Abstract

Flux balance analysis (FBA) is a standard optimization model that is used to study the metabolisms of cells that are in a steady state of optimal growth. The model optimizes a biologically defined objective while assuming (1) equilibria of a linear system of ordinary differential equations, and (2) deterministic data. However, the steady state assumption is imperfect from a biological point of view, and several of the coefficients are experimentally inferred from situations of inherent variation. Here, we propose a robust extension of FBA that removes the firm imposition of reaching steady state, instead capturing the innate variability of a cell culture probabilistically. Our mathematical study of the stochastic problem provides three key insights: 1) metabolic states are (Lipschitz) continuous with regards to the probabilistic modeling parameters, 2) convergent states of the stochastic model are solutions to the deterministic FBA paradigm as the stochastic elements dissipate, and 3) the stochastic model can help identify biological diversity of metabolic networks in an optimized culture. We benchmark our robust counterpart against traditional FBA on two genome-scale metabolic reconstructed models of *E. coli*, and the results show that the stochastic adaptation achieves results comparable to FBA.
1 Introduction and Motivation

Flux balance analysis (FBA) is a computational framework to study metabolic networks, and it has become a primary computational tool with a well established literature [5]. The original study of FBA investigated small portions of the central metabolism [24], but the use of FBA has expanded to encompass genome-scale metabolic models and a diverse range of applications. For example, modern adaptations have suggested combining metabolic models with gene-expression data to study the impacts of factors such as antimicrobial stress and/or temperature stress [14].

FBA is predicated on the fact that we can model a cell’s metabolism as a linear system of equations in the (unknown) fluxes of the metabolic reactions. Such models are widely trusted in the general field of computational chemistry. One premise of FBA is that a metabolic network has reached steady state, which results in a homogeneous system. Another premise is that optimizing an appropriate objective over the fluxes that achieve steady state accurately models how a cellular metabolism reacts as it is held in a constant environment [8, 11]. Different objectives have been proposed and studied [18, 19], with the most common being the maximization of the rate at which biomass is created.

The preponderance of FBA research has studied, updated, augmented, or adapted the basic FBA paradigm to investigate a variety of properties of cellular metabolic networks. However, the basic premises of optimality and steady state have remained intact, although these assumptions are well known to be inexact from a biochemistry perspective. That said, there are two somewhat evident reasons to leave these premises unchecked. First, the resulting optimization problems are most often linear and less often quadratic (convex), and in both cases the problems are easily solved with standard software. Second, traditional FBA models are useful even with questionable premises. For example, FBA models can accurately identify essential genes [10], can predict metabolic responses as pathways are interrupted [6], and can identify the central metabolism [1, 7]. Hence, a justified retort by the FBA community to those who might question the two defining premises is that traditional FBA yields meaningful and computationally tractable science even though the cornerstone premises are imperfect.

Here, we challenge and relax the FBA steady state assumption and instead introduce a robust counterpart to FBA, called robust analysis of metabolic pathways (RAMP). This robust extension allows us to model the problem in terms of probability, and instead of imposing the steady state condition, we alternatively limit the likelihood of deviating from steady state. Such probability statements better agree with biological reality. In the laboratory, cultures are held in a constant environment until they achieve a maximal growth rate: While the culture may be growing as rapidly as possible, sub-populations vary greatly in their (optimal) growth rates [11]. The traditional FBA paradigm aggregates the growth rates of individual cells across the entire culture. For this reason, FBA essentially models the metabolism of an average cell, which may not exist in the culture. By adjusting the paradigm to model the likelihood
of a cell being near steady state, we are in better agreement with the varying biological reality being studied.

Possible concerns with RAMP are that it could lose the successes of FBA and that it could be less computationally tractable. Addressing these concerns motivates our discussion. With regards to computational tractability, RAMP is a robust linear program, which is a type of second-order cone program (SOCPs) that is known to be solvable in polynomial time [3, 4]. Such robust models were developed to overcome problems with over-optimizing designs, meaning that a design could be optimal with respect to the estimated data on which it was built but (much) less so as the data varied over realistic possibilities [2]. A similar concern about over optimization has been expressed in the FBA literature [18], and RAMP addresses this issue by expressing variability within the model itself.

There are several publicly available solvers, e.g. SDPT3 [23] and SeDuMi [21], for SOCPs. However, as we discuss in Section 4, realistic genome-scale metabolic models do not lend themselves to these standard solvers and instead provide unique computational challenges that currently limit investigation. We design our computational experiments to compare RAMP with FBA so that both models can be solved with the default simplex algorithms of several linear solvers. Simplex algorithms have proven themselves to be the only computationally reliable option. Our results show that RAMP rivals FBA in its ability to identify essential genes.

The stochastic flexibility of RAMP permits us to postulate metabolic models for cultures that are away from steady state, and this option opens the possibility of computationally studying functional states of cellular metabolism in the transition to steady state. Such an investigation begs the questions of (1) whether or not RAMP solutions are continuous in their stochastic elements and (2) whether or not convergent trajectories of RAMP solutions yield deterministic FBA solutions as probabilistic assumptions abate. We mathematically answer both questions in the affirmative, and hence, from a RAMP perspective, FBA is justly thought of as the limit of the evolutionary adaptations of a cellular metabolic network that achieves steady state. Our mathematical analysis further identifies the probabilistic situations that coincide with our modeling framework once the limiting deterministic setting is reached. Consequently, if we assume that RAMP’s modeling paradigm aligns with the stochastic nature of a culture, then we can mathematically characterize the ways in which a culture may harbor variation.

Our conclusions are that RAMP deserves further study and that it has the possibility of becoming a computational alternative and complement to traditional FBA. We note that an alternative robust model has been suggested in [25]. While the authors of [25] use robust least squares in a bi-objective model that can be used to illustrate the trade-off between the optimal growth rate and the deviation from steady state, we use robust linear programming. Furthermore, the numerical work in [25] is conducted on a small, illustrative example, and there is no extrapolation of the bi-objective model to a realistic genome-scale metabolic reconstruction.
This article is organized as follows. Sections 2 and 3 introduce the RAMP model and present our mathematical results. Section 4 benchmarks RAMP against FBA. We conclude in Section 5.

2 Robust Analysis for Metabolic Pathways

A metabolism is defined by a system of \( n \) reactions, and we denote the \( j \)-th reaction as

\[
a_{1j}[x_1] + a_{2j}[x_2] + \ldots + a_{mj}[x_m] \xleftrightarrow{k^+_j \frac{k^-_j}{k^-_j}} b_{1j}[x_1] + b_{2j}[x_2] + \ldots + b_{mj}[x_m],
\]

where \([x_i]\) is the concentration of the \( i \)-th metabolite. This gives the following ordinary differential equation for the concentration of the \( i \)-th metabolite,

\[
\frac{d[x_i]}{dt} = \sum_{j=1}^{n} (b_{ij} - a_{ij}) \left( k^+_j [x_1]^{a_{1j}} [x_2]^{a_{2j}} \ldots [x_m]^{a_{mj}} - k^-_j [x_1]^{b_{1j}} [x_2]^{b_{2j}} \ldots [x_m]^{b_{mj}} \right)
\]

\[
= \sum_{j=1}^{n} S_{ij} v_j,
\]

where \( S_{ij} \) is the stoichiometric coefficient for metabolite \( i \) in reaction \( j \) and \( v_j \) is the flux of reaction \( j \). We let \( S \) be the matrix whose coefficients are \( S_{ij} \) and \( v \) be the vector whose components are \( v_j \). We further assume for notational convenience that \( S_i \) is the \( i \)-th row of \( S \) so that \( \frac{d[x_i]}{dt} = S_i v \).

Biological research supports the use of cellular metabolic networks as metabolisms evolve toward optimal flux activity as their environment is left unchanged and as populations are repeatedly sampled and re-grown [10, 11]. FBA assumes the limiting steady state outcome, and hence, fluxes satisfy \( \frac{d[x_i]}{dt} = S_i v = 0 \), which is the equilibrium condition as \( t \to \infty \). FBA further assumes that metabolisms have been tuned through evolutionary processes to best use their resources as assessed by a function of the fluxes, \( g(v) \). Hence, FBA studies metabolic processes through optimization problems, and adaptations thereof, of the form

\[
\max \{ g(v) : S v = 0, L \leq v \leq U \}.
\]

The objective function \( g(v) \) is typically chosen to estimate the rate at which biomass is created, a modeling feat achieved by inferring an effective growth reaction whose output is scaled to a unit of biomass as defined in terms of the cell’s molecular components. Unlike the known metabolic chemical reactions, the coefficients of the effective growth reaction are not integers established by standard stoichiometry and are instead estimated to match experimentally determined cell-culture growth rates. An illustrative example for \( E. coli \) is listed in Table 1. We assume \( g(v) \) is the sole flux of the growth reaction, called the growth rate and denoted by \( v_{\text{Growth}} \), so that \( g(v) = v_{\text{Growth}} = c^T v \).
Table 1: An example of the non-integer coefficients, in parentheses, of the input and output metabolites of a growth reaction.

is the rate at which biomass is created. The vector $c$ of the last notation is all zeros except for a single one at the location of the growth rate. While several alternative objectives have been proposed and investigated in many biological settings [18, 19], the production of biomass is most prevalent and is the default in the COBRA [17] toolbox. The vector of lower bounds, $L$, and upper bounds, $U$, may contain $\pm \infty$, or some suitably large value, to indicate that a flux is unbounded. If $L_j < 0$ and $U_j > 0$, then reaction $j$ is reversible.

## 2.1 A Probabilistic Development of Robust Analysis

The preponderance of stoichiometric coefficients in an FBA model are known with certainty; however, as the growth equation in Table 1 demonstrates, many are estimated and are less certain. In fact, these coefficients are estimated averages from experiments over entire cultures, and linear models, such as (1), ignore the inherit variation among individual cells and assume that a culture’s average properties represent the individuals of the population.

To counter the assumption that an individual mirrors its population’s averages, we instead model individual variability directly and recognize that some coefficients are random, which implies that $d[x_i]/dt = S_i v$ is a random variable. We assume that $S_i v$ is normally distributed with mean $\mu_i$ and variance $\sigma_i^2$, which is a reasonable assumption since the unknown coefficients are estimates from large populations. The modeling
paradigm is altered so that for positive scalars $M_i$ and $\varepsilon$ the deterministic constraint
$S_i v = 0$ is replaced with the following two probabilistic inequalities,
\begin{align*}
P(d[x_i]/dt > M_i) &= P(S_i v > M_i) \leq \varepsilon \quad \text{and} \
P(d[x_i]/dt < -M_i) &= P(S_i v < -M_i) \leq \varepsilon,
\end{align*}
which combine to ensure that $P(-M_i \leq d[x_i]/dt \leq M_i) \geq 1 - 2\varepsilon$. If $M_i$ and $\varepsilon$ are arbitrarily small, then this inequality is a probabilistic interpretation of the deterministic constraint $S_i v = 0$. If $M_i$ increases while $\varepsilon$ is near zero, then we are allowing variation in the steady state assumption but with the near certainty that we have contained the random deviations of $S_i v$ to the interval $[-M_i, M_i]$. If $\varepsilon$ increases, then we are decreasing our certainty that $S_i v$ is within this interval.

Since $(S_i v - \mu_i)/\sigma_i$ is a standard normal variable, constraints (2) and (3) are the same as
\[
\frac{M_i - \mu_i}{\sigma_i} \geq \delta_{1-\varepsilon} \quad \text{and} \quad \frac{-M_i - \mu_i}{\sigma_i} \leq \delta_{\varepsilon},
\]
where $\delta_{\varepsilon}$ and $\delta_{1-\varepsilon}$ are the $\varepsilon$ and $1-\varepsilon$ percentiles, e.g. $P((S_i v - \mu_i)/\sigma_i > \delta_{1-\varepsilon}) = \varepsilon$.

Rearranging these inequalities and using the fact that $\delta_{1-\varepsilon} = -\delta_{\varepsilon}$, we find that
\[
\sigma_i \leq \min \left\{ \frac{M_i - \mu_i}{\delta_{1-\varepsilon}}, \frac{M_i + \mu_i}{\delta_{1-\varepsilon}} \right\}.
\]

To express (4) in terms of the fluxes, we typically mimic the normal distribution with discrete scenarios. Assuming that each $S_i$ has $q$ random scenarios, we let $S_{ik}$ be the stoichiometric coefficients for the $i$-th metabolite in scenario $k$ and let $p_{ik}$ be the probability of scenario $k$, for $k = 1, 2, 3, \ldots, q$. Allowing $p_i$ to be the (positive) vector of probabilities indexed by $k$, $P_i$ to be the (positive definite) diagonal matrix formed by $p_i$, and $\hat{S}_i$ to be the matrix whose $k$-th row is $S_{ik}$, we have that the mean and variance of $S_i v$ are
\[
E(S_i v) = E(S_i)v = \sum_{k=1}^{q} p_{ik} S_{ik} v = p_i^T \hat{S}_i v
\]
and
\[
\text{Var}(S_i v) = \sum_{k=1}^{q} p_{ik} (S_{ik} v - E(S_{ik} v))^2 = (\hat{S}_i v)^T (I - e p_i^T)^T P_i (I - e p_i^T) \hat{S}_i v,
\]
where $e$ is a vector of ones. We let $R_i = \delta_{1-\varepsilon} \sqrt{P_i} (I - e p_i^T) \hat{S}_i v$ so that the variance calculation succinctly satisfies
\[
\text{Var}(S_i v) = \|R_i v\|^2/\delta_{1-\varepsilon}^2.
\]
Substituting the expected value and variance into (4), we have
\[
\|R_i v\| \leq \min \left\{ M_i - p_i^T \hat{S}_i v, \ M_i + p_i^T \hat{S}_i v \right\},
\]
which is the same as
\[ \|R_i v\| - M_i \leq p_i^T \hat{S}_i v \leq M_i - \|R_i v\|. \] (5)

Although (5) is more complicated than a traditional linear constraint, it is the combination of two second order cone constraints. Such constraints share several desirable properties with linear constraints, such as being convex. Replacing the traditional linear constraints \( Sv = 0 \) with (5), we have an SOCP that is our RAMP model:

\[
\begin{align*}
\text{max} & \quad v_{\text{Growth}} \\
\text{subject to} & \quad \|R_i v\| - M_i \leq p_i^T \hat{S}_i v \leq M_i - \|R_i v\|, \quad i = 1, 2, \ldots, m \\
& \quad L \leq v \leq U.
\end{align*}
\] (6)

The RAMP model has an alternative representation upon recognizing
\[ \|R_i v\| = \max \{ u^T R_i v : \|u\| \leq 1 \}, \]
which allows the right-hand side of the SOCP constraint to be re-written as
\[ p_i^T \hat{S}_i v + \|R_i v\| \leq M_i \iff p_i^T \hat{S}_i v + \max \{ u^T R_i v : \|u\| \leq 1 \} \leq M_i \]
\[ \iff \max \{ p_i^T \hat{S}_i v + u^T R_i v : \|u\| \leq 1 \} \leq M_i \]
\[ \iff S_i v \leq M_i, \quad \forall \; S_i \in \{ p_i^T \hat{S}_i + u^T R_i : \|u\| \leq 1 \}. \] (7)

The left-hand side SOCP constraint can be similarly re-written as,
\[ -p_i^T \hat{S}_i v + \|R_i v\| \leq M_i \iff S_i v \leq M_i, \quad \forall \; S_i \in \{ -p_i^T \hat{S}_i + u^T R_i : \|u\| \leq 1 \}. \] (8)

The sets from which \( S_i \) are drawn are called uncertainty sets in the robust optimization literature, and the re-expression of the SOCP constraints in terms of their uncertainty sets supports a geometric description. As an illustrative example, consider the two variable system
\[ S_1 v = s_1 v_1 + s_2 v_2 = 0, \quad -1 \leq v_1 \leq 1, \quad -1 \leq v_2 \leq 1. \]

Assume
\[ p_1 = \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \end{pmatrix} \quad \text{and} \quad \hat{S}_1 = \begin{bmatrix} 0.9 & -1.2 \\ 1 & -1 \\ 1.1 & -0.8 \end{bmatrix}, \]
which means
\[ P(s_1 = 0.9) = P(s_1 = 1.1) = P(s_2 = -1.2) = P(s_2 = -0.8) = 1/4 \]
and 

\[ P(s_1 = 1) = P(s_2 = -1) = 1/2. \]

The expected value of the constraint is

\[ p_1^T \hat{S}_1 v = \tilde{S}_1 v = v_1 - v_2, \]

which is the value of the original, static constraint. Setting \( \delta_{1-\varepsilon} = 3 \) so that \( \varepsilon \approx 0.0015 \), we have

\[
R_1 \approx \begin{bmatrix}
-0.15 & -0.30 \\
0 & 0 \\
0.15 & 0.30 \\
\end{bmatrix}.
\]

(9)

If \( M_1 = 0.2 \), then (7) and (8) combine to show that \( v_1 \) and \( v_2 \) are feasible in the stochastic model if and only if

\[-0.2 \leq (1 - 0.15(u_1 - u_3))v_1 + (-1 - 0.30(u_1 - u_3))v_2 \leq 0.2\]

for all vectors \( u \) such that \( u_1^2 + u_2^2 + u_3^2 \leq 1 \). Hence, \( v_1 \) and \( v_2 \) are feasible in the stochastic case only if they satisfy an infinite number of linear inequalities that are perturbations of a relaxed version of the original, static equality. This infinite collection of constraints is neither more nor less restrictive than the average constraint. One interpretation is that RAMP first relaxes the average equilibrium constraint of \( p_i^T \hat{S}_i v = 0 \) by replacing it with \( -M_i \leq p_i^T \hat{S}_i v \leq M_i \), but then RAMP further restricts the relaxed constraint by replacing it with \( \|R_i v\| - M_i \leq p_i^T \hat{S}_i v \leq M_i - \|R_i v\| \), which adds an infinite number of linear constraints (per original constraint). With regard to our computational work on identifying essential genes in Section 4, the relaxation suggests RAMP might identify fewer essential genes whereas the restriction generally suggests otherwise.

A depiction of the geometry for the above example is shown in Figure 1. The original, static constraint with the variable bounds is depicted by the dashed line segment through the origin. The shaded region is the feasibility set formed by the combined stochastic constraints, and the light dashed lines are samples of the infinite linear inequalities added by the stochastic model. The stochastic feasible region does not contain the static feasible region, for example \((0.75, 0.75)\) is feasible to the original static constraint but is infeasible to the stochastic constraints. Likewise, there are feasible solutions to the stochastic model that are infeasible in the static case.

### 3 Mathematical Observations

We establish three favorable mathematical qualities of the RAMP model. Our first result shows that feasible flux vectors satisfy a Lipschitz continuity property in their probabilistic elements, provided that bounding constraints can be relaxed. In particular, we show that feasible fluxes remain close to perturbed feasible sets as scenarios and probabilities adjust. The following lemma establishes a general property from which our result follows.
Figure 1: A depiction of the difference between an imposed, static equality (the dashed line segment through the origin) and its stochastic counterpart (the shaded region).

Lemma 1. Let \( \bar{A} \) be an \( n \)-element row vector, \( R \) be a \( q \times n \) matrix, and \( b \) be a positive scalar. Let \( \mathcal{F}(\bar{A}, R) = \{ v : \bar{A}v + \|Rv\| \leq b \} \). Then, for any \( v \in \mathcal{F}(\bar{A}, R) \), there are scalars \( c \) and \( \lambda \) satisfying \( 0 < c \leq 1 \) and \( \lambda \geq 0 \) such that

\[
\min \left\{ \|v - v'\| : v' \in \mathcal{F}(\bar{A} + \Delta \bar{A}, R + \Delta R) \right\} \leq \|v - cv\| \leq \lambda (\|\Delta \bar{A}\| + \|\Delta R\|),
\]

where \( cv \in \mathcal{F}(\bar{A} + \Delta \bar{A}, R + \Delta R) \) and \( \lambda \) is independent of \( \|\Delta \bar{A}\| \) and \( \|\Delta R\| \).

Proof. Should \( v \) satisfy \( (\bar{A} + \Delta \bar{A})v + \|(R + \Delta R)v\| \leq b \), then we immediately have

\[
\min \{\|v - v'\| : v' \in \mathcal{F}(\bar{A} + \Delta \bar{A}, R + \Delta R)\} = \|v - v\| = 0,
\]

and we are done with \( c = 1 \) and \( \lambda = 0 \).

Suppose instead that \( (\bar{A} + \Delta \bar{A})v + \|(R + \Delta R)v\| > b \). From the Intermediate Value Theorem there is a \( \tau \) such that \( 0 < \tau \leq 1 \) and

\[
0 < (\bar{A} + (1 - \tau)\Delta \bar{A})v + \|(R + (1 - \tau)\Delta R)v\| = b.
\]

To ease notation, let \( \bar{A}' = \bar{A} + (1 - \tau)\Delta \bar{A} \) and \( R' = R + (1 - \tau)\Delta R \). Also select \( c \) such that

\[
c = \frac{b}{b + \tau (\|\Delta \bar{A}\| + \|\Delta R\|) \|v\|}.
\]
Then,
\[
(\bar{A} + \Delta \bar{A}) (cv) + \| (R + \Delta R) (cv) \| - b \\
= (\bar{A}' + \tau \Delta \bar{A}) (cv) + \| (R' + \tau \Delta R) (cv) \| - b \\
\leq c (\bar{A}'v + \|R'v\| + \tau(\|\Delta \bar{A}\| + \|\Delta R\|)\|v\|) - b \\
= c (b + \tau(\|\Delta \bar{A}\| + \|\Delta R\|)\|v\|) - b \\
= 0.
\]

We conclude that \((\bar{A} + \Delta \bar{A}) (cv) + \| (R + \Delta R) (cv) \| \leq b\), and hence,
\[
cv \in \mathcal{F}(\bar{A} + \Delta \bar{A}, R + \Delta R).
\]

Let \(\lambda = \|v\|^2/b\), from which we have that
\[
\|v - cv\| = \left(1 - \frac{b}{b + \tau(\|\Delta \bar{A}\| + \|\Delta R\|)\|v\|}\right) \|v\|
\]
\[
\leq \left(\frac{\tau\|v\|^2}{b + \tau(\|\Delta \bar{A}\| + \|\Delta R\|)\|v\|}\right) (\|\Delta \bar{A}\| + \|\Delta R\|)
\]
\[
\leq \lambda(\|\Delta \bar{A}\| + \|\Delta R\|).
\]

Hence,
\[
\min \{\|v - v'\| : v' \in \mathcal{F}(\bar{A} + \Delta \bar{A}, R + \Delta R)\} \leq \|v - cv\| \leq \lambda(\|\Delta \bar{A}\| + \|\Delta R\|),
\]
and \(\lambda\) is independent of \(\|\Delta \bar{A}\|\) and \(\|\Delta R\|\).

Lemma 1 shows that vectors satisfying SOCP constraints of the form \(\bar{A}v + \|Rv\| \leq b\), with \(b > 0\), can be scaled to remain feasible, and moreover, that the magnitude of the adjustment to remain feasible is uniformly bounded by the magnitude of the perturbations. Importantly, this result immediately extends to finite collections of SOCP constraints of the same form, a fact we formalize in Corollary 1.

**Corollary 1.** For \(i = 1,\ldots, m\), let \(\bar{A}_i\) be an \(n\)-element row vector, \(R_i\) be a \(q \times n\) matrix, and \(b_i\) be a positive scalar. Let
\[
\mathcal{F} \left( \{(\bar{A}_i, R_i) : i = 1, 2, \ldots, m\} \right) = \left\{ v : \bar{A}_i v + \|R_i v\| \leq b_i, \text{ for } i = 1, 2, \ldots, m \right\}.
\]

Then, for any \(v \in \mathcal{F} \left( \{(\bar{A}_i, R_i) : i = 1, 2, \ldots, m\} \right)\), there are scalars \(c\) and \(\lambda\) satisfying \(0 < c \leq 1\) and \(\lambda \geq 0\) such that
\[
\min \{\|v - v'\| : v' \in \mathcal{F} \left( \{\bar{A}_i + \Delta \bar{A}_i, R_i + \Delta R_i, i = 1, 2, \ldots, m\} \right)\} \\
\leq \|v - cv\| \leq \lambda \max_i \left\{ (\|\Delta \bar{A}_i\| + \|\Delta R_i\|) \right\},
\]

where \(cv \in \mathcal{F} \left( \{\bar{A}_i + \Delta \bar{A}_i, R_i + \Delta R_i : i = 1, 2, \ldots, m\} \right)\) and \(\lambda\) is independent of all \(\|\Delta \bar{A}_i\|\) and \(\|\Delta R_i\|\).
Proof. The proof follows directly from Lemma 1 and its proof upon letting $c_i$ and $\lambda_i$ be the scalars for each $i$ and setting $c = \min_i\{c_i\}$ and $\lambda = \max_i\{\lambda_i\}$. 

Unfortunately, systems of linear equalities like those of FBA do not satisfy similar continuity properties, and changing the stoichiometric matrix in FBA can lead to discontinuities. As a simple example, the system of homogenous equations

\[ \begin{align*}
  v_1 - v_2 &= 0 \\
  v_1 - (1 + \alpha)v_2 &= 0
\end{align*} \]

has substantially different solution sets around $\alpha = 0$. Notice that $v_1 = v_2 = t$ solves the system for any $t$ if $\alpha = 0$, but that the only solution for $\alpha \neq 0$ is $v_1 = v_2 = 0$. Hence the solution $v_1 = v_2 = 1$ at $\alpha = 0$ is not arbitrarily close to a solution with $\alpha \neq 0$.

Corollary 1 suggests that RAMP’s SOCP constraints might provide the noted continuity of the feasible flux vectors with regard to the stochastic modeling elements $p_i$ and $\hat{S}_i$. However, the linear bounds $L \leq v \leq U$ prevent an immediate application of Corollary 1 because real models commonly enforce some fluxes with implied equalities, i.e. $L_i = U_i$ for some $i$. An example is the rate at which ATP is lost due to non-metabolic processes such as cellular repair. Such fixed rates are included to ensure that FBA models take non-growth associated cellular demands into account, which are not explicitly part of the genome-scale metabolic reconstructed model [22]. The continuity results depend on scaling, and Lemma 1 and Corollary 1 do not provide the Lipschitz continuity for RAMP since scaled fixed fluxes do not remain feasible.

The equalities imposed by the bounding constraints could be re-cast probabilistically, but unlike the steady state equations of $Sv = 0$, these constraints cannot be guaranteed to be of the form required by Lemma 1. For instance, if $v_i$ is the flux for the reaction that removes ATP for non-metabolic processes, then the equality $v_i = L_i$ would naturally transition to $P(L_i - M_i \leq v_i \leq L_i + M_i) \geq 1 - 2\varepsilon$, which would associate with an SOCP constraint of the form $L_i - M_i + \|R_i v\| \leq v \leq L_i + M_i - \|R_i v\|$. Since the signs of $L_i - M_i$ and $L_i + M_i$ would agree for small values of $M_i$, which would be required to remain realistic, the requirement of a positive right-hand side in Lemma 1 can not be ensured after writing both inequalities in the correct form. However, if we adjust the variable bounds dependent on the perturbed stochastic elements instead of prescribing them independent of perturbation, then we can relax the bounds to ensure feasibility. Probabilistically this means that we can set the values of $M_i$ for the variable bounds so that we satisfy them with probability 1.

**Theorem 1.** For a collection of probability vectors $p_i$ and scenario matrices $\hat{S}_i$ and for the lower and upper bounds $U$ and $L$, let $\mathcal{F}(\{(p_i, \hat{S}_i) : i = 1, 2, \ldots, m\}, L, U)$ be the nonempty set of feasible fluxes satisfying the constraints of RAMP model (6). Assuming that each $p_i + \Delta p_i$ is a probability vector, we then have for each

\[ v \in \mathcal{F}(\{(p_i, \hat{S}_i) : i = 1, 2, \ldots, m\}, L, U) \]
that there is a $\lambda \geq 0$ such that

$$
\min_{v'} \|v - v'\| \leq \lambda \Gamma
$$

where

$$
\Gamma = \max_i \left\{ \|\Delta p_i\| + \|\Delta \hat{S}_i\| + \|\Delta p_i\| \|\Delta \hat{S}_i\| \right\},
$$

$v' \in \mathcal{F}(\{(p_i + \Delta p_i, \hat{S}_i + \Delta \hat{S}_i) : i = 1, 2, \ldots, m\}, L - \lambda \Gamma e, U + \lambda \Gamma e)$, and

$\lambda$ is independent of all $\|\Delta \hat{S}_i\|$.

**Proof.** Let $\bar{A}_i = p_i^T \hat{S}_i$ and $\bar{A}_i + \Delta \bar{A}_i = (p_i + \Delta p_i)^T (\hat{S}_i + \Delta \hat{S}_i)$. Then,

$$
\|\Delta \bar{A}_i\| = \|(p_i^T + \Delta p_i^T)(\hat{S}_i + \Delta \hat{S}_i) - p_i^T \hat{S}_i\| \\
\leq \|p_i^T \Delta \hat{S}_i + \Delta p_i^T \hat{S}_i + \Delta p_i^T \Delta \hat{S}_i\| \|v\| \\
\leq \kappa_i^A \left( \|\Delta \hat{S}_i\| + \|\Delta p_i\| + \|\Delta \hat{S}_i\| \|\Delta p_i\| \right), \quad (10)
$$

where $\kappa_i^A = \|v\| \max\{\|\hat{S}_i\|, \|p_i\|, 1\}$. Further note that

$$
\|\Delta R_i\| = \|(R_i + \Delta R_i) - R_i\| \\
= \delta_{1-\varepsilon} \left\| \sqrt{P_i + \Delta P_i}(I - e(p_i + \Delta p_i)^T)(\hat{S}_i + \Delta \hat{S}_i) - \sqrt{P_i}(I - ep_i^T)\hat{S}_i \right\| \\
= \delta_{1-\varepsilon} \left\| \sqrt{P_i + \Delta P_i} - \sqrt{P_i} \right\| \|I - ep_i^T\| \|\hat{S}_i\| \\
+ \sqrt{q} \|\Delta p_i\| \|\hat{S}_i\| + \sqrt{q} \|\Delta p_i\| \|\Delta \hat{S}_i\| \\
\leq \delta_{1-\varepsilon} \left( \|\sqrt{P_i + \Delta P_i} - \sqrt{P_i}\| \|I - ep_i^T\| \|\hat{S}_i\| + \|I - ep_i^T\| \|\Delta \hat{S}_i\| \\
+ \sqrt{q} \|\Delta p_i\| \|\hat{S}_i\| + \sqrt{q} \|\Delta p_i\| \|\Delta \hat{S}_i\| \right). \quad (11)
$$

From the Mean Value Theorem there is a $\mu_{ik}$ between $p_{ik}$ and $p_{ik} + \Delta p_{ik}$ such that

$$
\sqrt{p_{ik} + \Delta p_{ik}} - \sqrt{p_{ik}} = \frac{\Delta p_{ik}}{2\sqrt{\mu_{ik}}},
$$

Hence,

$$
\left\| \sqrt{P_i + \Delta P_i} - \sqrt{P_i} \right\| = \sqrt{\sum_{k=1}^{q} \left( \sqrt{p_{ik} + \Delta p_{ik}} - \sqrt{p_{ik}} \right)^2} \\
= \sqrt{\sum_{k=1}^{q} \left( \frac{\Delta p_{ik}}{2\sqrt{\mu_{ik}}} \right)^2} \\
\leq \frac{1}{2\sqrt{\mu_{i}}} \|\Delta p_i\|,
$$

12
where \( \mu_i = \min_k \{\mu_{ik}\} \). We now have from inequality (11) that
\[
\| \Delta R_i \| \leq \kappa_i^R \left( \| \Delta \hat{S}_i \| + \| \Delta p_i \| + \| \Delta \hat{S}_i \| \| \Delta p_i \| \right),
\] (12)

where
\[
\kappa_i^R = \delta_{1-\varepsilon} \max_i \left\{ \frac{1}{2\sqrt{\mu_i}} \left( \| I - c p_i^T \| + \sqrt{q} \right) \| \hat{S}_i \|, \| I - c p_i^T \|, \sqrt{q} \right\}.
\]

Each of RAMP’s SOCP constraints may be re-written as
\[
(\bar{A}_i + \Delta \bar{A}_i)v + \|(R_i + \Delta R_i)v\| \leq M_i, \; i = 1, 2, \ldots, m \quad \text{and} \quad -(\bar{A}_i + \Delta \bar{A}_i)v + \|(R_i + \Delta R_i)v\| \leq M_i, \; i = 1, 2, \ldots, m.
\]

From Corollary 1 and inequalities (10) and (12) we know that there is a \( v' \) satisfying these perturbed SOCP constraints and a \( \hat{\lambda} \geq 0 \) defined independent of \( \| \Delta A_i \| \) and \( \| \Delta R_i \| \), and subsequently independent of \( \| \Delta \hat{S}_i \| \), such that
\[
\| v - v' \| \leq \hat{\lambda} \max_i \left\{ \| \Delta \bar{A}_i \| + \| \Delta R_i \| \right\}
\]
\[
\leq \hat{\lambda} \max_i \left\{ (\kappa_i^A + \kappa_i^R) \left( \| \Delta \hat{S}_i \| + \| \Delta p_i \| + \| \Delta \hat{S}_i \| \| \Delta p_i \| \right) \right\}
\]
\[
\leq \hat{\lambda} \max_i \left\{ (\kappa_i^A + \kappa_i^R) \right\} \max_i \left\{ (\| \Delta \hat{S}_i \| + \| \Delta p_i \| + \| \Delta \hat{S}_i \| \| \Delta p_i \| ) \right\}.
\]

Since \( \hat{\lambda}, \kappa_i^A, \text{ and } \kappa_i^R \) are all independent of \( \| \Delta \hat{S}_i \| \), the proof is complete upon noticing that \( v' \) must also satisfy
\[
L - \lambda \Gamma_{\varepsilon} \leq v' \leq U + \lambda \Gamma_{\varepsilon},
\]
where \( \lambda = \hat{\lambda} \max_i \{ \kappa_i^A + \kappa_i^R \} \).

Theorem 1 highlights that any questionable discontinuities that FBA might exhibit with regard to parametric update are due to the imposed linear equalities that are the byproduct of enforcing the biologically unrealistic assumption of a uniform steady state. RAMP adjusts FBA’s modeling paradigm to include its inherit stochastic nature, and in the process RAMP ensures the continuity that would be expected of the feasible flux states as long as the imposed linear bounds can also be relaxed commensurate with the magnitude of the perturbation. The conclusion is that any discontinuity of FBA that could be caused by minute model adjustments are the outcome of an overly rigid (linear) model.

A reasonable question is if the deterministic FBA model is a limiting case of RAMP as the stochastic elements diminish, and Theorems 2 and 3 show that convergent RAMP solutions are indeed FBA solutions and that interpreting FBA as a limiting RAMP model characterizes the possible random flux states among cells in an optimal growth, steady state culture. The convergence result of Theorem 2 assumes an interiority condition, which is tacit as long as the equalities imposed by the bounding constraints are relaxed to allow arbitrarily small adjustments, i.e. as long as \( L_i = v_i = U_i \) is
replaced with \( L_i - \eta \leq v_i \leq U_i + \eta \) for any arbitrarily small \( \eta > 0 \). The proof of Theorem 2 is straightforward and follows from the fact that if the primal and dual solutions converge as the stochastic elements dissipate, then the resulting necessary and sufficient conditions are that of FBA. However, the proof importantly identifies that not all random elements need to disappear, a point that prompts Theorem 3.

**Theorem 2.** Let \( p_i^t, \hat{S}_i^t, M_i^t \), be sequences such that for each \( t \) the corresponding RAMP model satisfies Slater’s interiority condition. Let \( v^t \) be an optimal solution of the RAMP model corresponding to \( p_i^t, \hat{S}_i^t, \) and \( M_i^t \), and assume \( v^t \to v \) as \( t \to \infty \). Assume likewise that a corresponding dual sequence of optimal solutions converges. Further assume that as \( t \to \infty \) we have \( (p_i^t)^T \hat{S}_i^t \to \bar{S}_i, R_i^t \to 0, \) and \( M_i^t \to 0 \). Then \( v \) is a solution to the FBA model

\[
\max \{ v_{\text{Growth}} : \bar{S}v = 0, \ L \leq v \leq M \},
\]

where \( \bar{S} \) is the matrix whose \( i \)-th row is \( \bar{S}_i \).

**Proof.** For each \( t \) let \((\hat{y}^t, y^t, \hat{w}^t, w^t, \rho^t, \sigma^t)\) be the assumed dual solution that converges to \((\hat{y}, y, \hat{w}, w, \rho, \sigma)\). Should the dual also satisfy Slater’s interiority condition, then the necessary and sufficient primal-dual conditions would be satisfied by the convergent sequences:

\[
-M_i^t + \| R_i^t v^t \| \leq (p_i^t)^T \hat{S}_i^t v^t \leq M_i^t - \| R_i^t v^t \|, \ \forall \ i
\]

\[
L \leq v^t \leq U
\]

\[
\sum_i \left((\hat{S}_i^t)^T p_i^t (\hat{y}_i^t - \hat{w}_i^t) - (R_i^t)^T (y_i + w_i)\right) + \rho^t - \sigma^t = c
\]

\[
\| y_i^t \| \leq \hat{y}_i^t, \ \forall \ i
\]

\[
\| w_i^t \| \leq \hat{w}_i^t, \ \forall \ i
\]

\[
\rho^t, \sigma^t \geq 0
\]

\[
\sum_i M_i^t (\hat{y}_i^t + \hat{w}_i^t) + U^T \rho^t - L^T \sigma^t - c^T v^t = 0.
\]

Dual feasibility is defined by the 3rd through 6th constraints, from which we can see that the dual is always strictly feasible. Select any \( y \) and \( w \) variables such that \( \| y_i^t \| < \hat{y}_i^t \) and \( \| w_i^t \| < \hat{w}_i^t \). Since \( \rho^t - \sigma^t \) can attain any vector, these two variables can be chosen so that \( \rho^t > 0, \sigma^t > 0 \) and that

\[
\sum_i \left((\hat{S}_i^t)^T p_i^t (\hat{y}_i^t - \hat{w}_i^t) - (R_i^t)^T (y_i^t + w_i^t)\right) + \rho^t - \sigma^t = c.
\]

Hence the system is indeed necessary and sufficient for optimality.

Allowing \( t \to \infty \), we have by assumption that \((p_i^t)^T \hat{S}_i^t \to \bar{S}_i, R_i^t \to 0, \) and \( M_i^t \to 0 \), from which we have

\[
\bar{S}v = 0
\]

\[
L \leq v \leq U
\]

\[
\hat{y}_i, \hat{w}_i, \rho, \sigma \geq 0
\]

\[
U^T \rho - L^T \sigma - c^T v = 0.
\]
Since these are the necessary and sufficient conditions for FBA, \( v \) is an optimal solution to the FBA model. \( \square \)

The assumption that \( R^i_t \to 0 \) in Theorem 2 could be relaxed, and the proof remains valid as long as the limiting matrices \( R_i \) satisfy \( R_i v = 0 \) and \( R^T_i (y_i + w_i) = 0 \). Since \( R_i \) can have low rank, e.g. the rank of \( R_1 \) in (9) is 1, we see that \( R_i \) need not generally vanish. This observation suggests that not all uncertainty needs to be removed to recover an FBA solution with RAMP, and it is this observation that motivates our third result.

Suppose \( \hat{v} \) is an optimal FBA solution with \( S_i \) being the average stoichiometric vector \( p^T \hat{S}_i \). FBA models are well known to be highly degenerate [12], which points to substantial variability among optimal, steady state metabolic fluxes. Might it be possible to identify scenarios and probabilities so that \( \hat{v} \) is an optimal RAMP solution as the likelihood of steady state increases as \( M_i \downarrow 0 \)? If so, then while \( \hat{v} \) is a solution to a deterministic FBA model, it is also the solution to a stochastic counterpart that models the steady state probabilistically. If such probabilities and scenarios exist, then the optimal flux state \( \hat{v} \) would identify stochastic variation within an optimal growth, steady state culture. We say that scenarios \( \hat{S}_i \) are \textit{biologically possible} for \( \hat{v} \) if there exists probability vectors \( p_i \) such that \( \hat{v} \) maximizes cellular growth under the conditions that \( \hat{v} \) satisfies \( L \leq \hat{v} \leq U \) and

\[
\lim_{M_i \to 0} P\left(-M_i \leq p^T_i \hat{S}_i \hat{v} \leq M_i \right) \leq 1 - 2\varepsilon, \forall i.
\]

Theorem 3 characterizes the biologically possible scenarios of any FBA solution. The argument arranges and partitions \( \hat{v} \) into \((\hat{v}', \hat{v}'')\), where \( \hat{v}' \neq 0 \) and \( \hat{v}'' = 0 \). This way \( S \hat{v} = S' \hat{v}' \), where \( S' \) is the submatrix of \( S \) whose columns correspond with \( \hat{v}' \).

\textbf{Theorem 3.} Let \( \hat{v} = (\hat{v}', 0) \), with \( \hat{v}' \neq 0 \), be a solution to the FBA problem

\[
\max \{v_{\text{Growth}} : p^T_i \hat{S}_i v = 0, \forall i, \ L \leq v \leq U\}.
\]

Then the scenarios of \( \hat{S}_i \) are biologically possible for \( \hat{v} \) if and only if for all \( i \) we have \( \hat{S}_i' \hat{v}' = \alpha_i e \) for some scalar \( \alpha_i \neq 0 \).

\textit{Proof.} Let \( \hat{v} = (\hat{v}', 0) \) be as stated above. Then,

\[
R_i \hat{v} = R_i' \hat{v}' = \sqrt{P_i}(I - ep_i^T) \hat{S}_i \hat{v}' = 0 \quad \text{if and only if} \quad ep_i^T \left( \hat{S}_i' \hat{v}' \right) = \left( \hat{S}_i' \hat{v}' \right).
\]

The last equality shows that \( R_i \hat{v} = 0 \) if and only if \( \hat{S}_i \hat{v}' \) is an eigenvector of \( ep_i^T \) for the eigenvalue 1. Since \( p_i \) is a probability vector, \( ep_i^T \) has only two eigenspaces, one of dimension \( q - 1 \) for the eigenvalue 0, and one of dimension 1 for the eigenvalue 1. All eigenvectors for the eigenvalue of 1 are scalar multiples of the all ones vector \( e \). Hence, \( R_i \hat{v} = 0 \) if and only if the scenarios for each \( i \) satisfy \( \hat{S}_i' \hat{v}' = \alpha_i e \) for some \( \alpha_i \neq 0 \).
RAMP's SOCP constraints as $M_i \downarrow 0$ are

$$\|R_i v\| \leq p_i^T \hat{S}_i v \leq -\|R_i v\| \iff \begin{cases} p_i^T \hat{S}_i v = 0 \\ R_i v = 0. \end{cases}$$

Hence, the limiting RAMP model as $M_i \downarrow 0$ is the linear program

$$\max \{ v \text{Growth} : p_i^T \hat{S}_i v = 0, \ R_i v = 0, \ \forall \ i, \ L \leq v \leq U \}. \quad (15)$$

The necessary and sufficient conditions are

$$p_i \hat{S}_i v = 0, \ \forall \ i$$

$$R_i v = 0, \ \forall \ i$$

$$L \leq v \leq U$$

$$\sum_i \left( S_i^T p_i y_i + R_i^T w_i \right) - \rho + \sigma = c$$

$$\rho, \sigma \geq 0$$

$$U^T \sigma - L^T \rho - c^T v = 0$$

Suppose for each $i$ that $S'_i v' = \alpha_i e$ for some $\alpha_i \neq 0$. Then $\hat{v}$ satisfies $R_i v = 0$ for all $i$. Moreover, since $\hat{v}$ solves (13), the strong duality theorem of linear programming guarantees a solution to (16) with $v = \hat{v}$ and $w_i = 0$. Alternatively, if for some $i$ we have $S'_i v' \neq \alpha_i e$ for all nonzero $\alpha_i$, then $R_i \hat{v} \neq 0$ and $\hat{v}$ is infeasible in (15). Hence, in this case $\hat{S}_i$ is not biologically possible for $\hat{v}$.

In biological terms, Theorem 3 identifies the probabilistic variations that are possible in the stoichiometric matrix for cells with the same optimal, steady state fluxes. So, if we trust that FBA solutions do indeed identify the limiting behavior of cells as they optimize to their environmental resources, then Theorem 3 concludes that optimized cells with the same flux state can be described by different probabilistic models. Moreover, since (14) only requires that $p_i$ be a probability vector with nonzero entries, we have a test to see if any particular collection of scenarios is possible. For example, suppose we want to query if some of the stoichiometric coefficients can vary for the cells having a flux state $\hat{v}$ in an optimized culture. Any collection of scenarios for which $\hat{S}_i \hat{v}$ does not have identical components is impossible since there is no way to assign probabilities to make the flux state optimal once probabilistic variation is included in the model. If our mathematical models accurately assess biological reality, then we can claim that it is biologically impossible to have a collection of optimized cells with a common flux state that vary according to the suggested scenarios.

4 RAMP as a Computational Model

RAMP’s computational ability is compared with FBA below. The computational results herein only initiate the broad comparisons that would benchmark RAMP against
the many ways in which FBA has been used. Since one of the typical tests to assess a metabolic model’s potential for producing biologically relevant predictions is it’s ability to identify essential and viable genes, this is the metric we use in our comparisons. We discuss some options for future work in Section 5. All tests were conducted on the iJO1366 and iAF1260 metabolic models of *Escherichia coli* [9, 16]. These models have, respectively, 1366 and 1260 genes, 2583 and 2382 metabolic reactions, and 1805 and 1668 metabolites and are considered to be very high quality and realistic reconstructions of the *E. coli* genome-scale metabolic network.

Gene essentiality is decided by simulating gene knockouts. If a gene’s activity is blocked, then a set of fluxes is inhibited. The simulation removes the fluxes associated with a knockout from the metabolic model to mimic the gene knockout. If the reduced metabolic model affects the optimal growth rate, typically by at least a 50% reduction, then the gene is considered to be essential to the organism. If a predictive model correctly identifies a gene as essential, we label this as a true positive. Similarly, a true negative indicates that a predictive model correctly identifies a non-essential gene. A false positive occurs if the predictive model incorrectly labels a gene as essential, and a false negative occurs if the predictive model incorrectly labels a gene as non-essential. The predictive power of the model is the ratio of the sum of true positives and negatives to the number of genes.

### 4.1 Computational Stability

The computational burden to conduct the identification of essential genes has been substantial, and the authors have had a multi-year effort to achieve stable numerical results. Describing our effort will hopefully aid others who might want to pursue robust extensions of FBA. We initially attempted several of the standard re-formulations of SOCPs as nonlinear programs. Since these models proved intractable, we drew the reasonable conclusion that a native SOCP solver might be required. The COBRA toolbox [17] is written in Matlab, and we attempted to solve RAMP models with the well established Matlab solvers SDPT3 and SeDuMi, only to find that computational concerns persisted. After testing a battery of different metabolic models, alternate formulations, and parameter settings, we found no combination that reliably solved RAMP models as SOCPs.

Our computational frustration identified a likely culprit in the underlying interior-point methods themselves. Indeed, interior point methods regularly failed on linear FBA models. We believe it is the case that metabolic models contain two particular challenges: First, a substantial number of variables are unsigned and, essentially, unbounded. The default models use $\pm 10^5$ as surrogates for $\pm \infty$, and we were unsuccessful in remodeling or tightening these bounds to reach stability. Second, and already noted, is the fact that metabolic problems are highly degenerate and often have high-dimensional solution sets. Whether due to these problems or other reasons, we were unable to find trustworthy success with interior-point algorithms for LP, SOCP, or
NLP.

Our computational success relies on the fact that if the probabilistic variability of each constraint is restricted to a single coefficient, then the associated SOCP remains linear. The linear constraint is constructed by, for example, assuming that the last coefficient of the \(i\)-th constraint is the sole random variable. Then for some \(q\)-vector \(\hat{s}\) we have

\[
\|R_i v\| = \|\delta_{1-\varepsilon} \sqrt{P_i} (I - e p_i^T) \hat{S}_{i} v\|
\]

\[
= \|\delta_{1-\varepsilon} \sqrt{P_i} (I - e p_i^T) [S_{i,1} e, \ldots, S_{i,(m-1)} e, \hat{s}] v\|
\]

\[
= \|\delta_{1-\varepsilon} \sqrt{P_i} [S_{i,1}(e - e p_i^T e), \ldots, S_{i,(m-1)}(e - e p_i^T e), (I - e p_i^T) \hat{s}] v\|
\]

\[
= \|\delta_{1-\varepsilon} \sqrt{P_i} (I - e p_i^T) \hat{s}\| \nu_n.
\]

This calculation shows that if we restrict probabilistic variability to, for example, the coefficients of the growth equation, then each of the SOCP constraints of the form

\[
\|R_i v\| - M_i \leq p^T \hat{S}_i v \leq M_i - \|R_i v\|,
\]

can be re-written linearly as

\[
P^T \hat{S}_i v - \|\delta_{1-\varepsilon} \sqrt{P_i} (I - e p_i^T) \hat{S}_{i,Growth}\| \nu_{Growth} \geq M_i
\]

\[
P^T \hat{S}_i v + \|\delta_{1-\varepsilon} \sqrt{P_i} (I - e p_i^T) \hat{S}_{i,Growth}\| \nu_{Growth} \leq M_i,
\]

(17)

where \(\hat{S}_{i,Growth}\) is the column of \(\hat{S}_i\) containing the scenarios for the growth coefficient. We choose to use the growth coefficients for two reasons. First, we have already noted that this is an empirically derived reaction that is based on the aggregate properties of a cell culture, and thus, the stoichiometric values are inherently associated with uncertainty. Furthermore, it is well known that biomass composition (and thus the values of the stoichiometric coefficients of the growth reaction) of a cellular culture is affected by nutrient conditions and the culture’s growth phase [13, 15].

The linear re-formulation in (17) allows RAMP to be solved with standard linear simplex solvers, which proved to be computationally stable. All numerical work was conducted with the freeware GLPK, which worked well with the COBRA toolbox [17].

### 4.2 Probabilistic Models

We considered four probabilistic models to assess RAMP’s sensitivity to probabilistic variation. Each of these tests are benchmarked against FBA’s predictions as calculated by the COBRA toolbox.
**Model 1 (default)** Our default RAMP model assumes that the means of the growth coefficients are the values stated in the FBA model and that probabilistic variation is restricted to the first unspecified significant digit. Assuming $10^{-d_i}$ identifies this digit for the $i$-th growth coefficient, we further assume that each growth coefficient has the 5 scenarios in which it is perturbed by $\pm \eta \cdot 10^{-d_i}$, where $\eta$ is one of 0, 1, or 2. The probability of the scenario with no perturbation is $P(-1/2 \leq z \leq 1/2) = 0.4520$, with perturbation $\pm 10^{-d_i}$ is $P(1/2 \leq z \leq 3/2) = 0.2389$, and with perturbation $\pm 2 \cdot 10^{-d_i}$ is $P(3/2 \leq z) = 0.0351$, where $z$ is a standard normal variable. A couple of examples are illustrated in Table 2. This model sets $\delta_{1-\varepsilon} = 3$ so that the probabilistic constraints have a 99.97% guarantee of satisfaction.

**Model 2** The second probabilistic model multiplies $\eta$ in the default model by the scalar $\rho$ to see how RAMP adjusts as the scenarios deviate away from those of FBA. Eight tests with $\rho = 2, 3, \ldots, 9$ were considered. Integers beyond 9 resulted in sign changes and were not considered. The probabilities from the default model were used, and $\delta_{1-\varepsilon}$ was similarly set to 3.

**Model 3** Model 2’s scaling by $\rho$ disproportionately effects small growth coefficients, and we counter this by replacing the default scenarios with percentages of the growth coefficient itself. We let $\sigma$ range over

{0.001, 0.002, \ldots, 0.009, 0.01, 0.02, \ldots, 0.09, \ldots 0.1, 0.2, \ldots, 0.9},

and for any of these values of $\sigma$ the scenarios are $(1 \pm \sigma \cdot \eta) \cdot \hat{S}_{i,Growth}$, where $\eta$ is one of 0, $\pm 1/2$, or $\pm 1$. The probabilities are unchanged from the default case, and $\delta_{1-\varepsilon}$ is 3.

**Model 4** To assess how RAMP reacts to changes in the certainty of satisfying the probabilistic constraints, model 4 changes $\delta_{1-\varepsilon}$ as $2\varepsilon$ ranges over

{0.01, 0.02, \ldots, 0.09, 0.1, 0.2 \ldots, 0.9}.

All other model parameters are inherited from the default model. We note that since $\varepsilon$ is used as an upper and lower bound on the probability constraint, this set collectively samples the full set of options with regard to changes in $\varepsilon$.

The selection of $M_i$ is yet another stochastic parameter to decide. However, instead of imposing these bounds arbitrarily, these parameters are calculated so that each RAMP model accurately returns the targeted, optimal growth rate. Let $\gamma^*$ be the optimal growth rate as calculated by the FBA model. The $M_i$ values are calculated by...
<table>
<thead>
<tr>
<th>Metabolite</th>
<th>FBA Growth Coefficient</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
<th>Scenario 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ohph[c]</td>
<td>-0.00023</td>
<td>-0.0002232</td>
<td>-0.0002231</td>
<td>-0.000223</td>
<td>-0.0002229</td>
<td>-0.0002228</td>
</tr>
<tr>
<td>adp[c]</td>
<td>53.95</td>
<td>53.948</td>
<td>53.499</td>
<td>53.95</td>
<td>53.951</td>
<td>53.952</td>
</tr>
<tr>
<td>probability</td>
<td>0.0351</td>
<td>0.2389</td>
<td>0.4520</td>
<td>0.2389</td>
<td>0.0351</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: The first column is the metabolite name from the iJO1366 model, and the second column contains the associated growth coefficient from the FBA model. The sign indicates whether the metabolite is an input (negative) or output of the growth equation. The remaining columns are the scenarios for the default case along with their probabilities.

\[
\begin{align*}
\text{min} & \quad \|M\|_1 \\
\text{subject to} & \\
\tau_{\text{Growth}} & \geq \gamma^* \\
\|R_i v\| - M_i & \leq p^T \hat{S}_i v \leq M_i - \|R_i v\|, \quad i = 1, 2, \ldots, m \\
L & \leq v \leq U \\
M & \geq 0,
\end{align*}
\]  

where \( M \) is the vector whose \( i \)-th component is \( M_i \). Setting \( M \) to be the calculated optimal solution of (18) tightens the SOCP constraints while ensuring that the optimal growth rate is held at its desired value. We experimented with \( L_2 \) and \( L_\infty \) counterparts; however, the \( L_2 \) norm suffered from inconsistent solves, and the \( L_\infty \) norm overly relaxed constraints, which was somewhat expected.

4.3 Computational Results

The essential gene predictions of RAMP and FBA are tabulated in Tables 3 through 5. All experiments assumed an oxygen rich environment with glucose as the sole, and limiting, carbon source as the default environment. We further conducted the same tests with glycerol as the only carbon source, but we forego their presentation since the results mirrored those of glucose.

The tallies in Table 3 show that the default RAMP model precisely replicates the essential gene predictions of FBA. This fact computationally substantiates Theorem 2 and shows that minor probabilistic variations mimic FBA. Model 2 magnifies the default perturbations by \( \rho \) and ‘spreads’ the scenarios in the last couple of digits, which impacts RAMP’s predictive ability. For the iJO1366 model, all values of \( \rho \) decreased the true positives by 9 and increased the true negatives by 1, which meant that RAMP had 8 fewer correct predictions than FBA with increased scenario variability as magni-
<table>
<thead>
<tr>
<th>Computational</th>
<th>Essential</th>
<th>Nonessential</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Positive</td>
<td>False Negative</td>
</tr>
<tr>
<td>FBA</td>
<td>171 (160)</td>
<td>77 (78)</td>
</tr>
<tr>
<td>Default</td>
<td>162 (160)</td>
<td>86 (78)</td>
</tr>
<tr>
<td>$2 \leq \rho \leq 9$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAMP</td>
<td>171 (160)</td>
<td>77 (78)</td>
</tr>
<tr>
<td></td>
<td>162 (160)</td>
<td>86 (78)</td>
</tr>
<tr>
<td>FBA</td>
<td>44 (35)</td>
<td>44 (35)</td>
</tr>
<tr>
<td>Default</td>
<td>43 (35)</td>
<td>43 (35)</td>
</tr>
<tr>
<td>$2 \leq \rho \leq 9$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAMP</td>
<td>1074 (987)</td>
<td>1074 (987)</td>
</tr>
<tr>
<td></td>
<td>1075 (987)</td>
<td>1075 (987)</td>
</tr>
<tr>
<td>$2 \leq \rho \leq 9$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Stochastic model 1 and 2 for RAMP as compared to FBA. Results for the iAF1260 metabolic model are in parentheses, all other results are for the iJO1366 model.

The iAF1260 model agreed with FBA for all values of $\rho$. RAMP’s predictive power for model 2 decreased from 91.14% to 90.56% with the iJO1366 model and matched FBA’s 91.32% with the iAF1260 model. The similarity between FBA and RAMP for stochastic models 1 and 2 shows that probabilistic deviations in the growth coefficients beyond what is reported in the standard model lead to minor adjustments in predictive power.

Scaling the scenarios proportionately has a slightly more varied effect than multiplying the default variations by $\rho$. Results with $\sigma \geq 0.4$ had little value since these models relaxed the FBA constraints to the point at which few, if any, essential genes were identified, and for this reason we only report the results for $\sigma \leq 0.3$. The trend as $\sigma$ increases is that RAMP reduces the number of predicted essential genes, which lowers the number of true and false positives but increases the number of true and false negatives. FBA’s predictive ability is 91.14% for the iJO1366 model and 91.03% for the iAF1260 model, which RAMP’s default case mirrors. Five true positives move to false negatives in the iJO1366 model for $0.001 \leq \sigma \leq 0.02$, which reduces RAMP’s predictive ability to 90.78%. However, for larger $\sigma$ RAMP increases it’s number of true negatives and reduces its number of false positives, which increased its predictive ability to 90.92%. The trend for the iAF1260 model is similar.

Altering $\varepsilon$ had no effect on RAMP’s ability to predict essential genes, and for all values of $\varepsilon$, RAMP agreed with FBA. This suggests that it is more important to tune RAMP with regard to the scenarios than it is with the probabilistic guarantee of satisfying the near equilibrium constraints (2) and (3).
<table>
<thead>
<tr>
<th></th>
<th>Essential</th>
<th>Nonessential</th>
<th>FBA</th>
<th>RAMP</th>
<th>Default</th>
<th>0.01 ≤ σ ≤ 0.01</th>
<th>0.01 ≤ σ ≤ 0.1</th>
<th>0.01 ≤ σ ≤ 0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Computational</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Positive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essential</td>
<td>171 (160)</td>
<td>166 (155)</td>
<td>0.001 ≤ σ ≤ 0.01</td>
<td>166 (152)</td>
<td>163 (152)</td>
<td>0.03 ≤ σ ≤ 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonessential</td>
<td>77 (78)</td>
<td>82 (83)</td>
<td>0.001 ≤ σ ≤ 0.01</td>
<td>82 (86)</td>
<td>85 (86)</td>
<td>0.03 ≤ σ ≤ 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essential</td>
<td>171 (160)</td>
<td>171 (160)</td>
<td>0.01 ≤ 2ε ≤ 0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonessential</td>
<td>77 (78)</td>
<td>77 (78)</td>
<td>0.01 ≤ 2ε ≤ 0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Positive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Stochastic model 3 for RAMP as compared with FBA. Results for the iAF1260 metabolic model are in parentheses, all other results are for the iJO1366 model.

<table>
<thead>
<tr>
<th></th>
<th>Essential</th>
<th>Nonessential</th>
<th>FBA</th>
<th>RAMP</th>
<th>Default</th>
<th>0.01 ≤ σ ≤ 0.01</th>
<th>0.01 ≤ σ ≤ 0.1</th>
<th>0.01 ≤ σ ≤ 0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Computational</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Positive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essential</td>
<td>171 (160)</td>
<td>166 (155)</td>
<td>0.001 ≤ σ ≤ 0.01</td>
<td>166 (152)</td>
<td>163 (152)</td>
<td>0.03 ≤ σ ≤ 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonessential</td>
<td>77 (78)</td>
<td>82 (83)</td>
<td>0.001 ≤ σ ≤ 0.01</td>
<td>82 (86)</td>
<td>85 (86)</td>
<td>0.03 ≤ σ ≤ 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essential</td>
<td>171 (160)</td>
<td>171 (160)</td>
<td>0.01 ≤ 2ε ≤ 0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonessential</td>
<td>77 (78)</td>
<td>77 (78)</td>
<td>0.01 ≤ 2ε ≤ 0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Positive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: RAMP was insensitive to changes in ε.
5 Conclusions & Continuations

The SOCP constraints of RAMP appropriately inject statistical variation into the FBA paradigm, and RAMP’s stochastic adaptation of FBA results in a model whose trend is to predict fewer essential genes. This computational outcome is not immediate mathematically, which is illustrated by the fact that feasible FBA fluxes for which $\|Rv\|$ dominates $M_i$ are infeasible in RAMP. Hence RAMP is not an immediate relaxation of FBA. However, the computational results demonstrate that RAMP is practically a relaxation of FBA. This outcome favorably agrees with the biological sentiment that cellular metabolic networks are more robust in their wild-type setting and increasingly fragile as they evolve toward an optimal, steady-state. Hence, the number of essential genes should reduce as statistical variation increases, which is predicted by RAMP. Moreover, our mathematical analysis and computational outcomes show that RAMP mimics FBA if the stochastic elements are sufficiently small. In particular, if variation is only allowed beyond the assumed accuracy of an FBA model, then RAMP behaves like FBA.

The computational limitation of allowing only the growth coefficients to vary restricts our ability to vet RAMP. While theoretically we should be able to solve nonlinear RAMP models as SOCPs, this has not proved to be a computational reality. We did achieve interspersed success with native SOCP solvers for general RAMP models, and specifically, we have attempted to

- solve $L_2$ versions of (18);
- predict essential genes with a RAMP version of minimization of metabolic adjustment (MOMA) [20], which is a quadratic model; and
- increase variation beyond the growth coefficients to replicate stochastic environmental bounds.

We were unable to adjust settings and convergence criteria to achieve steadfast success for any of these extensions. Establishing a stable computational platform for the general RAMP paradigm is an important goal for future work.

Theorem 3 provides a way to discriminate FBA’s optimal flux states. For example, one FBA solution may have no biologically possible scenarios while another may have many. This suggests that the former is less stable with regard to random variation whereas the latter is optimal under a wider range of stochastic possibilities. One may intuit that actual flux states should be robust against random variation, and hence, learning to calculate FBA solutions that can witness as much variability as possible is a promising direction of future research.

References


