Convex representation of metabolic networks with Michaelis-Menten kinetics

Josh A. Taylor[a], Alain Rapaport[b], Denis Dochain[c]

[a] *The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada*
[b] *MISTEA, Université de Montpellier, INRAE, Institut Agro, France*
[c] *Université Catholique de Louvain, Belgium*

Abstract

Polyhedral models of metabolic networks are computationally tractable and can predict some cellular functions. A longstanding challenge is incorporating metabolites without losing tractability. In this paper, we do so using a new second-order cone representation of the Michaelis-Menten kinetics. The resulting model consists of linear stoichiometric constraints alongside second-order cone constraints that couple the reaction fluxes to metabolite concentrations.

We formulate several new problems around this model: conic flux balance analysis, which augments flux balance analysis with metabolite concentrations; dynamic conic flux balance analysis; and finding minimal cut sets of networks with both reactions and metabolites. Solving these problems yields information about both fluxes and metabolite concentrations. They are second-order cone or mixed-integer second-order cone programs, which, while not as tractable as their linear counterparts, can nonetheless be solved at practical scales using existing software.

Keywords: Michaelis-Menten kinetics, Second-order cone programming, Flux balance analysis, Minimal cut set, Metabolite concentrations

1. Introduction

The structure of a metabolic network contains useful information about its cellular functions. Two techniques for analyzing this structure are

- flux balance analysis (FBA), in which optimization is used to predict reaction fluxes [1], and
- minimal cut set (MCS) analysis, which attempts to find critical subsets of reactions that, when removed, disable certain functions [2].

Standard formulations of FBA and MCS analysis are based on a linear approximation in which only the reaction fluxes are variables. The benefit of this simplification is that FBA is a linear program (LP), which can be reliably solved at large scales, and powerful analytical tools like Farkas' Lemma are available for finding the MCSs of polyhedral systems. Quoting [1], "FBA has limitations, however. Because it does not use kinetic parameters, it cannot predict metabolite concentrations." For the same reason, MCS analysis cannot explicitly identify critical metabolites.

In this paper, we augment FBA and MCS analysis with Michaelis-Menten kinetics and metabolite concentrations. Using the results of [3], we represent the Michaelis-Menten kinetics as a second-order cone (SOC) constraint. This leads to several original problem formulations.

- *Conic FBA (CFBA).* CFBA predicts both reaction fluxes and metabolite concentrations in steady state. It is a single SOC program (SOCP), which, while not as tractable as LP, can be solved at practical scales [4]. We also formulate dynamic CFBA, in which the SOC representations of the reaction kinetics are used in dynamic FBA [5].
- *Conic MCS (CMCS).* A CMCS is a cut set through a network of paths from reactions to metabolites, as specified by the stoichiometric matrix, and from metabolites to reactions, as specified by the Michaelis-Menten kinetics. To solve for CMCSs, we follow the strategy of [6], which uses Farkas' Lemma [4] and results from [7, 8] on irreducible infeasible subsystems (IIS) to identify MCSs. We generalize this to CMCSs using the recent results of [9] on the IISs of semidefinite systems. We also formulate a linear approximation that, due to the discrete nature of cut sets, produces similar results.

We remark that CFBA and CMCS analysis might not produce better predictions of reaction fluxes than existing methods. The main benefit of CFBA and CMCS analysis is that they incorporate metabolite concentrations and reaction kinetics while retaining much of the tractability of standard linear formulations.

We now describe how our contributions relate to the existing literature. Using LP to analyze metabolic networks was first suggested in [10]. Since then FBA has gained wide acceptance [11, 1], and is available in open source implementations [12, 13]. Dynamic FBA (DFBA) is an extension that incorporates reaction kinetics, which are typically nonlinear, and metabolite concentrations, potentially

as well as other transient information such as reprogramming and light intensity. Reference [5] first formulated the two main types of DFBA. The 'dynamic optimization approach' is a large nonlinear program, in which the reaction kinetics constrain the fluxes through time. In the 'static optimization approach', one solves a sequence of LPs and integrates the solution between time periods. There have been several refinements such as putting the optimization directly into the ODE simulation [14], and using lexicographic optimization to improve robustness when the LP has multiple solutions [15].

CFBA is similar to DFBA in that it also captures metabolite concentrations and reaction kinetics, but only when modeled as Michaelis-Menten. They differ in that CFBA is in steady state, and hence does not capture transients. If one approximates the biomass concentration as a constant, or approximates Michaelis-Menten with the Contois function [16], one can formulate the dynamic optimization approach of [5] as an SOCP. We refer to this as Dynamic CFBA. Dynamic CFBA can capture transients and accommodate time-varying parameters. CFBA and dynamic CFBA are not necessarily more accurate than dynamic FBA, but are more tractable in that they are single SOCPs.

CFBA is also related to resource balance analysis (RBA) [17, 18], a more general problem that can predict fluxes, metabolites, macromolecular cellular processes, and proteins. RBA differs from CFBA in that it does not contain the Michaelis-Menten function or any other nonlinearities, and as a result is an LP. In principle, our SOC representation of the Michaelis-Menten function could be incorporated into RBA, leading to an SOCP.

Another relevant literature stream focuses on the analysis of pathways through metabolic networks (see, e.g., [19, 20]). In the linear case, the nonnegativity and stoichiometric constraints form a polyhedral cone, the extreme rays of which correspond to the elementary flux modes of the metabolic network [21]. It is not clear that this perspective extends to our setup because SOC constraints are nonpolyhedral, and the SOC representation of the Michaelis-Menten kinetics is not in fact a cone. The Minkowski-Weyl Theorem states that a cone is finitely generated if and only if it is polyhedral [22]. Therefore, such a system could have an infinite number of elementary modes if it has any at all. If elementary modes do exist, at present we are not aware of any reliable techniques for obtaining them.

For these reasons, we focus on the adjacent problem of identifying MCSs. Reference [6] showed that the MCSs of a metabolic network are the elementary modes of a dual network specified by Farkas' Lemma [4]. We make use of the results of [9] to generalize this strategy to networks with metabolite concentrations coupled to the reaction fluxes through Michaelis-Menten kinetics. By then using a linear approximation of the Michaelis-Menten kinetics, we recover the use of tools for polyhedral systems, which we find produce more reliable results.

The paper is organized as follows. Section 2 reviews metabolic network modeling and the SOC representation of the Michaelis-Menten kinetics. Section 3 presents FBA, CFBA, and dynamic CFBA. Section 4 presents the MCS analysis of [6] and our extensions to systems with Michaelis-Menten kinetics. In Section 5, we apply CFBA, dynamic CFBA, and CMCS analysis to a model of *Escherichia coli*.

2. Background

2.1. Metabolic networks

The reactions are indexed by the set $\mathcal{N}$, where $n = |\mathcal{N}|$. Let $\mathcal{R} \subseteq \mathcal{N}$ and $\mathcal{I} \subseteq \mathcal{N}$ be the sets of reversible and irreversible reactions, where $\mathcal{R} \cup \mathcal{I} = \mathcal{N}$ and $\mathcal{R} \cap \mathcal{I} = \emptyset$. There are a total of m' metabolites.

Let $z \in \mathbb{R}^{m'}$ be a vector of metabolite concentrations and $v \in \mathbb{R}^n$ a vector of fluxes due to the reactions. $S \in \mathbb{R}^{m' \times n}$ is the stoichiometric matrix. The dynamics of the metabolites are given by

$$\dot{z} = z_{\mathrm{b}} S v, \tag{1}$$

where z_{b} is the element of z corresponding to biomass. We note that while the biomass is not a metabolite, its evolution can be written as z_{b} times a linear combination of the reaction fluxes, and hence can be represented as a row of (1).

In quasi-steady state, $\dot{z} = 0$. We then have $z_{\mathrm{b}} S v = \mathbf{0}$, which, because z_{b} is a nonzero scalar, implies $Sv = \mathbf{0}$. The fluxes are subject to the bounds $\underline{v} \leq v \leq \overline{v}$. Typically, $\underline{v}_i = 0$ for $i \in \mathcal{I}$ and $-\infty$ otherwise. When $\underline{v} = \mathbf{0}$ and $\overline{v} = \infty$, the system is a polyhedral cone [21].

Quasi-steady state models generally do not explicitly model the metabolites. Here we will include a subset of them, $\mathcal{M}$, where $m = |\mathcal{M}|$. We will henceforth let $z \in \mathbb{R}^m$.

Some of the fluxes are also bounded by a nonlinear function, usually the Michaelis-Menten kinetics [23]. We denote this set by $\mathcal{Q} \subseteq \mathcal{N}$, where $q = |\mathcal{Q}|$. We expect that $q \geq m$, because otherwise there are metabolites that appear in no reaction, and thus have no coupling to the rest of the model. To simplify notation, we order $\mathcal{N}$ so that its first q elements are $\mathcal{Q}$. Let $V^{\max} \in \mathbb{R}^q$ and $K^{\mathrm{m}} \in \mathbb{R}^q$ be vectors of maximum reaction rates and Michaelis constants. We can write this bound for each reaction $i \in \mathcal{Q}$ as

$$v_i \leq \frac{V_i^{\max} z_{\sigma(i)}}{K_i^{\mathrm{m}} + z_{\sigma(i)}}. \tag{2}$$

Here $\sigma(i)$ identifies the index of the metabolite concentration appearing in reaction i. In general σ is not invertible because the same metabolite can appear in multiple reactions. Note that if i is a reversible reaction, (2) may be applied in the reverse direction by putting a minus sign in front of v_i.

When a reaction has multiple reactants, its flux may be limited by more complicated functions such as the product of several Michaelis-Menten kinetics (cf. Table IV in [24]).

We can straightforwardly generalize our notation to this case. Suppose that reaction i has p_i reactants. For $k = 1, ..., p_i$, let $\sigma_k(i)$ be the index of the kth metabolite in the reaction. Then we can write the upper bound on reaction $i \in \mathcal{Q}$ as

$$v_{ik} \leq \frac{V_{ik}^{\max} z_{\sigma_k(i)}}{K_{ik}^{\text{m}} + z_{\sigma_k(i)}} \tag{3a}$$

$$v_i \leq \prod_{k=1}^{p_i} v_{ik}. \tag{3b}$$

The intermediary variables, v_{ik}, $k = 1, ..., p_i$, will be convenient for representing (3) in SOC form.

2.2. Convex representation of the Michaelis-Menten kinetics

As shown in [3], we can represent the inequality (2) as the SOC constraint

$$\left\| \begin{bmatrix} V_i^{\max} z_{\sigma(i)} \\ K_i^{\text{m}} v_i \\ V_i^{\max} K_i^{\text{m}} \end{bmatrix} \right\| \leq V_i^{\max} z_{\sigma(i)} - K_i^{\text{m}} v_i + V_i^{\max} K_i^{\text{m}}, \tag{4}$$

for $i \in \mathcal{Q}$. One can confirm equivalence by squaring both sides and simplifying. SOCPs consisting of linear and SOC constraints can be solved at large scales [4]. We provide a brief overview of SOCP in Appendix A.

Suppose now there are multiple reactants. As with (2), we can write (3a) as an SOC constraint in the form of (4). Unfortunately, (3b) is nonconvex. We can instead approximate it with

$$v_i \leq \prod_{k=1}^{p_i} v_{ik}^{\pi_{ik}}, \tag{5}$$

which has an SOC representation if $\sum_{k=1}^{p_i} \pi_{ik} \leq 1$ [25]. When $\pi_{ik} = 1/p_i$ for $k = 1, ..., p_i$, it is the geometric mean. (5) is ad hoc in that it cannot in general be derived from first principles. However, we believe it could be a useful approximation because there is considerable room to tune the fit via the exponents and Michaelis-Menten parameters; and, similar ad hoc expressions have been used in the past when Michaelis-Menten alone was not sufficiently descriptive (cf. Table IV in [24]).

In the simple case where $p_i = 2$ and $\pi_{i1} = \pi_{i2} = 1/2$, (5) is a hyperbolic constraint with SOC representation

$$\left\| \begin{bmatrix} 2v_i \\ v_{i1} - v_{i2} \end{bmatrix} \right\| \leq v_{i1} + v_{i2}. \tag{6}$$

Sometimes LP is more practical to work with than SOCP, e.g., when there are numerous discrete constraints or commercial solvers are too expensive. Fortunately an SOCP can always be approximated to arbitrary accuracy with an LP, albeit a potentially large one. Reference [26] provides a constructive procedure for approximating a generic SOC constraint with a family of linear constraints. Alternatively, the right hand side of (2) can be straightforwardly approximated by a family of line segments.

3. Flux balance analysis

In this section, we first review the standard formulation of FBA, and then formulate CFBA and Dynamic CFBA. We also use the duals to derive several sensitivities.

3.1. Linear FBA

The below LP is a basic FBA routine.

$$\max_{v} \quad c^\top v \tag{7a}$$

$$\text{subject to} \quad Sv = \mathbf{0} \tag{7b}$$

$$\underline{v} \leq v \leq \overline{v}. \tag{7c}$$

Here $c \in \mathbb{R}^n$ selects and/or weights the fluxes for maximization. Let $\mathcal{F}$ denote the optimal objective value.

Let $\lambda \in \mathbb{R}^m$ be the vector of dual multipliers of (7b), and let $\delta_{\text{L}} \in \mathbb{R}_+^n$ and $\delta_{\text{U}} \in \mathbb{R}_+^n$ be the vectors of dual multipliers associated with the upper and lower bounds in (7c). The dual of (7) is

$$\min_{\lambda, \delta_{\text{L}}, \delta_{\text{U}}} \quad \overline{v}^\top \delta_{\text{U}} - \underline{v}^\top \delta_{\text{L}} \tag{8a}$$

$$\text{subject to} \quad S^\top \lambda = c + \delta_{\text{L}} - \delta_{\text{U}} \tag{8b}$$

$$\delta_{\text{L}} \geq \mathbf{0}, \delta_{\text{U}} \geq \mathbf{0}. \tag{8c}$$

The dual variables, or shadow prices, can be interpreted as the sensitivities of the objective in (7) to changes in the constraints. More precisely, for each $i \in \mathcal{N}$,

$$\delta_{\text{L}i} = \frac{d\mathcal{F}}{d\underline{v}_i}, \quad \delta_{\text{U}i} = \frac{d\mathcal{F}}{d\overline{v}_i}.$$

λ is similarly interpretable as the sensitivity of the objective to perturbations to (7b). This, for example, can be used to determine which reactions are most influential [27, 28].

We can also obtain insight by observing that the objectives of (7) and (8) must be equal due to strong duality:

$$v^\top c = \overline{v}^\top \delta_{\text{U}} - \underline{v}^\top \delta_{\text{L}}.$$

This breaks down the optimal objective into contributions from each constraint. In the common case where we maximize a single reaction, v_s, and $\underline{v} = \mathbf{0}$, this simplifies to $v_s = \delta_{\text{U}}^\top \overline{v}$, a convex combination of the upper flux bounds.

3.2. Conic FBA

We now generalize the FBA (7) to include metabolite concentrations, which are coupled to the reactions by the Michaelis-Menten kinetics.

Let $\underline{z} \in \mathbb{R}^m$ and $\overline{z} \in \mathbb{R}^m$ be vectors of upper and lower bounds on the metabolite concentrations. Let $d \in \mathbb{R}^m$ be a vector weighting the metabolites in the objective.

Consider the below optimization.

$$\max_{v,z} \quad c^\top v + d^\top z \tag{9a}$$
$$\text{subject to} \quad Sv = \mathbf{0} \tag{9b}$$
$$\underline{v} \le v \le \overline{v} \tag{9c}$$
$$\underline{z} \le z \le \overline{z} \tag{9d}$$
$$v_i \le \frac{V_i^{\max} z_{\sigma(i)}}{K_i^{\mathrm{m}} + z_{\sigma(i)}}, \quad i \in \mathcal{Q}. \tag{9e}$$

We make the following observations.

- This is an SOCP if we write (9e) in the form of (4).
- (9e) is only meaningful if the corresponding metabolite concentrations are minimized in the objective, i.e., $d < 0$. Otherwise, it is trivially optimal to fix the metabolites at their maximum concentrations, $\overline{z}$, in which case (9e) can be represented as an upper bound on v in (9c).
- (9) allows for both internal and external metabolites [29]. However, if the external metabolites are buffered, which is to say fixed in our model, then the corresponding Michaelis-Menten function in (9e) is also fixed, and is then more efficiently represented as part of the vector $\overline{v}$.

We now derive the dual. For matrices A and B of the same size, let $A \circ B$ denote the element-wise product. We let $A_{i:}$ and $A_{:j}$ denote the ith row and j^{th} column of A.

Let $\lambda \in \mathbb{R}^m$ be the vector of dual multipliers of (9b), $\delta_{\mathrm{L}} \in \mathbb{R}^n_+$ and $\delta_{\mathrm{U}} \in \mathbb{R}^n_+$ of (9c), and $\gamma_{\mathrm{L}} \in \mathbb{R}^m_+$ and $\gamma_{\mathrm{U}} \in \mathbb{R}^m_+$ of (9d).

Let $\Phi \in \mathbb{R}^{3\times q}$ and $\phi \in \mathbb{R}^q_+$. For a given reaction $i \in \mathcal{Q}$, the constraint (9e) (in SOC form (4)) has dual multipliers $\Phi_{:i} \in \mathbb{R}^3$ and $\phi_i \in \mathbb{R}_+$, which satisfy the SOC constraint $\|\Phi_{:i}\| \le \phi_i$.

Let $M \in \mathbb{R}^{q\times m}$ be such that $M_{ij} = 1$ if $\sigma(i) = j$ and zero otherwise. The interpretation of M is complementary to that of S. S encodes a network representing how the reactions influence the evolution of the metabolite concentrations. Similarly, M encodes a network representing which metabolites appear in each reaction.

The dual of (9), which is also an SOCP, is below.

$$\min_{\lambda,\delta_{\mathrm{L}},\delta_{\mathrm{U}},\gamma_{\mathrm{L}},\gamma_{\mathrm{U}},\Phi,\phi} \quad \overline{v}^\top \delta_{\mathrm{U}} - \underline{v}^\top \delta_{\mathrm{L}} + \overline{z}^\top \gamma_{\mathrm{U}} - \underline{z}^\top \gamma_{\mathrm{L}} - \left(\Phi_{3:} - \phi^\top\right)\left(V^{\max} \circ K^{\mathrm{m}}\right) \tag{10a}$$
$$\text{subject to} \quad \begin{bmatrix} \left(\Phi_{2:}^\top + \phi\right) \circ K^{\mathrm{m}} \\ \mathbf{0} \end{bmatrix} + S^\top \lambda = c + \delta_{\mathrm{L}} - \delta_{\mathrm{U}} \tag{10b}$$
$$M^\top \left(\left(\Phi_{1:}^\top - \phi\right) \circ V^{\max}\right) = d + \gamma_{\mathrm{L}} - \gamma_{\mathrm{U}} \tag{10c}$$
$$\delta_{\mathrm{L}} \ge \mathbf{0}, \delta_{\mathrm{U}} \ge \mathbf{0}, \gamma_{\mathrm{L}} \ge \mathbf{0}, \gamma_{\mathrm{U}} \ge \mathbf{0} \tag{10d}$$
$$\|\Phi_{:i}\| \le \phi_i, \quad i \in \mathcal{Q}. \tag{10e}$$

Strong duality holds if a constraint qualification is satisfied, e.g., Slater's condition [4]. In this case $\lambda, \delta_{\mathrm{L}}, \lambda, \delta_{\mathrm{U}}, \gamma_{\mathrm{L}}$, and γ_{U} all have the usual LP sensitivity interpretations and complementary slackness with their respective constraints.

We can similarly interpret Φ and ϕ; see, e.g., Section 5.9.3 of [4]. ϕ_i is the sensitivity of the optimal objective value, $\mathcal{F}$, to perturbations to the right hand side of (4). $\Phi_{:i}$ is a vector of sensitivities of $\mathcal{F}$ to perturbations to each element in the left hand side. We can use the chain rule to derive sensitivities for the parameters of the Michaelis-Menten function. For $i \in \mathcal{Q}$ we have

$$\frac{d\mathcal{F}}{dV_i^{\max}} = \Phi_{1i} z_{\sigma(i)} + \Phi_{3i} K_i^{\mathrm{m}} + \phi_i \left(K_i^{\mathrm{m}} + z_{\sigma(i)}\right) \tag{11a}$$
$$\frac{d\mathcal{F}}{dK_i^{\mathrm{m}}} = \Phi_{2i} v_i + \Phi_{3i} V_i^{\max} + \phi_i \left(V_i^{\max} - v_i\right). \tag{11b}$$

As in the previous section, the objectives of (9) and (14) match if strong duality holds. If we assume that $\underline{v} = \mathbf{0}$, $\underline{z} = \mathbf{0}$, and $\overline{z} = \infty$, the equality simplifies to

$$\begin{aligned} c^\top v + d^\top z &= \overline{v}^\top \delta_{\mathrm{U}} - \left(\Phi_{3:} - \phi^\top\right)\left(V^{\max} \circ K^{\mathrm{m}}\right) \\ &= \overline{v}^\top \delta_{\mathrm{U}} - \sum_{i\in\mathcal{Q}} \left(\Phi_{3i} - \phi_i\right) V_i^{\max} K_i^{\mathrm{m}}. \end{aligned}$$

3.2.1. Alternative objectives

A shortcoming of (9) is that the value of the optimal objective—flux rates minus metabolite concentrations—does not have a clear physical meaning. We could instead maximize $d^\top z$ alone subject to a constraint on v, e.g., $v_s \ge 1$, where s is the index of a reaction of interest. This is interpretable as the minimum metabolite concentration necessary to carry out a certain function. Such a constraint could be incorporated into (9c).

An objective consisting only of metabolite concentrations might be more physically interpretable. For instance, minimizing metabolite concentrations was used as a model of cellular function in [30]. A second advantage is that the dual sensitivities in (11) also have physical meanings. For these reasons, we maximize $-\mathbf{1}^\top z$ in the example in Section 5.1.

We could also maximize $c^\top v$ alone and add constraints on z. This would be nontrivial, e.g., with coupled polyhedral constraints. However, box constraints like (9d) would simply lead to z binding at its upper bounds.

Note that while the dual is slightly different for these alternative formulations, the expressions for the sensitivities to $V^{\max}$ and K^{m} in (11) are unchanged.

3.3. Dynamic FBA

Dynamic FBA extends conventional FBA in two ways: the incorporation of metabolite concentrations, and changes over time not captured in a steady state model. Metabolite concentrations are present in CFBA, but not transients. We thus formulate dynamic CFBA for the purpose of capturing transient phenomena.

There are two main approaches to dynamic FBA, as described in [5].

- In the dynamic optimization approach, the full trajectory is optimized. The concentrations across time periods are coupled by a finite difference approximation of the derivative; note that we could use a more accurate approximation, e.g., a higher order Runge-Kutta scheme [31]. The optimization is nonlinear due to the Michaelis-Menten kinetics and bilinearities between the fluxes and biomass variables.

- In the static optimization approach, a conventional FBA is solved in each time period. The resulting flux vector is used to propagate the concentrations to the next time period.

The first is harder to solve because it is larger and nonlinear. Here, we optimize the full trajectory as in the dynamic optimization approach. We render the problem more computationally tractable by representing the Michaelis-Menten kinetics in SOC form, (4).

A difficulty not present in steady state FBA is the biomass, z_{b}, which we recall is an element of the vector z. Because we are not setting the derivatives in (1) to zero, we cannot divide it out, and the product of the Michaelis-Menten function and z_{b} does not have an SOC representation. We can remedy this in two ways.

1. If the biomass evolution is predictable, we can approximate it as a fixed parameter in each time period, $\bar{z}_{\mathrm{b}}(t)$. Then the kinetics can be represented in the form of (4).
2. Instead of Michaelis-Menten, we can model the reaction limits with the Contois function [16]:

$$v_i \leq \frac{V_i^{\max} z_{\sigma(i)} z_{\mathrm{b}}}{K_i^{\mathrm{m}} z_{\mathrm{b}} + z_{\sigma(i)}}, \quad i \in \mathcal{Q}. \tag{12}$$

The Contois function is commonly used to model biochemical processes [32], and differs from Michaelis-Menten only in that the denominator depends on z_{b}. Given its similarity to the Michaelis-Menten kinetics, (12) could also be an acceptable approximation. The advantage of the Contois function is that, unlike the product of z_{b} and the Michaelis-Menten function, the resulting kinetics has an SOC representation. As shown in [3], we can represent the inequality (12) as the SOC constraint

$$\left\| \begin{bmatrix} V_i^{\max} z_{\sigma(i)} \\ K_i^{\mathrm{m}} v_i \\ V_i^{\max} K_i^{\mathrm{m}} z_{\mathrm{b}} \end{bmatrix} \right\| \leq V_i^{\max} z_{\sigma(i)} - K_i^{\mathrm{m}} v_i + V_i^{\max} K_i^{\mathrm{m}} z_{\mathrm{b}}. \tag{13}$$

Which approximation is more appropriate depends on the context. For example, if the biomass does not vary much, holding it constant in Michaelis-Menten is a clear choice; one can test this assumption by subsequently simulating (1). If the evolution of the biomass does depend strongly on the the other metabolites, then the Contois function might be a better choice. Alternatively, we could fit a generic SOC constraint to reaction data, which would retain compatibility with SOCP and potentially be more accurate than either of the above approximations [33].

There are time periods $t = 0, ..., \tau$, each of length Δ. Let $v(t)$ and $z(t)$ denote the fluxes and metabolite concentrations in period t. We thus obtain the below SOCP for dynamic CFBA.

$$\begin{aligned} \max_{v(\cdot), z(\cdot)} \quad & \sum_{t=1}^{\tau} c^\top v(t) + d^\top z(t) && \text{(14a)} \\ \text{subject to} \quad & z(0) = z_0 && \text{(14b)} \\ & z(t) - z(t-1) = \Delta S v(t-1) && \text{(14c)} \\ & \underline{z} \leq z(t) \leq \overline{z} && \text{(14d)} \\ & \dot{\underline{v}} \leq v(t) - v(t-1) \leq \overline{\dot{v}} && \text{(14e)} \\ & t = 1, ..., \tau. \end{aligned}$$

If we approximate the biomass as a fixed parameter, we also have for $t = 1, ..., \tau$:

$$\bar{z}_{\mathrm{b}}(t)\underline{v} \leq v(t) \leq \bar{z}_{\mathrm{b}}(t)\overline{v} \tag{14f}$$

$$v_i(t) \leq \frac{V_i^{\max} z_{\sigma(i)}(t) \bar{z}_{\mathrm{b}}(t)}{K_i^{\mathrm{m}} + z_{\sigma(i)}(t)}, \quad i \in \mathcal{Q}. \tag{14g}$$

If we model the reaction limits with the Contois function, (12), then instead of (14f) and (14g), we have for $t = 1, ..., \tau$:

$$z_{\mathrm{b}}(t)\underline{v} \leq v(t) \leq z_{\mathrm{b}}(t)\overline{v} \tag{14h}$$

$$v_i(t) \leq \frac{V_i^{\max} z_{\sigma(i)}(t) z_{\mathrm{b}}(t)}{K_i^{\mathrm{m}} z_{\mathrm{b}}(t) + z_{\sigma(i)}(t)}, \quad i \in \mathcal{Q}. \tag{14i}$$

(14b) is the initial condition. (14c) is the Euler approximation of (1), and couples the variables across time periods. (14d) and (14f) (or (14h)) are bounds on the reactions and metabolites. (14e) limits the rate of change of the reactions. In (14g) (or (14i)), $v_i(t)$ is less than either the Michaelis-Menten kinetics with fixed biomass or the Contois function, depending on which of the above approximations is used.

Note that in dynamic FBA, the evolution of the metabolites generally depends on more than just the stoichiometry, as in (14c). The same should also be the case for dynamic CFBA. For example, there could be inflows and outflows with endogenous or exogenous metabolite concentrations, or biomass death [15].

4. Minimal cut sets

In (11) in Section 3.2, we used dual variables to compute sensitivities to kinetic parameters. This was one of several potential applications of the CFBA dual system. We now explore another in which we use Farkas' Lemma to identify CMCSs.

An MCS is the smallest set of reactions which, if constrained to be zero, disables some function of interest. In this regard, the reactions in an MCS are the lynchpins of the system. In [6], it was shown that the MCSs of a metabolic network correspond to the elementary modes of a dual network, which is specified by Farkas' Lemma [22]. This was an application of a result from [7, 8], which states that there is a one-to-one mapping between the IISs (irreducible infeasible subsystems) of a linear system and the vertices of its dual polyhedron. One can therefore identify the MCSs of a metabolic network by solving for the vertices of the dual polyhedron, which can be done via mixed-integer LP (MILP).

In [9], a weaker version of the result of [7, 8] was extended to semidefinite systems, which generalize SOC systems. We make use of this in Section 4.1 to extend the results of [6] to networks with metabolites linked by Michaelis-Menten kinetics. Whereas an MC contains only reactions, a CMCS contains the smallest set of reactions and/or metabolites that the system cannot function without. We then formulate a linear approximation that produces similar results in Section 4.2.

4.1. Conic minimal cut sets

We now describe the setup, starting with the linear part as given in [6]. Let $\mathcal{T}_v \subseteq \mathcal{N}$ denote the polyhedral set of target reactions, parametrized by the matrix T and vector v^*.The target set encodes some function of interest, which the removal of a cut set disables. The constraint

$$T^\top v \geq v^* \tag{15a}$$

forces the reactions in $\mathcal{T}_v$ to be active. The below two constraints encode the steady state operation of the metabolic network:

$$Sv = \mathbf{0} \tag{15b}$$

$$v_i \geq 0, \quad i \in \mathcal{I}. \tag{15c}$$

We say that $\mathcal{C}_v$ is a cut set for $\mathcal{T}_v$ if $v_i = 0$ for $i \in \mathcal{C}_v$ implies that $v_i = 0$ for $i \in \mathcal{T}_v$ under (15b) and (15c). It is an MCS if it contains no smaller cut sets, i.e., cut sets with fewer elements, for $\mathcal{T}_v$.

To make (15a)-(15c) infeasible, following [34], we add the constraints

$$v_i \leq 0, \quad i \in \mathcal{I} \tag{15d}$$

$$v_i = 0, \quad i \in \mathcal{R}. \tag{15e}$$

Note that together, (15c)-(15e) imply $v = \mathbf{0}$.

Lemma 1 in [6] states that each MCS corresponds to an IIS of (15a)-(15e). Note that an MCS could correspond to multiple IISs, but no two MCSs correspond to the same IIS.

We now incorporate metabolite concentrations and reaction kinetics. We assume there is also a target set of metabolites, $\mathcal{T}_z$, each of which is constrained to be in some concentration range by

$$W^\top z \geq z^*. \tag{15f}$$

The below two constraints prohibit negative concentrations and relate the concentrations to the reactions:

$$z \geq \mathbf{0} \tag{15g}$$

$$v_i \leq \frac{V_i^{\max} z_{\sigma(i)}}{K_i^{\mathrm{m}} + z_{\sigma(i)}}, \quad i \in \mathcal{Q}. \tag{15h}$$

Similar to (15d) and (15e), we make this portion of the system infeasible by adding the constraint

$$z \leq \mathbf{0}, \tag{15i}$$

which, together with (15g), implies $z = \mathbf{0}$.

We note that infeasibility is not always precise enough when moving from linear to nonlinear systems because in the latter case, the set of solutions might not be closed. As in [9], we say that a semidefinite system is weakly feasible if any positive perturbation to its eigenvalues makes it feasible, and weakly infeasible if it is not weakly feasible. Feasibility implies weak feasibility, and weak infeasibility implies infeasibility. (15) is weakly infeasible, and hence also infeasible, because there are positive perturbations that do not make it feasible. We henceforth take the definition of an IIS to refer to weak infeasibility.

We now extend the definition of an MCS.

Definition 1. *$\mathcal{C} = \{\mathcal{C}_v, \mathcal{C}_z\}$ is a cut set for $\{\mathcal{T}_v, \mathcal{T}_z\}$ under (15b), (15c), (15g), and (15h) if the additional constraints $v_i = 0$ for $i \in \mathcal{C}_v$ and $z_i = 0$ for $i \in \mathcal{C}_z$ imply that $v_i = 0$ for $i \in \mathcal{T}_v$ and $z_i = 0$ for $i \in \mathcal{T}_z$. It is a CMCS if it contains no smaller cut sets.*

The following lemma relates Definition 1 to the IISs of the second-order cone system, (15).

Lemma 1. *Each CMCS $\mathcal{C} = \{\mathcal{C}_v, \mathcal{C}_z\}$ for target set $\{\mathcal{T}_v, \mathcal{T}_z\}$ under (15b), (15c), (15g), and (15h) corresponds to an IIS of (15).*

The proof is similar to that of Lemma 1 in [6].

Proof. Consider the CMCS $\mathcal{C} = \{\mathcal{C}_v, \mathcal{C}_z\}$ for target set $\{\mathcal{T}_v, \mathcal{T}_z\}$. The definition specifies an infeasible system consisting of (15a)-(15c), (15f)-(15h), and

$$v_i = 0, \quad i \in \mathcal{C}_v \tag{16a}$$

$$z_i = 0, \quad i \in \mathcal{C}_z. \tag{16b}$$

Denote this system Ψ. Ψ is a subsystem of (15) because (16) is a subsystem of $v = \mathbf{0}$ and $z = \mathbf{0}$. If Ψ is not irreducible, it must contain an IIS. Because the cut set is minimal, the removal of any element from $\mathcal{C}_v$ or $\mathcal{C}_z$ makes Ψ feasible. Therefore, any IIS of Ψ must contain (16). Because $\mathcal{C} = \{\mathcal{C}_v, \mathcal{C}_z\}$ is distinct to the CMCS, any IIS of Ψ is also distinct to the CMCS. □

Note the following two implications of Lemma 1: while each IIS corresponds to at most one CMCS, a CMCS can correspond to multiple IISs; and there may be IISs that do not correspond to a CMCS.

We now seek to relate the IISs of (15) to some dual system, which we recall was specified by Farkas' Lemma in the linear case in [6]. Farkas' Lemma does not apply to (15) due to the SOC constraint, (15h). There are several extensions to semidefinite systems, e.g., [35], which [9] employs to generalize the results of [7, 8] to semidefinite systems. The dual system in this case is also referred to as the alternative spectrahedron. Because any SOC constraint can be written as a semidefinite constraint, the results of [9] specialize to SOC systems like (15) without modification.

The dual system of (15) is

$$\left(\Phi_{3:} - \phi^\top\right)\left(V^{\max} \circ K^{\mathrm{m}}\right) + \rho_v^\top v^* + \rho_z^\top z^* = 1 \tag{17a}$$

$$\begin{bmatrix} \left(\Phi_{2:}^\top + \phi\right) \circ K^{\mathrm{m}} \\ \mathbf{0} \end{bmatrix} + S^\top \lambda + \delta = T^\top \rho_v \tag{17b}$$

$$M^\top \left(\left(\Phi_{1:}^\top - \phi\right) \circ V^{\max}\right) + \gamma = W^\top \rho_z \tag{17c}$$

$$\rho_v \geq \mathbf{0}, \rho_z \geq \mathbf{0} \tag{17d}$$

$$\|\Phi_{:i}\| \leq \phi_i, \quad i \in \mathcal{Q}. \tag{17e}$$

λ is the dual variable of (15b); δ of $v = \mathbf{0}$; γ of $z = \mathbf{0}$; ρ_v of (15a); ρ_z of (15f); and (Φ_i, ϕ_i), $i \in \mathcal{Q}$, of the SOC form of (15h).

Theorem 3.2 in [9] states that if (15) is weakly infeasible, then each of its IISs corresponds to an extremal point of (17). More precisely, each IIS determines the nonzero entries of some extremal point of (17). We can thus identify CMCSs by finding the extremal points of (17).

Unfortunately, (17) may have extremal points that do not correspond to an IIS of (15). Theorem 4.1 in [9] provides conditions under which the alternative spectrahedron has a single solution; however, it does not appear to apply in general to (17).

Lemma 3.1 in [9] states that the indices of a minimal cardinality solution of (17) correspond to an IIS of (15). In our case, similar to [6], we seek solutions in which the vectors δ and γ have minimal cardinality.

In the linear case, there are several different MILPs for solving for IISs. Likewise, one could formulate multiple mixed-integer SOCPs (MISOCP) for finding the IISs of (15); e.g., (9) in [9] can be expressed as an MISOCP when specialized to our problem. The MISOCP below differs from [9] in that we only seek minimal cardinality in δ and γ, and not the other variables. The portion of the problem corresponding to reaction fluxes is based on the MILP (14) in [36].

$$\min_{\lambda,\delta,\gamma,\rho_v,\rho_z,\Phi,\phi,\beta^\delta,\beta^\gamma} \quad \mathbf{1}^\top \beta^\delta + \mathbf{1}^\top \beta^\gamma \tag{18a}$$

$$\text{subject to} \quad (\lambda, \delta, \gamma, \rho_v, \rho_z, \Phi, \phi) \text{ solves (17)} \tag{18b}$$

$$\beta_i^\delta \in \{0,1\}, \beta_i^\gamma \in \{0,1\} \tag{18c}$$

$$|\delta_i| \leq \Upsilon \beta_i^\delta, \quad i \in \mathcal{R} \tag{18d}$$

$$0 \leq \delta_i \leq \Upsilon \beta_i^\delta, \quad i \in \mathcal{I} \tag{18e}$$

$$0 \leq \gamma_i \leq \Upsilon \beta_i^\gamma, \quad i \in \mathcal{M} \tag{18f}$$

$$|\phi_i| + \|\Phi_{:i}\| \leq \Upsilon \beta_{\sigma(i)}^\gamma, \; i \in \mathcal{Q}. \tag{18g}$$

(18c)-(18g) are disjunctive constraints that, for sufficiently large Υ, either force their left hand sides to be zero or have no effect. The left side of (18g) is the 2-norm of the SOC variable $(\Phi_{:i}, \phi_i)$ [25]. We include this constraint because if $\beta_i^\gamma = 0$, the dual variables associated with both (15h) and (15i) should be zero. Given a solution, the IIS is specified by the entries of β^δ and β^γ that are equal to one.

There are a number of further refinements one can make to (18). For example, by weighting the terms in the objective, one can promote cuts with reactions or with metabolites. To exclude either the reactions or metabolites in the target set, or a cut that has already been found, say $\left(\hat{\beta}^\delta, \hat{\beta}^\gamma\right)$, we can add the constraint

$$\hat{\beta}^{\delta\top} \beta^\delta + \hat{\beta}^{\gamma\top} \beta^\gamma \leq \mathbf{1}^\top \beta^\delta + \mathbf{1}^\top \beta^\gamma - 1. \tag{19}$$

4.2. Linear approximation

There are severals reasons why solving (18) might not be the best way to identify CMCSs. First, as discussed, the underlying theoretical results are weaker than in the linear case—an extremal point of the alternative spectrahedron might not correspond to an IIS [9]. Second, MISOCP is less tractable than MILP. And third, given the discrete nature of cut sets, it is not clear that the nonlinearity could not be replaced with something simpler.

The following is a standard linear approximation for when the metabolite concentration is smaller than the Michaelis constant:

$$v_i \leq \eta_i z_{\sigma(i)}, \quad i \in \mathcal{Q}, \tag{20}$$

where $\eta_i = V_i^{\max}/K_i^{\mathrm{m}}$. If $z_{\sigma(i)}$ is in a cut set, i.e., we set it to zero, then $v_i \leq 0$ under either (15h) or (20). Definition 1 and Lemma 1 also hold in both cases. Note that while this definition of η has precedent, any positive value would serve similarly.

Let (15)* denote (15) with (20) instead of (15h), and let $\alpha \in \mathbb{R}^q$ be the dual variable of (20). The dual system

of (15)* is

$$\rho_v^\top v^* + \rho_z^\top z^* = 1 \tag{21a}$$

$$\begin{bmatrix}\alpha \\ \mathbf{0}\end{bmatrix} + S^\top \lambda + \delta = T^\top \rho_v \tag{21b}$$

$$\gamma = M^\top (\alpha \circ \eta) + W^\top \rho_z \tag{21c}$$

$$\rho_v \geq \mathbf{0}, \rho_z \geq \mathbf{0}, \alpha \geq \mathbf{0}. \tag{21d}$$

Because (15)* is polyhedral, the results of [7, 8] apply—each IIS of (15)* corresponds to exactly one extreme point of (21). Lemma 1 establishes that each CMCS corresponds to some IIS of (15)*.

Similar to (18), we can identify the IISs by solving the below MILP.

$$\min_{\lambda,\delta,\gamma,\rho_v,\rho_z,\alpha,\beta^\delta,\beta^\gamma} \quad \mathbf{1}^\top \beta^\delta + \mathbf{1}^\top \beta^\gamma \tag{22a}$$

$$\text{subject to} \quad (\lambda, \delta, \gamma, \rho_v, \rho_z, \alpha) \text{ solves (21)} \tag{22b}$$

$$(18c) - (18f) \tag{22c}$$

$$\alpha_i \leq \Upsilon \beta^\gamma_{\sigma(i)}, \; i \in \mathcal{Q}. \tag{22d}$$

Here (22d) serves the same role as (18g), but is linear because (20) is a linear inequality.

5. *Escherichia coli* example

We now apply the tools we have developed to an example based on the model `e_coli_core` in the BiGG database [37], which corresponds to *Escherichia coli str. K-12 substr. MG1655*. The model has 95 reactions and 72 metabolites.

We take all parameters for the FBA routine (7) from the COBRA Toolbox [12, 13]. We augment the model with Michaelis-Menten kinetics for the reactions listed in Table 1. The values of K^{m} and $V^{\max}$ for each reaction were taken from Table 1 in [38].

All optimizations were carried out in Python using CVXPy [39] and the solver Gurobi [40]. All figures were made with Matplotlib [41].

5.1. CFBA

We first apply CFBA. We solve the alternative formulation described in Section 3.2.1, which is identical to (9) except in two respects.

- The objective is $-\mathbf{1}^\top z$, which corresponds to minimizing the sum of the concentrations of the metabolites in Table 1.
- We add the constraint $v_{\mathrm{b}} \geq \Xi$. v_{b} is the flux of the biomass reaction, which is `BIOMASS_Ecoli_core_w_GAM` in the BiGG database. Here, instead of maximizing the biomass as in FBA, we constrain it to be above $\Xi > 0$.

The SOCP for CFBA took 0.08 seconds to solve. For comparison, the LP for FBA took 0.007 seconds, roughly an order of magnitude less.

Figure 1 shows the sensitivities of the optimal objective to $V^{\max}$ and K^{m} for three of the reaction in Table 1 for $\Xi = 0.1$. Given primal and dual CFBA solutions, the sensitivities are computed via (11). Each represents the change in total metabolite usage resulting from a small change in a Michaelis-Menten parameter, given that the biomass reaction flux cannot go below Ξ. Of the reactions that are not shown, 1 and 2 have sensitivities on the order of 10^{-4}, and the rest 10^{-8}.

We can see that the objective is most sensitive to the kinetics of Reaction 3 (`ACONTa`), which depends on the concentration of citrate. This reaction produces H_20 for a number of other reactions. Via inspection of the network structure and optimal reaction fluxes, we can see that it also directly enables the reactions aconitase (half-reaction B, Isocitrate hydrolyase) and then isocitrate dehydrogenase (NADP).

The objective is also sensitive to K^{m} for Reactions 12 (`GLCpts`) and 17 (`O2t`), which depend on glucose and oxygen, indicating that an increase in K^{m} for either reaction will significantly increase the amount of metabolite needed to keep the biomass reaction flux at Ξ.

This indicates that at the optimal solution, Reactions 3, 12, and 17 are successively further to the left on the steeply increasing part of the Michaelis-Menten function; e.g., a slight increase in oxygen would significantly increases the maximum rate of Reaction 17, where as a slightly increase in citrate would moderately increase the maximum rate of Reaction 3. We can conclude that, for this setup, citrate, glucose, and oxygen are the limiting metabolites for biomass production.

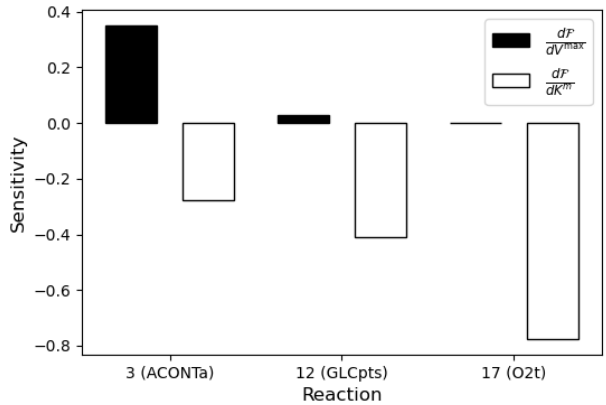

Figure 1: Sensitivities of the optimal CFBA objective, the sum of the metabolite concentrations, to $V^{\max}$ and K^{m} for `ACONTa`, `GLCpts`, and `O2t`. The lower bound on the biomass reaction flux is $\Xi = 0.1$.

Figure 2 shows the concentrations of citrate, glucose, and oxygen as Ξ, the required biomass flux, increases from 0.05 to 0.5. Because these are limiting metabolites, their concentrations increase with Ξ, and do so at a greater than linear rate. The concentrations of citrate and glucose increase rapidly with ξ because, at the optimal solution, they are further to the right on flatter part of the Michaelis-

#	K^{m}	$V^{\max}$	Metabolite	Reaction
1	6	2	Acetate (`ac_c`)	Acetate kinase (`ACKr`)
2	6	2	Acetate (`ac_e`)	Acetate reversible transport via proton symport (`ACt2r`)
3	1	1	Citrate (`cit_c`)	Aconitase (half-reaction A, Citrate hydro-lyase) (`ACONTa`)
4	1	1	Fumarate (`fum_c`)	Fumarase (`FUM`)
5	1	1	Fumarate (`fum_c`)	Fumarate reductase (`FRD7`)
6	1	1	Fumarate (`fum_e`)	Fumarate transport via proton symport (2 H) (`FUMt2_2`)
7	1	2.2	Formate (`for_e`)	Formate transport in via proton symport (`FORt2`)
8	1	2.2	Formate (`for_e`)	Formate transport via diffusion (`FORti`)
9	0.1	1	L-Glutamine (`gln__L_c`)	Glutaminase (`GLUN`)
10	0.1	1	L-Glutamine (`gln__L_e`)	L-glutamine transport via ABC system (`GLNabc`)
11	0.1	1	L-Glutamine (`gln__L_c`)	Glutamate synthase (NADPH) (`GLUSy`)
12	0.53	10	D-Glucose (`glc__D_e`)	D-glucose transport via PEP:Pyr PTS (`GLCpts`)
13	1	1	Succinate (`succ_e`)	Succinate transport via proton symport (2 H) (`SUCCt2_2`)
14	1	1	Succinate (`succ_c`)	Succinate transport out via proton antiport (`SUCCt3`)
15	1	1	Succinate (`succ_c`)	Succinate dehydrogenase (irreversible) (`SUCDi`)
16	1	1	Succinate (`succ_c`)	Succinyl-CoA synthetase (ADP-forming) (`SUCOAS`)
17	0.0024	20	Oxygen (`o2_e`)	O2 transport diffusion (`O2t`)

Table 1: Michaelis-Menten parameters. The units of K^{m} and $V^{\max}$ are mM and mmol/g dw/h, and the values are from [38]. The metabolites and reactions are listed along with their ID in the BiGG database [37].

Menten function. The concentration of oxygen increases least rapidly because, as described above, a slight increase dramatically increases the reaction rate.

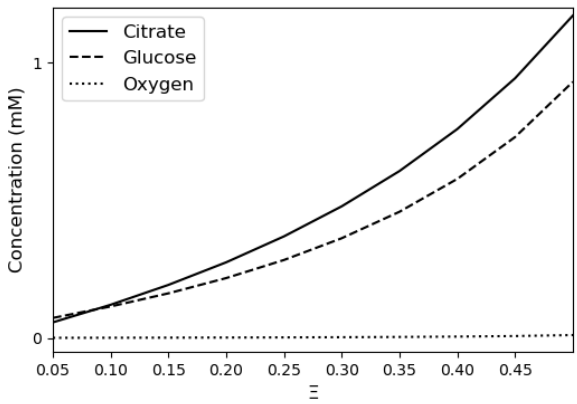

Figure 2: Optimal concentrations of citrate, glucose, and oxygen as a function of the lower bound on the biomass reaction flux, Ξ, in CFBA.

These concentrations are consistent with reported ranges. When $\Xi = 0.5$, the predicted concentration of citrate is 1.17 mM, within the range of 1.1 to 3.5 mM reported in Supplementary Table 3 of [42]. *Escherichia coli*'s metabolism can function over a wide range of oxygen concentrations. When $\Xi = 0.5$, the predicted oxygen concentration is 0.01 mM, which, for example, falls well within the range depicted in Figure 4A of [43].

5.2. Dynamic CFBA

We now test dynamic CFBA by solving (14). Our secondary goal in this section is to understand the scalability of CFBA and its dynamic extension, which we do by varying the time horizon in (14), τ.

The objective, (14a), is to maximize the biomass concentration in the last period, $z_{\mathrm{b}}(\tau)$, as in equation (6b) in Case 2 of [5]. The initial biomass concentration is $z_{\mathrm{b}0} = 0.001$, and the remaining elements of z_0 are ones. The biomass concentration evolves as

$$z_{\mathrm{b}}(t) - z_{\mathrm{b}}(t-1) = \Delta v_{\mathrm{b}}(t-1),$$

where $v_{\mathrm{b}}(t)$, as described in Section 5.1, is the biomass reaction flux.

As described in Section 3.3, we approximate the Michaelis-Menten kinetics with the Contois function, parameterized using the values in Table 1. This means using constraints (14h) and (14i) (and not (14f) or (14g)).

We solved (14) for time horizons ranging from $\tau = 50$ to $\tau = 500$. In each instance the time step was $\Delta = 0.001$ hours. Figure 3 shows the time taken by the solver as a function of τ. When $\tau = 50$, there are 8,473 variables, and when $\tau = 500$, there are 84,073 variables. The trend is increasing with problem size, though some smaller instances take longer than larger ones, presumably due to specific problem structure and solver behavior. The longest time taken by the solver was roughly six minutes. This confirms that, by virtue of being SOCPs, CFBA and dynamic CFBA are highly tractable problems.

The upper plot in Figure 4 shows the biomass, citrate, glucose, and glutamine concentrations through time for the case when $\tau = 500$. Recall that in Section 5.1, citrate and glucose were two of the metabolites CFBA identified as most important. Glutamine enables the glutaminase reaction, v_9 (`GLUN`), and citrate enables aconitase, v_3 (`ACONTa`), both of which are shown in the lower plot.

The biomass concentration increases from near zero, first exponentially, and then more gradually—this is a

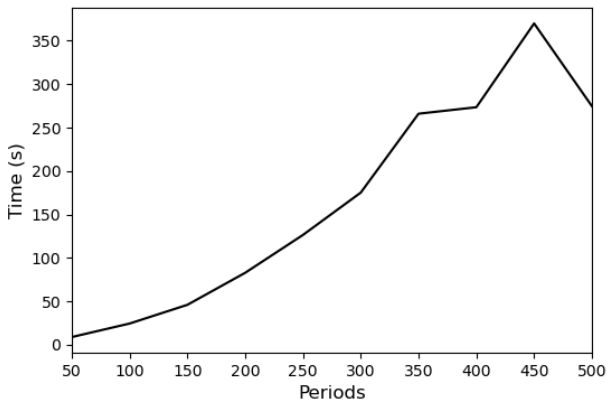

Figure 3: Computation time of dynamic CFBA as a function of the number of periods, τ.

standard behavior. The glutaminase reaction first increases as the biomass concentration increases, and then decreases as the concentration of glutamine drops. The aconitase reaction also increases as biomass concentration increases and then levels out.

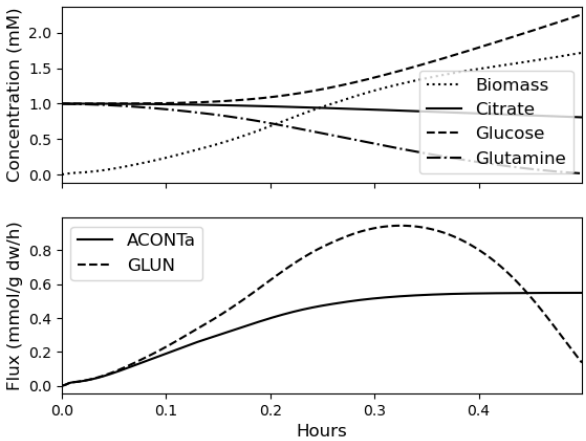

Figure 4: Concentrations (upper) and reactions (lower) through time produced by dynamic CFBA with $\tau = 500$ periods.

These observations point to several potential refinements of dynamic CFBA. First, to induce more complicated behaviors like loss or death of biomass, one must include more complicated exogenous conditions; e.g., in [38] there are five different metabolic phases, some of which have distinct exogenous inputs.

A second potential shortcoming is that dynamic CFBA assumes too much foresight, in that the entire trajectory of 500 periods is optimized at once, and with full knowledge of future exogenous inputs. Full foresight is consistent with the dynamic optimization approach, as described in Section 3.3. However, it is the static optimization approach that finds more usage today, wherein optimization only occurs within individual time periods. A potential remedy is to implement dynamic CFBA with a shorter horizong in receding horizon fashion [44]. This would both limit foresight and enhance scalability.

5.3. CMCS analysis

We compute CMCSs of our example by solving (18), an MISOCP, and (22), an MILP. In addition to the setup at the start of this section, we must also specify target sets, $\mathcal{T}_v$ and $\mathcal{T}_z$, in the form of (15a) and (15f).

Table 2 lists reaction 'knockouts' and the corresponding CMCSs found by (18) and (22), specified by their IDs in the BiGG database [37]. The target set in each case consists of constraining the flux to be greater than one. Note that we only list the first CMCSs found by the solver, that one target set can have many CMCSs, and that a given CMCS can be a solution for both (18) and (22).

Target	CMCS from (18)	CMCS from (22)
GLNS	EX_nh4_e, GLUDy, GLUN	gln__L_c, GLUDy, NH4t
FUM	fum_c	fum_c
SUCDi	succ_c	succ_c
CS	ACONTb	cit_c
MDH	Biomass_Ecoli_core, ACONTa, ADK1, FBA, PYK	ADK1, FBA, PIt2r PYK, cit_c
ME2	Did not converge.	EX_nh4_e, EX_pi_e FORti, MALS, O2t AKGDH, AKGt2r, ALCD2x, ATPS4r, NADTRHD, TALA, succ_e

Table 2: Target knockout reactions and the CMCSs found via (18) and (22).

Whereas an MCS contains only critical reactions, the CMCSs in Table 2 consist of reactions, metabolites, or combinations of both. The prevalence of metabolites in the CMCSs depends heavily on which reactions are limited by Michaelis-Menten kinetics. For instance, fum_c and succ_c are both CMCSs because they limit FUM and SUCDi, among other reactions.

The computation time in all but the final case was under one second. In the last case, (22) took six seconds to solve, and (18) was still not solved after an hour. This highlights the fact that MILP and MISOCP are NP-hard, and similar instances of a problem can take very different times to solve.

Problem (22) often has the same or similar solutions to (18) and, due to the higher tractability and maturity of MILP, is substantially easier to solve in some cases. For these reasons, (22) appears to be more practical than (18).

6. Conclusion

By representing the Michaelis-Menten kinetics as a second-order cone constraint, we can add metabolite concentrations to standard models of metabolic networks without losing much tractability. This has enabled us to formulate several new tools: conic flux balance analysis, dynamic conic flux balance analysis, and conic minimal cut set analysis. In our numerical examples, we demonstrated that each of these new problems is tractable and can provide insight into both reaction fluxes and metabolite concentrations.

There are several directions for future work. We believe that there are numerous potential applications to the many different organisms there are. Such studies could both provide new insights into metabolic networks and further clarify when these tools are appropriate. A starting point for this is applying them to larger metabolic network models. This entails augmenting more existing models with Michaelis-Menten parameter data. There is also room for methodological advancements. For example, a receding horizon implementation [44] could make dynamic conic flux balance analysis both more realistic and tractable. The SOC representation of Michaelis-Menten could be incorporated into flux variability analysis [45] so as to find near optimal ranges of both fluxes and metabolites. There is certainly more to understand about the basic geometry of our setup, which, though convex, is not amenable to many of the techniques used to analyze polyhedral models.

Code availability

The code for the examples is available at: `https://github.com/JAT38/conic-metabolic`.

Acknowledgements

We thank Professor Christopher Lawson for helpful discussion.

Appendix A. Second-order cone programming

We here provide cursory background on SOCP, and refer the reader to [25] and [4] for in-depth coverage. Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and $d \in \mathbb{R}$. A standard form SOC constraint on the variable $x \in \mathbb{R}^n$ is written

$$\|Ax + b\| \leq c^\top x + d,$$

where the left-hand side is the two-norm. If $A = \mathbf{0}$, this reduces to a linear constraint.

A commonly occurring constraint with SOC form is the hyperbolic constraint $x_1^2 \leq x_2 x_3$. It can be written

$$\left\| \begin{bmatrix} 2x_1 \\ x_2 - x_3 \end{bmatrix} \right\| \leq x_2 + x_3.$$

We must also require $x_2 \geq 0$ and $x_3 \geq 0$.

An optimization problem with a linear objective and p SOC constraints is an SOCP, and can in general be written

$$\begin{aligned} \min_x \quad & f^\top x \\ \text{subject to} \quad & \|A_i x + b_i\| \leq c_i^\top x + d_i, \quad i = 1, ..., p, \end{aligned}$$

where $A_i \in \mathbb{R}^{m_i \times n}$, $b_i \in \mathbb{R}^{m_i}$, $c_i \in \mathbb{R}^n$, and $d_i \in \mathbb{R}$. Note that the number of rows in A_i and b_i can be different for each i. SOCP is a generalization of LP because any linear constraint can be written as an SOC constraint.

The SOC is self-dual, i.e., its dual cone is the SOC. The dual of an SOCP is therefore also an SOCP. The dual of the above SOCP is

$$\begin{aligned} \min_{u_i, v_i,\ i=1,...,p} \quad & \sum_{i=1}^p b_i^\top u_i + d_i v_i \\ \text{subject to} \quad & \sum_{i=1}^p A_i^\top u_i + v_i c_i = f \\ & \|u_i\| \leq v_i, \quad i = 1, ..., p. \end{aligned}$$

Strong duality is attained if a constraint qualification holds, e.g., the existence of a Slater point. In this case the two optimizations have the same objective.

References

[1] J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis?, Nature Biotechnology 28 (3) (2010) 245–248.

[2] S. Klamt, E. D. Gilles, Minimal cut sets in biochemical reaction networks, Bioinformatics 20 (2) (2004) 226–234.

[3] J. Taylor, A. Rapaport, Second-order cone optimization of the gradostat, Computers & Chemical Engineering 151 (2021) 107347. `doi:10.1016/j.compchemeng.2021.107347`.

[4] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[5] R. Mahadevan, J. S. Edwards, F. J. Doyle, Dynamic flux balance analysis of diauxic growth in escherichia coli, Biophysical Journal 83 (3) (2002) 1331–1340. `doi:https://doi.org/10.1016/S0006-3495(02)73903-9`.

[6] K. Ballerstein, A. von Kamp, S. Klamt, U.-U. Haus, Minimal cut sets in a metabolic network are elementary modes in a dual network, Bioinformatics 28 (3) (2012) 381–387.

[7] J. Gleeson, J. Ryan, Identifying minimally infeasible subsystems of inequalities, ORSA Journal on Computing 2 (1) (1990) 61–63.

[8] M. Parker, J. Ryan, Finding the minimum weight IIS cover of an infeasible system of linear inequalities, Annals of Mathematics and Artificial Intelligence 17 (1) (1996) 107–126.

[9] K. Kellner, M. E. Pfetsch, T. Theobald, Irreducible infeasible subsystems of semidefinite systems, Journal of Optimization Theory and Applications 181 (3) (2019) 727–742.

[10] M. Watson, Metabolic maps for the Apple II (1984).

[11] A. Varma, B. Ø. Palsson, Metabolic flux balancing: basic concepts, scientific and practical use, Bio/Technology 12 (10) (1994) 994–998.

[12] J. Schellenberger, R. Que, R. M. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, et al., Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0, Nature protocols 6 (9) (2011) 1290–1307.

[13] A. Ebrahim, J. A. Lerman, B. Ø. Palsson, D. R. Hyduke, COBRApy: COnstraints-Based Reconstruction and Analysis for Python, BMC Systems Biology 7 (1) (2013) 1–6.

[14] T. J. Hanly, M. A. Henson, Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures, Biotechnology and Bioengineering 108 (2) (2011) 376–385.

[15] J. A. Gomez, K. Höffner, P. I. Barton, DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis, BMC Bioinformatics 15 (1) (2014) 1–10.

[16] D. Contois, Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures, Microbiology 21 (1) (1959) 40–50.

[17] A. Goelzer, V. Fromion, G. Scorletti, Cell design in bacteria as a convex optimization problem, Automatica 47 (6) (2011) 1210–1218, special Issue on Systems Biology. doi:https://doi.org/10.1016/j.automatica.2011.02.038.

[18] A. Goelzer, V. Fromion, Bacterial growth rate reflects a bottleneck in resource allocation, Biochimica et Biophysica Acta (BBA) - General Subjects 1810 (10) (2011) 978–988, systems Biology of Microorganisms. doi:https://doi.org/10.1016/j.bbagen.2011.05.014.

[19] B. L. Clarke, Stability of Complex Reaction Networks, John Wiley & Sons, 1980, Ch. 1, pp. 1–215.

[20] C. H. Schilling, J. S. Edwards, D. Letscher, B. Ø. P. Palsson, Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective, Journal of Theoretical Biology 203 (3) (2000) 229–248. doi:https://doi.org/10.1006/jtbi.2000.1073.

[21] S. Schuster, C. Hilgetag, On elementary flux modes in biochemical reaction systems at steady state, Journal of Biological Systems 2 (02) (1994) 165–182.

[22] A. Schrijver, Theory of Linear and Integer Programming, John Wiley & Sons, 1998.

[23] L. Michaelis, M. L. Menten, Die kinetik der invertinwirkung, Biochem. Z 49 (333-369) (1913) 352.

[24] C. Chassagnole, N. Noisommit-Rizzi, J. W. Schmid, K. Mauch, M. Reuss, Dynamic modeling of the central carbon metabolism of *Escherichia coli*, Biotechnology and Bioengineering 79 (1) (2002) 53–73.

[25] F. Alizadeh, D. Goldfarb, Second-order cone programming, Mathematical Programming 95 (2003) 3–51.

[26] A. Ben-Tal, A. Nemirovski, On polyhedral approximations of the second-order cone, Mathematics of Operations Research 26 (2) (2001) 193–205.

[27] C. H. Schilling, J. S. Edwards, D. Letscher, B. Ø. Palsson, Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems, Biotechnology and Bioengineering 71 (4) (2000) 286–306.

[28] E. Reznik, P. Mehta, D. Segrè, Flux imbalance analysis and the sensitivity of cellular growth to changes in metabolite pools, PLoS computational biology 9 (8) (2013) e1003195.

[29] R. Heinrich, S. Schuster, The regulation of cellular systems, Springer Science & Business Media, 2012.

[30] S. Schuster, R. Heinrich, Minimization of intermediate concentrations as a suggested optimality principle for biochemical networks: I. Theoretical analysis, Journal of mathematical biology 29 (1991) 425–442.

[31] J. Betts, Survey of numerical methods for trajectory optimization, Journal of Guidance, Control, and Dynamics 21 (2) (1998) 193–207.

[32] G. Bastin, D. Dochain, On-line estimation and adaptive control of bioreactors, Elsevier, 1990.

[33] S. Tan, E. Krichen, A. Rapaport, E. Passeport, J. Taylor, Fitting second-order cone constraints to microbial growth data, Journal of Process Control 118 (2022) 165–169. doi:10.1016/j.jprocont.2022.08.018.

[34] L. Tobalina, J. Pey, F. J. Planes, Direct calculation of minimal cut sets involving a specific reaction knock-out, Bioinformatics 32 (13) (2016) 2001–2007.

[35] I. Klep, M. Schweighofer, An exact duality theory for semidefinite programming based on sums of squares, Mathematics of Operations Research 38 (3) (2013) 569–590.

[36] S. Klamt, R. Mahadevan, A. von Kamp, Speeding up the core algorithm for the dual calculation of minimal cut sets in large metabolic networks, BMC Bioinformatics 21 (1) (2020) 1–21.

[37] Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. Ø. Palsson, N. E. Lewis, BiGG Models: A platform for integrating, standardizing and sharing genome-scale models, Nucleic acids research 44 (D1) (2016) D515–D522.

[38] A. L. Meadows, R. Karnik, H. Lam, S. Forestell, B. Snedecor, Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation, Metabolic Engineering 12 (2) (2010) 150–160, metabolic Flux Analysis for Pharmaceutical Production Special Issue. doi:https://doi.org/10.1016/j.ymben.2009.07.006.

[39] S. Diamond, S. Boyd, CVXPY: A Python-embedded modeling language for convex optimization, The Journal of Machine Learning Research 17 (1) (2016) 2909–2913.

[40] L. Gurobi Optimization, Gurobi optimizer reference manual (2021).
URL http://www.gurobi.com

[41] J. D. Hunter, Matplotlib: A 2D graphics environment, Computing in science & engineering 9 (03) (2007) 90–95.

[42] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, J. D. Rabinowitz, Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*, Nature Chemical Biology 5 (8) (2009) 593–599.

[43] S. G. Henkel, A. t. Beek, S. Steinsiek, S. Stagge, K. Bettenbrock, M. J. T. de Mattos, T. Sauter, O. Sawodny, M. Ederer, Basic regulatory principles of *Escherichia coli*'s electron transport chain for varying oxygen conditions, PLoS ONE 9 (9) (2014) e107640.

[44] J. Mattingley, Y. Wang, S. Boyd, Receding horizon control, IEEE Control Systems Magazine 31 (3) (2011) 52–65.

[45] S. Gudmundsson, I. Thiele, Computationally efficient flux variability analysis, BMC Bioinformatics 11 (1) (2010) 1–3.