$dx_2/dt$</td>
<td>0.0315</td>
</tr>
<tr>
<td>$\omega_3 = \omega(R_3)$</td>
<td>0.0129</td>
<td>$dx_3/dt$</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\omega_4 = \omega(R_4)$</td>
<td>0.0156</td>
<td>$dx_4/dt$</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\omega_5 = \omega(R_5)$</td>
<td>-0.0101</td>
<td>$dx_5/dt$</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\omega_6 = \omega(R_6)$</td>
<td>0.1000</td>
<td>$dx_6/dt$</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\omega_7 = \omega(R_7)$</td>
<td>0.0764</td>
<td>$dx_7/dt$</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\omega_8 = \omega(R_8)$</td>
<td>0.1000</td>
<td>$dx_8/dt$</td>
<td>0.0000</td>
</tr>
<tr>
<td>Biomass Reaction</td>
<td>0.1593</td>
<td>$dx_9/dt$</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

TABLE III

<table>
<thead>
<tr>
<th>Flux</th>
<th>Value</th>
<th>Flux</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1 = \omega(R_1)$</td>
<td>0.0327</td>
<td>$dx_1/dt$</td>
<td>0.0131</td>
</tr>
<tr>
<td>$\omega_2 = \omega(R_2)$</td>
<td>0.0234</td>
<td>$dx_2/dt$</td>
<td>0.0400</td>
</tr>
<tr>
<td>$\omega_3 = \omega(R_3)$</td>
<td>0.0091</td>
<td>$dx_3/dt$</td>
<td>0.0383</td>
</tr>
<tr>
<td>$\omega_4 = \omega(R_4)$</td>
<td>0.0057</td>
<td>$dx_4/dt$</td>
<td>0.0058</td>
</tr>
<tr>
<td>$\omega_5 = \omega(R_5)$</td>
<td>0.0018</td>
<td>$dx_5/dt$</td>
<td>0.0052</td>
</tr>
<tr>
<td>$\omega_6 = \omega(R_6)$</td>
<td>0.0876</td>
<td>$dx_6/dt$</td>
<td>0.0125</td>
</tr>
<tr>
<td>$\omega_7 = \omega(R_7)$</td>
<td>0.0727</td>
<td>$dx_7/dt$</td>
<td>0.0034</td>
</tr>
<tr>
<td>$\omega_8 = \omega(R_8)$</td>
<td>0.0730</td>
<td>$dx_8/dt$</td>
<td>0.0114</td>
</tr>
<tr>
<td>Biomass Reaction</td>
<td>5.27 $\cdot 10^{-6}$</td>
<td>$dx_9/dt$</td>
<td>0.0039</td>
</tr>
</tbody>
</table>

be obtained by the solution of a bi-criterion semidefinite problem, which can be solved to optimality using available techniques and scales well to large networks due its convex nature. We illustrated our approach on synthetic data and studied the effect of regularization on the final solution. Future work involves application of our method to real experimental data and comparison with existing techniques.