A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks

R. M. T. Fleming
Center for Systems Biology, University of Iceland
+354 618 6245, ronan.mt.fleming@gmail.com

C. M. Maes
Institute for Computational and Mathematical Engineering, Stanford University

M. A. Saunders and Y. Ye
Department of Management Science and Engineering, Stanford University

B. Ø. Palsson
Department of Bioengineering, University of California, San Diego

Technical Report SOL 2011-1

Abstract

We derive a convex optimization problem on a steady-state nonequilibrium network of biochemical reactions, with the property that energy conservation and the second law of thermodynamics both hold at the problem solution. This suggests a new variational principle for biochemical networks that can be implemented in a computationally tractable manner. We derive the Lagrange dual of the optimization problem and use strong duality to demonstrate that a biochemical analogue of Tellegen's theorem holds at optimality. Each optimal flux is dependent on a free parameter that we relate to an elementary kinetic parameter when mass action kinetics is assumed.

Keywords: constraint-based modeling, flux balance analysis, thermodynamics, convex optimization, entropy function

1. Introduction

The biochemical system of any organism can be represented mathematically by a network of chemicals (nodes) and reactions (edges). To analyze these networks at genome scale, systems biologists often use a linear optimization technique called flux balance analysis (FBA) [38]. Flux balance requires that the sum of fluxes into and out of each node in the network be zero. This is equivalent to Kirchhoff's current law in an electrical network.

Recent work has sought to augment flux balance analysis with Kirchhoff's loop law for energy conservation as well as the second law of thermodynamics [5, 32, 46, 27, 17, 39].

The incorporation of thermodynamic constraints into genome-scale models has produced models that are biologically more realistic and reveal greater insight into the control mechanisms operating in these complex biological systems [5, 25, 45, 18, 26]. See [41] for a recent broad review of the application of thermodynamic constraints to biochemical networks. The second law of thermodynamics may be applied to each reversible elementary reaction by constraining net reaction flux to be in the direction of a negative change in chemical potential [9]. Constraints on net reaction flux can be incorporated within flux balance analysis as additional linear inequality constraints [16]. Software packages to quantitatively assign reaction directionality for genome-scale models of metabolism are available [15, 40].

In contrast to the addition of constraints arising from the second law of thermodynamics, the addition of energy conservation constraints has been problematic because the resulting equations are nonlinear and/or nonconvex. Previous attempts required computing the global solution of a nonconvex continuous optimization problem [5, 27, 17], solving an NP-hard problem [46], or solving a mixed integer linear program [19, 39]. Mixed integer programs have unpredictable computational complexity.

The purpose of this work is to show that Kirchhoff's loop law and the second law of thermodynamics arise naturally from the optimality conditions of a convex optimization problem with flux balance constraints. Furthermore, every set of reaction fluxes that satisfies Kirchhoff's loop law and the second law of thermodynamics must be optimal for some instance of this problem. This suggests that there is an underlying variational principle operating in biochemical networks, and leads to an efficient (scalable) method for computing steady sate fluxes that also satisfy energy conservation and the second law of thermodynamics.

2. Linear resistive networks

Consider a simple electrical circuit consisting of current sources, batteries, and resistors, as illustrated in Figure 1.

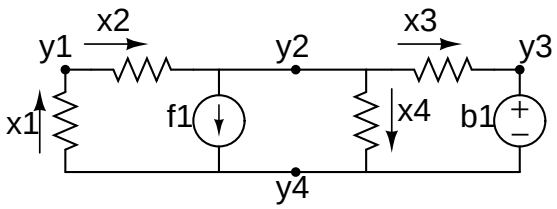

Figure 1: A linear resistive network with currents x, potentials y, batteries b, and current sources f.

This is a linear resistive network with m nodes and n edges, where the node variables

$y \in \mathbb{R}^m$ represent potentials and the edge variables $x \in \mathbb{R}^n$ represent flows (or currents) in the network. The circuit topology is defined by a node-edge incidence matrix $A \in \mathbb{R}^{m \times n}$, and properties of the network are encoded in a set of data vectors: $f \in \mathbb{R}^m$ is a vector of current sources, $b \in \mathbb{R}^n$ is a vector of batteries, and $r \in \mathbb{R}^n$ is a vector of resistances ($r > 0$).

To solve for the voltages and currents in the circuit we use three fundamental laws: Kirchhoff's current law (KCL) $Ax = f$, Kirchhoff's voltage (or loop) law (KVL) $w = b + A^T y$, and Ohm's law $w = Rx$ (where $w \in \mathbb{R}^n$ is a vector of voltages and $R = \mathrm{diag}(r)$ is a positive-definite matrix).

A variational principle underlies the circuit, which seeks a set of currents that minimize the heat (or power) dissipated subject to KCL. This is the convex optimization problem

$$\begin{array}{ll} \text{minimize} & F(x) \equiv \frac{1}{2} x^T R x - b^T x \\ \text{subject to} & Ax = f \quad : y \end{array} \tag{QP}$$

where the node variables y are Lagrange multipliers for the equality constraints. The optimality conditions $\nabla F(x) = A^T y$ yield equations that enforce KVL and Ohm's Law, and the optimal variables $x^\star$ and $y^\star$ are a set of consistent potentials and currents for the circuit.

Biochemical networks are significantly more complicated than linear resistive networks. However, some of the same underlying network concepts apply [29]. In this work we construct an optimization problem in a form similar to problem (QP), where the potentials are Lagrange multipliers for an equality constraint on the flow variables, and the optimality conditions of the problem yield equations that enforce Kirchhoff's loop law and the second law of thermodynamics. Previous work noted a connection between Lagrange multipliers and chemical potentials in flux balance analysis [44]. Here, we establish a quantitative relation between the Lagrange multipliers and chemical potential for this new optimization problem.

3. Biochemical networks

The mathematical representation of a biochemical network is the *stoichiometric matrix* S. Like A above, $S \in \mathbb{R}^{m \times n}$ is a sparse incidence matrix that encodes the network topology. However, biochemical networks, unlike linear resistive networks, are nonlinear networks or hypergraphs. That is, a single edge may link many nodes to many nodes, and the entries in S, which are integer stoichiometric coefficients, are not confined to the set $\{-1, 0, 1\}$. Each row of S corresponds to an individual chemical compound, and each column of S corresponds to an individual elementary reaction. In practice, $m < n$ and S does not have full row-rank. A model of a system is called genome-scale if a large proportion of the system's genes are represented. In current genome-scale models of the metabolic system of *E. coli*, m and n are several thousand.

Flux balance analysis (FBA) computes a set of fluxes that satisfy steady state mass-conservation constraints and are optimal for a biological objective function [38, 28]. A *flux* is a reaction rate; it represents flow through the network and is analogous to current in an electrical circuit. We denote the net flux of the jth reaction by the variable $v_j \in \mathbb{R}$. The concentration of the ith chemical in the network is denoted $x_i \in \mathbb{R}$. A fundamental equation in flux balance analysis is the dynamic mass conservation equation—a differential equation relating the change in chemical concentration to reaction fluxes via the stoichiometric matrix:

$$Sv = \frac{dx}{dt}.$$

Here $x \in \mathbb{R}^m$ is a vector of chemical concentrations and $v \in \mathbb{R}^n$ is a vector of net reaction fluxes. Each row of this vector equation states that the rate of change in concentration for a chemical is the sum of fluxes that synthesize or degrade that chemical.

So far we have considered net flux for a stoichiometric matrix of reactions, each of which conserves mass. A living biochemical system operates in a *nonequilibrium state* [33] and exchanges mass with its surroundings. This can be modeled by including exchange reactions that do not conserve mass. The model is augmented with a matrix $S_e \in \mathbb{R}^{m\times k}$ and a corresponding set of *exchange fluxes* v_e, which are sources and sinks of chemicals and are analogous to current sources in an electrical network. If we assume that the biochemical system is operating at a steady state, then the concentrations of chemicals within the system remain constant. Thus, we have

$$Sv + S_e v_e = \frac{dx}{dt} \equiv 0. \tag{1}$$

This equation is analogous to Kirchhoff's current law in an electrical network. Henceforth we assume that there exists a feasible solution (v, v_e) for (1). Ensuring the existence of a steady state flux is part of a quality control process during reconstruction of S and S_e from experimental literature [43].

Any reversible reaction $A + 2B \rightleftharpoons C$ may be split into two one-way reactions: forward $(A + 2B \rightarrow C)$ and reverse $(A + 2B \leftarrow C)$. To distinguish between forward, reverse, and exchange reactions we split the augmented stoichiometric matrix into three components: $[S \;\; -S \;\; S_e]$. Here S contains all the columns corresponding to forward reactions, and $-S$ contains all columns corresponding to the reverse reactions. The fluxes for these one-way reactions form the vectors v_f and v_r. The *net flux* is then $v = v_f - v_r$.

Flux balance analysis has been implemented using linear programming [30]. We take the flux balance analysis problem to be

$$\begin{array}{ll} \underset{v_f, v_r, v_e}{\text{maximize}} & d^T v_e \\ \text{subject to} & Sv_f - Sv_r + S_e v_e = 0 \\ & v_f, v_r \geq 0, \quad \ell \leq v_e \leq h. \end{array} \tag{FBA}$$

Often, the lower bounds ℓ and upper bounds h on the exchange fluxes come from laboratory measurements (*e.g.* the uptake of glucose in a particular culture of *E. coli*). The vector d is chosen to optimize a biological objective (*e.g.* maximizing replication rate in unicellular organisms). Note that flux balance analysis does not explicitly solve for v_f and v_r but rather $v = v_f - v_r$.

We now prove a lemma about alternative optimal solutions to problem (FBA).

Lemma 1. *If there is an optimal solution to problem (FBA), there is an optimal solution with strictly positive internal fluxes v_f and v_r.*

Proof. Let $(v_f^\star, v_r^\star, v_e^\star)$ be an optimal solution to problem (FBA). It could be the case that one or more components of $v_f^\star$ or $v_r^\star$ are zero; that is, they are sitting on their lower bounds. We want to show that we can construct a new optimal solution $(v_f^o, v_r^o, v_e^\star)$ with strictly positive fluxes v_f^o and v_r^o. To do this, let $v_f^o = v_f^\star + \alpha e$ and $v_r^o = v_r^\star + \alpha e$, where α is any positive scalar and $e \in \mathbb{R}^n$ is the vector of all ones. The internal fluxes (v_f^o, v_r^o) are strictly positive, and the set of fluxes $(v_f^o, v_r^o, v_e^\star)$ is feasible because $\ell \leq v_e^\star \leq h$ and

$$\begin{aligned} Sv_f^o - Sv_r^o + S_e v_e &= S(v_f^\star + \alpha e) - S(v_r^\star + \alpha e) + S_e v_e^\star \\ &= Sv_f^\star - Sv_r^\star + S_e v_e^\star = 0. \end{aligned}$$

Finally, the set of fluxes $(v_f^o, v_r^o, v_e^\star)$ is optimal because it has the same objective value $d^T v_e^\star$ as the optimal solution $(v_f^\star, v_r^\star, v_e^\star)$. Here we used the fact that the objective function in problem (FBA) depends only on the exchange fluxes v_e; not on the internal fluxes v_f and v_r. □

4. Thermodynamic constraints

Flux balance analysis predicts fluxes that satisfy steady state mass conservation but not necessarily energy conservation or the second law of thermodynamics. Whilst the domain of steady state mass conserved fluxes includes those that are thermodynamically feasible, additional constraints are required in order to guarantee a flux that additionally satisfies energy conservation and the second law of thermodynamics. Recent work has tried to add constraints based on Kirchhoff's loop law and the second law of thermodynamics to problem (FBA) [5, 32, 27]. However, these constraints are problematic because they are nonlinear and nonconvex. We now describe these constraints.

The loop law for chemical potentials in a biochemical network is directly analogous to Kirchhoff's voltage law for electrical circuits. It states that the stoichiometrically weighted sum of chemical potentials around any closed loop of chemical reactions is zero. We explicitly model a chemical potential $u_i \in \mathbb{R}$ for each of the chemicals in the network, and we define the *change in chemical potential* for all internal reactions in the network as the vector

$$\Delta u \equiv S^T u \quad (\in \mathbb{R}^n), \tag{2}$$

where $u \in \mathbb{R}^m$ is the vector of chemical potentials. This ensures that Kirchhoff's loop law is satisfied, as energy conservation requires that an injective relation exist between each row of S and a chemical potential [31]. It follows that the change in chemical potential around a stoichiometrically balanced loop will be zero.

Assuming mass-action kinetics, constant temperature and pressure, and uniform spatial concentrations (*i.e.* a well mixed system), it is known (*e.g.* [37]) that the change in chemical potential may be expressed in terms of elementary one-way reaction rates as

$$\Delta u = \rho \log \left(v_r \,./\, v_f \right), \tag{3}$$

where $./$ denotes component-wise division of vectors, and $\rho = RT > 0$ is the gas constant multiplied by temperature. Equation (3) leads directly to a macroscopic (long-term) application of the second law of thermodynamics:

$$-\Delta u_j \, v_j = -\Delta u_j \left(v_{fj} - v_{rj} \right) \geq 0,$$

which says that the net flux for each elementary reaction must be down a gradient of chemical potential, that the system must dissipate heat, and that entropy must increase as a result of work being done on the system through the exchange fluxes [33]. The total heat dissipation rate of the biochemical system in a non-equilibrium steady state is given by $-\Delta u^T v \geq 0$. Henceforth we consider each flux as a dimensionless quantity. Dimensioned flux can be rendered dimensionless by division with a standard flux of the same dimensions. This standard may be defined according to a convenient timescale in the same way that a standard concentration may be defined according to a convenient abundance scale.

Under the specified assumptions, we now define a thermodynamically feasible flux.

Definition 1. For a network described by an augmented stoichiometric matrix $[S \;\; -S \;\; S_e]$ with a given set of exchange fluxes v_e, a set of *thermodynamically feasible fluxes* is a pair of internal flux vectors $(v_f, v_r) > 0$ that satisfy steady-state mass-balance,

$$S v_f - S v_r = -S_e v_e, \tag{4}$$

and for which there exists an underlying vector of chemical potentials $u \in \mathbb{R}^m$ that satisfies (2) and (3):

$$\Delta u \equiv S^T u = \rho \log \left(v_r \,./\, v_f \right) = \rho \log v_r - \rho \log v_f. \tag{5}$$

5. A variational principle

We now present the main theorem of this paper. The theorem introduces a new convex optimization problem with the same flux balance constraints as problem (FBA) but with a negative entropy objective function. It states that the thermodynamic constraints (4) and (5) hold at its unique solution. We use e to denote a vector of ones.

Theorem 1. *Let $v_e^\star$ be any set of optimal exchange fluxes from problem (FBA). Define $b = -S_e v_e^\star$, and let c be any vector in $\mathbb{R}^n$. The convex equality-constrained problem*

$$\begin{array}{ll} \underset{v_f, v_r > 0}{\text{minimize}} & \phi \equiv v_f^T(\log(v_f) + c - e) + v_r^T(\log(v_r) + c - e) \\ \textit{subject to} & Sv_f - Sv_r = b \quad : y \end{array} \tag{EP}$$

is then feasible, and its solution $(v_f^\star, v_r^\star)$ is a set of thermodynamically feasible internal fluxes. The combined vector $(v_f^\star, v_r^\star, v_e^\star)$ is thermodynamically feasible and optimal for problem (FBA). The associated chemical potentials u may be obtained from the optimal Lagrange multiplier $y^\star \in \mathbb{R}^m$ for the equality constraints according to $u = -2\rho y^\star$.

Proof. First note that the constraints $v_f, v_r > 0$ are implied by the domain of the logarithm; they have no associated Lagrange multipliers. From Lemma 1, we know that if $v_e^\star$ is optimal for problem (FBA) and $b = -S_e v_e^\star$, there must be corresponding positive internal fluxes v_f and v_r that satisfy $Sv_f - Sv_r = b$. Therefore, problem (EP) with this choice of b is always feasible.

Define the objective function as $\phi(v_f, v_r)$ and note that it is strictly convex because $\nabla^2\phi(v_f, v_r)$ is positive-definite for all $v_f, v_r > 0$, and it is bounded below. Problem (EP) is thus a convex linear equality-constrained problem with a unique optimal solution $(v_f^\star, v_r^\star)$ that satisfies the optimality conditions

$$S^T y^\star = \nabla_{v_f}\phi = \log(v_f^\star) + c, \tag{6}$$

$$-S^T y^\star = \nabla_{v_r}\phi = \log(v_r^\star) + c, \tag{7}$$

$$Sv_f^\star - Sv_r^\star = b \tag{8}$$

for some vector $y^\star$ (which is not unique because S has low row rank). Subtracting (6) from (7) gives $S^T(-2y^\star) = \log(v_r^\star ./ v_f^\star)$. Taking $u = -2\rho y^\star$ we see from (5) that $(v_f^\star, v_r^\star)$ is a pair of thermodynamically feasible fluxes with underlying chemical potentials u for the exchange fluxes $v_e^\star$. The combined vector $(v_r^\star, v_f^\star, v_e^\star)$ is feasible for problem (FBA), and is optimal because the objective $d^T v_e^\star$ is unchanged. □

In summary, to compute an optimal solution $(v_f^\star, v_r^\star, v_e^\star)$ to problem (FBA) that is thermodynamically feasible, perform the following steps: solve problem (FBA) to find an optimal exchange flux vector $v_e^\star$, form $b = -S_e v_e^\star$, choose a vector c, and solve problem (EP) to find $v_f^\star, v_r^\star > 0$. Since S is row rank deficient, problem (EP) defines Δu uniquely, but not the part of u in the nullspace of S^T.

Note the structural similarity of problems (QP) and (EP). In both cases the primal variables are the flows in the networks, the constraints impose Kirchhoff's current law, and the dual variables for these constraints are the potentials in the network. There is a clear physical interpretation of the objective function in problem (QP) and a variational principle in operation. We believe there must be a variational principle in operation in

the biochemical networks as well. However, it is not yet clear to the authors what this is, or why it appears in the form of the negative entropy objective. Perhaps, by maximizing the entropy of the internal fluxes, the optimization problem produces the most unbiased prediction [24] of internal elementary fluxes, subject to mass-conservation constraints and boundary conditions imposed by exchange fluxes.

The next theorem proves that if there is a set of thermodynamically feasible fluxes in a biochemical network, it *must* be the solution to an optimization problem in the form of problem (EP).

Theorem 2. *Every set of thermodynamically feasible fluxes v_f, v_r (and chemical potentials u) is the solution (and corresponding Lagrange multiplier) of a convex optimization problem in the form of problem (EP).*

Proof. We show how to choose the vector $c \in \mathbb{R}^n$ so that the given v_f, v_r, and u are optimal. Since v_f and v_r satisfy (4), they are feasible. From (5) we have $S^T u = \rho \log(v_r ./ v_f)$. Letting $y = -\frac{1}{2\rho}u$ we have $u = -2\rho y$ and hence

$$-2S^T y = \log(v_r) - \log(v_f).$$

If we take $2c = -\log(v_r) - \log(v_f)$, this gives

$$\begin{aligned} 2S^T y &= 2\log(v_f) + 2c \\ -2S^T y &= 2\log(v_r) + 2c, \end{aligned}$$

which with (4) are the optimality conditions for problem (EP) as given by (6)–(8). Therefore, with $c = -\frac{1}{2}\log(v_r) - \frac{1}{2}\log(v_f)$, the given v_f and v_r are the optimal solution, with Lagrange multiplier $y = -\frac{1}{2\rho}u$. □

The vector c in problem (EP) is a set of free parameters. Theorem 1 tells us that regardless of the value of c, (EP) has a unique solution that is a thermodynamically feasible flux. Theorem 2 states that given any thermodynamically feasible flux there exists at least one associated vector c.

6. Dual variational principle

Given the primal optimization problem (EP) it is instructive to consider the equivalent dual optimization problem, as its objective function lends insight into the properties of the optimal dual variables for the primal problem.

By Lemma 1, the constraints of Problem (EP) have feasible solutions that are strictly positive. Similarly, there must be feasible solutions that are finite, and these have finite objective values. Hence the optimum of the strictly convex objective function of problem (EP) is finite and attainable.

Proposition 1. *The Lagrange dual of problem (EP) is the unconstrained convex optimization problem*

$$\underset{y}{\text{maximize}} \;\; \psi(y) \equiv b^T y - e^T \exp(S^T y - c) - e^T \exp(-S^T y - c). \tag{DEP}$$

Proof. The dual for problem (EP) is derived in terms of the associated Lagrangian,

$$\mathcal{L}\left(v_f, v_r, y\right) = v_f^T(\log(v_f) + c - e) + v_r^T(\log(v_r) + c - e) - y^T(Sv_f - Sv_r - b).$$

The dual function $\psi(y)$ is the set of greatest lower bounds (denoted inf for infimum) of the Lagrangian over the primal variables. Equivalently, it is the set of least upper bounds (denoted sup for supremum) of the negative Lagrangian:

$$\psi(y) = \inf_{v_f, v_r} \; \mathcal{L}\left(v_f, v_r, y\right) = - \sup_{v_f, v_r} \; -\mathcal{L}\left(v_f, v_r, y\right).$$

The supremum of the linear function $y^T b$ is itself, and the other terms may be grouped to give

$$\psi(y) = b^T y - \sup_{v_f} \left(y^T S v_f - v_f^T(\log(v_f) + c - e)\right) - \sup_{v_r} \left(-y^T S v_r - v_r^T(\log(v_r) + c - e)\right). \tag{9}$$

The first supremum is attained when the partial derivative with respect to v_f is zero:

$$S^T y - \log(v_f) - c = 0 \qquad \Leftrightarrow \qquad v_f = \exp(S^T y - c). \tag{10}$$

Similarly for the second supremum,

$$-S^T y - \log(v_r) - c = 0 \qquad \Leftrightarrow \qquad v_r = \exp(-S^T y - c). \tag{11}$$

Substituting (10)–(11) into (9) gives the Lagrange dual problem (DEP). □

The first and second derivatives of $\psi(y)$ are

$$\begin{aligned} \nabla\psi(y) &= b - S\exp(S^T y - c) + S\exp(-S^T y - c), \\ \nabla^2\psi(y) &= -SDS^T, \end{aligned}$$

where $D \equiv \text{diag}(\exp(S^T y - c) + \exp(-S^T y - c))$ is a positive-definite diagonal matrix for all finite y (so that $\nabla^2\psi(y)$ is negative semidefinite). The necessary first- and second-order conditions for a point $y = y^\star$ to be an optimum are that $\nabla\psi(y^\star) = 0$ and $\nabla^2\psi(y^\star)$ be negative semidefinite. With $v_f^\star \equiv \exp(S^T y^\star - c)$ and $v_r^\star \equiv \exp(-S^T y^\star - c)$ we see that $(v_f^\star, v_r^\star, y^\star)$ satisfy the optimality conditions for both (EP) and (DEP).

The dual objective $\psi(y)$ offers a complementary insight into the meaning of our vari-

ational principle. The chemical potential is defined by $u = -2\rho y^\star$. Up to scalar multiplication, $-b^T y = \frac{1}{2\rho} b^T u$ corresponds to the rate of work being done by the environment to maintain the system away from equilibrium [34]. Therefore, one may interpret the Lagrange dual problem as a minimization of this rate of work, balanced against minimization of the sum of thermodyamically feasible forward and reverse fluxes. We note that minimization of a weighted linear combination of forward or reverse fluxes, where thermodynamic equilibrium constants are used as weighting factors, has previously been explored as an optimality principle for metabolic networks [21].

6.1. Strong duality and Tellegen's theorem

Since problem (EP) is convex with a finite attainable optimum objective and linear constraints, it is known that strong duality holds [7]. Among other things, strong duality means that the values of the primal and dual objectives are equal at the optimum: $\phi(v_f^\star, v_r^\star) = \psi(y^\star)$. Omitting the *s and using (10)–(11), we have

$$\begin{aligned} v_f^T(\log(v_f) + c - e) + v_r^T(\log(v_r) + c - e) &= b^T y - e^T \exp(S^T y - c) - e^T \exp(-S^T y - c) \\ &= b^T y - e^T v_f - e^T v_r. \end{aligned}$$

Thus,

$$v_f^T(\log(v_f) + c) + v_r^T(\log(v_r) + c) = b^T y. \tag{12}$$

Also from (10)–(11) we have

$$c = -\log(v_f) + S^T y = -\log(v_r) - S^T y.$$

With $u = -2\rho y$, (12) becomes

$$u^T S(v_f - v_r) = u^T b. \tag{13}$$

On the left of (13) is the rate of entropy production by the biochemical network, and on the right is the rate of work done by the environment to maintain the system away from equilibrium. Their equality means that the rate at which heat is produced by chemical reactions equals the rate at which heat is dissipated into the environment, so the temperature of the system is time-invariant. In the thermodynamic literature, (13) is known as the isothermal Clausius equality [34], whereas from electrical network literature one may recognize (13) as a biochemical analogue of Tellegen's theorem.

A comprehensive introduction to convex optimization is given in [8]. Problem (EP) is an example of a monotropic optimization problem [35]. Further discussion of such problems (and duality) can be found in [7].

7. Thermodynamic feasibility and mass action kinetics

From the optimality conditions for problem (EP), observe that forward flux is a function of substrate and product chemical potentials. For an elementary reaction, one would expect that forward flux should depend only on substrate chemical potential(s) and reverse flux should depend only on product chemical potential(s) [13]. According to mass action kinetics, the rate of an elementary forward reaction only depends on a forward kinetic parameter and substrate concentration(s) [10]. In the objective of problem (EP), for simplicity of exposition, consider replacing $c^T(v_f + v_r)$ with $c_f^T v_f + c_r^T v_r$. When $c_f - c_r$ is constrained to lie in the range of S^T, this provides another mechanism for satisfying (5). Here we show that one may choose c_f, c_r such that mass action kinetics also holds.

The chemical potential of the ith metabolite in dilute solution at constant temperature and pressure can be expressed as

$$u_i = u_i^o + \rho \log\left(x_i / x_i^o\right), \tag{14}$$

where u_i^o is the standard chemical potential, x_i is the molar concentration of the metabolite, and x_i^o is a reference concentration that we define to be one molar. The standard chemical potential is therefore the chemical potential of a metabolite in a reference state. Assume we are given forward and reverse elementary kinetic parameter vectors $k_f, k_r \in \mathbb{R}^n$ satisfying the thermodynamic feasibility condition

$$S^T u^o = \rho \log(k_r \,./\, k_f) \tag{15}$$

for a particular standard chemical potential vector $u^o \in \mathbb{R}^m$. Mass action kinetics may be expressed as a set of vectors v_f, v_r, k_f, k_r, x that satisfy

$$\begin{aligned} \log(v_f) &= \log(k_f) + F^T \log(x), \\ \log(v_r) &= \log(k_r) + R^T \log(x), \end{aligned}$$

where the stoichiometric matrix has been decomposed into the entry-wise difference between forward and reverse stoichiometric matrices, respectively $F, R \in \mathbb{R}^{m,n}_{\geq 0}$, defined by

$$\begin{array}{rcl} S_{ij} = 0 & \rightarrow & \{F_{ij} = 0, \quad R_{ij} = 0\}, \\ S_{ij} < 0 & \rightarrow & \{F_{ij} = -S_{ij}, \; R_{ij} = 0\}, \\ S_{ij} > 0 & \rightarrow & \{F_{ij} = 0, \quad R_{ij} = S_{ij}\}. \end{array}$$

With respect to the forward reaction direction, each column of F contains the absolute value of the stoichiometric coefficient for each substrate, and the corresponding column of R contains the stoichiometric coefficient for each product.

Given a consistent stoichiometric matrix and thermodynamically feasible kinetic parameters, it is still an important open question whether for all $b \in \mathcal{R}(S)$ there exists a

metabolite concentration vector x satisfying steady-state mass action kinetics, namely

$$S\exp(\log(k_f) + F^T\log(x)) - S\exp(\log(k_r) + R^T\log(x)) = b. \tag{16}$$

Ongoing work aims to establish the necessary and sufficient conditions for this to be so [1]. Nonetheless, assuming (16) is satisfied, there exist c_f and c_r such that

$$c_f = -\log(k_f) + R^T y^\star, \qquad c_r = -\log(k_r) + F^T y^\star,$$

and optimality conditions (10)–(11) become

$$\log(v_f^\star) = \log(k_f) - F^T y^\star, \qquad \log(v_r^\star) = \log(k_r) - R^T y^\star.$$

The optimal dual vector may then be equated with the negative of logarithmic concentration $y^\star = -\log(x)$, leading to satisfaction of mass action kinetics in a non-equilibrium steady state.

8. Discussion

With problem (EP) and Theorem 1 we provide a new variational principle that enforces steady state mass conservation, energy conservation and the second law of thermodynamics for genome-scale biochemical networks. Moreover, we prove that all thermodynamically feasible steady state fluxes are instances of problem (EP) for a different free parameter vector $c \in \mathbb{R}^n$ (Theorem 2). The values of the free parameters influence the optimal forward and reverse fluxes through the relation $v_f^\star .* v_r^\star = \exp(-2c)$. Varying c may change $v_f^\star$ and $v_r^\star$ but not the fact that the corresponding solution is thermodynamically feasible and optimal for flux balance analysis.

To compute a thermodynamically feasible flux we must first choose a value for the vector c. Thus, there is some freedom left in the model. We see this property of the model as an advantage rather than a disadvantage, as thermodynamic feasibility is necessary but not sufficient for satisfaction of mass action kinetics [22]. In the objective of problem (EP), one could replace $c^T(v_f + v_r)$ by $c_f^T v_f + c_r^T v_r$. Assuming the existence of a mass action kinetic non-equilibrium steady state, section 7 demonstrates that this state corresponds to a particular choice of c_f, c_r given particular kinetic parameters k_f, k_r. Kinetic parameters evolve, in the biological sense, subject to thermodynamic feasibility (15). Even if c_f, c_r could be chosen such that mass action kinetics holds, the actual k_f, k_r that pertain to a particular organism's biochemical network would remain unknown. This problem is not particular to our modeling approach—it reflects a paucity of suitable enzyme kinetic data in biochemistry generally [11].

In addition to the constraints in problem (FBA), flux balance analysis often includes extra inequalities on the net flux of internal reactions. It is not possible to incorporate explicit inequality constraints on net flux into problem (EP) because if any such inequalities

were active at the optimum, the corresponding nonzero dual variables would appear in a modification of the optimality conditions (10)–(11) and interfere with satisfaction of the thermodynamic constraints (2) that we know must hold. In practical applications of problem (EP) to genome-scale biochemical networks, for arbitrarily chosen c, the omission of explicit bounds on net flux leads to a subset of net fluxes with directions opposite to that known biochemically.

At this point, one might think that an obvious application would be to search for free parameters c_f, c_r that minimize the Euclidean distance between predicted and experimentally determined change in chemical potentials or fluxes. In *E. coli* metabolism [16], for example, one can experimentally quantify transformed reaction Gibbs energy (*in vivo* change in chemical potential) using quantitative measurement of absolute concentrations [6], combined with experimentally derived [2, 3] or group contribution estimates [20, 19, 23, 14] of standard transformed reaction Gibbs energy. However, obtaining *in vivo* chemical potentials is not a limitation to validation of the practical utility of problem (EP). Even if one could find free parameters c_f, c_r corresponding to predicted chemical potentials close to experimental data, the corresponding fluxes might not satisfy mass action kinetics, which is known to hold for elementary chemical reactions.

A reliable algorithm for satisfaction of mass action kinetics at a non-equilibrium steady state is the subject of ongoing work [1]. Given a consistent stoichiometric matrix S, for there to exist a non-equilibrium steady state it is necessary that the boundary condition be in the range of the stoichiometric matrix: $b \in \mathcal{R}(S)$. Given thermodynamically feasible kinetic parameters in addition, it is an important open problem to establish if $b \in \mathcal{R}(S)$ is also sufficient for there to exist a non-equilibrium steady state satisfying mass action kinetics. It may be that the set of feasible boundary conditions is a (proper) subset of the range of S. In fact, the boundary condition obtained from flux balance analysis $b = -S_e v_e^\star$ is an underdetermined parameter for problem (EP) because $v_e^\star$ may not be unique, given the absence of strict convexity in problem (FBA).

We consider problem (EP) to represent a mapping between a set of parameters and the corresponding set of thermodynamically feasible fluxes and potentials. More precisely, when S is row reduced (to have full row rank), problem (EP) represents a single-valued surjective saddle point mapping [36] between a parameter vector (c, b) and the corresponding primal-dual optimal vector $(v_f^\star, v_r^\star, y^\star)$. That is, each parameter vector corresponds to a unique primal-dual optimal vector, and every optimal vector is associated with at least one parameter (c, b). Moreover, under the same conditions, this saddle point mapping is locally Lipschitz continuous [12]. That is, the optimal vector of fluxes and potentials is smoothly perturbed by a smooth perturbation of the parameter vector.

The properties discussed in the previous paragraph motivate future efforts to design algorithms for gradient-based optimization of the parameter vector. We envisage a convergent sequence of parametric convex optimization problems whose final optimum satisfies mass-action kinetic constraints. However, the necessary and sufficient conditions for convergence of a sequence of parametric convex optimization problems is still an active research

area within convex analysis. Exploitation of the properties of problem (EP) may provide the route to such an algorithm.

The computational tractability of *linear* flux balance analysis (problem (FBA)) is one of the key reasons for its widespread use as a genome-scale modeling tool. The theorems of this paper provide the theoretical basis for efficiently computing thermodynamically feasible fluxes, even for the largest genome-scale biological networks [42]. The convexity of problem (EP) is a crucial property. Efficient polynomial-time algorithms [4] exist for solving convex optimization problems of this form, based on interior methods [47]. These algorithms are guaranteed to return a solution that is optimal to within roundoff error, or a certificate that the original problem is infeasible. In practice, computing thermodynamically feasible fluxes using *convex* flux balance analysis described herein should take no longer than performing *linear* flux balance analysis.

Although the ultimate value of this work lies in what is accomplished with it in future biological studies, we believe it makes an important step forward by providing the first computationally tractable method for implementing energy conservation and the second law of thermodynamics for genome-scale biochemical networks in a non-equilibrium steady state. We also establish, in an exact manner, the duality relationship between reaction rates and chemical potentials. Furthermore, our theorems extend to any potential network where flux balance holds and the change in potential can be written as a difference in a monotone function of fluxes: $\Delta u = g(v_r) - g(v_f)$, where $g(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ is monotone [35]. These potential networks admit a convex optimization model whose solution is unique and whose potentials are Lagrange multipliers for the flux balance constraint.

Acknowledgements

We would like to acknowledge invaluable assistance from Ines Thiele and some helpful thermodynamic discussions with Hong Qian. This project was supported by the U.S. Department of Energy (Office of Advanced Scientific Computing Research and Office of Biological and Environmental Research) as part of the Scientific Discovery Through Advanced Computing program, grant DE-SC0002009.

References

[1] Akle, S., Dalal, O., Fleming, R. M. T., Saunders, M. A., Taheri, N. A. & Ye, Y. 2011. Existence of positive equilibria for mass conserving and mass-action chemical reaction networks with a single-terminal-linkage class. , (in preparation).

[2] Alberty, R. A. 2003. *Thermodynamics of Biochemical Reactions*. Wiley-Interscience, Hoboken, NJ.

[3] Alberty, R. A. 2006. *Biochemical Thermodynamics: Applications of Mathematica*. Wiley-Interscience, Hoboken, NJ.

[4] Andersen, E. D. & Ye, Y. 1998. A computational study of the homogeneous algorithm for large-scale convex optimization. Comput Optim Appl, 10 (3), 243–269.

[5] Beard, D. A., Liang, S. & Qian, H. 2002. Energy balance for analysis of complex metabolic networks. Biophys J, 83 (1), 79–86.

[6] Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J. & Rabinowitz, J. D. 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat Chem Biol, 5 (8), 593–9.

[7] Bertsekas, D. P. 1999. *Nonlinear Programming*. 2nd edition, Athena Scientific, Belmont, Massachusetts.

[8] Boyd, S. & Vandenberghe, L. 2004. *Convex Optimization*. Cambridge University Press, Cambridge, England.

[9] Burton, K., Krebs, H. & Kornberg, H. 1957. *Energy Transformations in Living Matter*. Springer-Verlag, Berlin.

[10] Cornish-Bowden, A. 2004. *Fundamentals of Enzyme Kinetics*. 3rd edition, Portland Press, London.

[11] Cornish-Bowden, A. & Hofmeyr, J. H. S. 2005. Enzymes in context: kinetic characterization of enzymes for systems biology. The Biochemist, 27 (1), 11–13.

[12] Dontchev, A. L. & Rockafellar, R. T. 2001. Primal-dual solution perturbations in convex optimization. Set-Valued Analysis, 9 (1), 49–65.

[13] Ederer, M. & Gilles, E. 2007. Thermodynamically feasible kinetic models of reaction networks. Biophys J, 92 (6), 1846–1857.

[14] Finley, S., Broadbelt, L. & Hatzimanikatis, V. 2009. Thermodynamic analysis of biodegradation pathways. Biotechnol Bioeng, 103 (3), 532–541.

[15] Fleming, R. M. T. & Thiele, I. 2011. von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models. Bioinformatics, 27 (1), 142.

[16] Fleming, R. M. T., Thiele, I. & Nasheuer, H. P. 2009. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*. Biophys Chem, 145 (2-3), 47–56.

[17] Fleming, R. M. T., Thiele, I., Provan, G. & Nasheuer, H. P. 2010. Integrated stoichiometric, thermodynamic and kinetic modeling of steady state metabolism. J Theor Biol, 264, 683–92.

[18] Garg, S., Yang, L. & Mahadevan, R. 2010. Thermodynamic analysis of regulation in metabolic networks using constraint-based modeling. BMC Res Notes, 3, 125.

[19] Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. 2007. Thermodynamics-based metabolic flux analysis. Biophys J, 92 (5), 1792–1805.

[20] Henry, C. S., Jankowski, M. D., Broadbelt, L. J. & Hatzimanikatis, V. 2006. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. Biophys J, 90 (4), 1453–1461.

[21] Holzhütter, H. 2004. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur J Biochem, 271 (14), 2905–2922.

[22] Jamshidi, N. & Palsson, B. Ø. 2008. Formulating genome-scale kinetic models in the post-genome era. Mol Sys Bio, 4, 171.

[23] Jankowski, M. D., Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. 2008. Group contribution method for thermodynamic analysis of complex metabolic networks. Biophys J, 95 (3), 1487–1499.

[24] Jaynes, E. T. 2003. *Probability Theory – The Logic of Science*. Cambridge University Press, Cambridge.

[25] Kümmel, A., Panke, S. & Heinemann, M. 2006. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. Mol Sys Bio, 2, e34.

[26] Liebermeister, W., Uhlendorf, J. & Klipp, E. 2010. Modular rate laws for enzymatic reactions: Thermodynamics, elasticities and implementation. Bioinformatics, 26 (12), 1528.

[27] Nagrath, D., Avila-Elchiver, M., Berthiaume, F., Tilles, A. W., Messac, A. & Yarmush, M. L. 2007. Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. Ann. Biomedical Eng. 35 (6), 863–885.

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

[29] Oster, G., Perelson, A. & Katchalsky, A. 1971. Network thermodynamics. Nature, 234 (5329), 393–399.

[30] Palsson, B. Ø. 2006. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, Cambridge.

[31] Planck, M. 1945. *Treatise on Thermodynamics*. Courier Dover Publications, Chelmsford, MA.

[32] Price, N. D., Thiele, I. & Palsson, B. Ø. 2006. Candidate states of *Helicobacter pylori*'s genome-scale metabolic network upon application of "loop law" thermodynamic constraints. Biophys J, 90, 3919–3928.

[33] Qian, H. 2007. Phosphorylation energy hypothesis: Open chemical systems and their biological functions. Annu Rev Phys Chem, 58, 113–142.

[34] Qian, H. & Beard, D. A. 2005. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. Biophys Chem, 114 (2-3), 213–220.

[35] Rockafellar, R. T. 1984. *Network Flows and Monotropic Optimization*. Wiley, New York.

[36] Rockafellar, R. T. & Wets, R. J. B. 1997. *Variational Analysis*. Springer-Verlag, Berlin.

[37] Ross, J. 2008. *Thermodynamics and fluctuations far from equilibrium*, vol. 80, of *Springer Series in Chemical Physics*. Springer, New York.

[38] Savinell, J. M. & Palsson, B. Ø. 1992. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. J Theor Biol, 154 (4), 421–454.

[39] Schellenberger, J., Lewis, N. E. & Palsson, B. Ø. 2011*a*. Elimination of thermodynamically infeasible loops in steady-state metabolic models. Biophys J, 100 (3), 544–553.

[40] Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E. R., Kang, J., Hyduke, D. & Palsson, B. Ø. 2011*b*. Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2.0. Nat Protoc, .

[41] Soh, K. & Hatzimanikatis, V. 2010. Network thermodynamics in the post-genomic era. Curr Opin Microbiol, 13 (3), 350–357.

[42] Thiele, I., Jamshidi, N., Fleming, R. M. T. & Palsson, B. Ø. 2009. Genome-scale reconstruction of *E. coli*'s transcriptional and translational machinery: A knowledge-base, its mathematical formulation, and its functional characterization. PLoS Comput Biol, 5 (3), e1000312.

[43] Thiele, I. & Palsson, B. Ø. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc, 5, 93–121.

[44] Warren, P. B. & Jones, J. L. 2007. Duality, thermodynamics, and the linear programming problem in constraint-based models of metabolism. Phys Rev Lett, 99 (10), 108101.

[45] Xu, M., Smith, R. & Sadhukhan, J. 2008. Optimization of productivity and thermodynamic performance of metabolic pathways. Industrial & Engineering Chemistry Research, 47 (15), 5669–5679.

[46] Yang, F., Qian, H. & Beard, D. A. 2005. Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry. Metab Eng, 7 (4), 251–259.

[47] Ye, Y. 1997. *Interior Point Algorithms: Theory and Analysis*. Interscience Series in Discrete Mathematics and Optimization, Wiley, New York.