A Pivoting Algorithm for Metabolic Networks in the Presence of Thermodynamic Constraints

R. Nigam
W.W. Hansen Experimental Physics Laboratory
Stanford University
Stanford, CA 94305-4085, USA
rakesh@quake.stanford.edu

S. Liang
Department of Biostatistics and Applied Mathematics
The University of Texas M.D Anderson Cancer Center
1515 Holcombe Blvd, Houston, TX 77030, USA
shoudan@mdanderson.org

Abstract

A linear programming algorithm is presented to constructively compute thermodynamically feasible fluxes and change in chemical potentials of reactions for a metabolic network. It is based on physical laws of mass conservation and the second law of thermodynamics that all chemical reactions should satisfy. As a demonstration, the algorithm has been applied to the core metabolic pathway of E. coli.

1. Introduction

Constraint-based optimization approaches to modeling metabolic pathways are powerful methods for quantitative analysis of behavior of cells, since the constraints are based on fundamental physical laws and allow some predictions to be made without using any unknown parameters. For a given system of reactions in steady state, mass balance on the reactions restricts the space of possible fluxes or rates to the null space of the stoichiometric matrix [8]. This is a constraint that has been used in the flux balance analysis (FBA). In the absence of detailed knowledge about the kinetic parameters and enzyme concentrations, the FBA assumes that the metabolic flux vector in a biological network optimizes some objective, for example, the production of growth or biomass, subject to the mass balance constraint. Other objectives have been used, but the production of biomass can be easily compared with experiment. Inspite of these constraints, the FBA can still give rise to an infinite number of flux vectors. Hence, we need to further constrain the system.

The network structure of metabolic pathways imposes an additional thermodynamic constraint on the flux vector. Recently, these constraints have been applied to metabolic networks to remove fluxes that violate the second law of thermodynamics. The thermodynamic constraint, applicable to non-equilibrium systems in steady state, is a consequence of the second law of thermodynamics according to which the direction of a chemical reaction is from a higher chemical potential to a lower chemical potential [6]. This is similar to the flow of currents from a higher voltage to a lower voltage in an electrical circuit. This constraint is also called the energy balance analysis (EBA) in the literature [1]. The stoichiometric matrix contains all the information regarding the flux balance and energy balance constraints.

In [5] it was discussed that the FBA solution can be decomposed into weightings of three types of pathways. The Type I pathway for example deals with the cycling of ATP, which then drives other cellular processes. Type II pathways are those that have exchange fluxes corresponding to metabolites like ATP, NADH, with the rest of the pathway being an internal cycle. These pathways represent futile cycles. Type III pathways have no exchange fluxes, and these represent internal cycles corresponding to the internal fluxes. It is the presence of these type III loops in the flux direction that violate the second law of thermodynamics [5]. For example, the pyruvate kinase reaction flux could be a part of the type III loop. A simple solution to satisfy the thermodynamic constraint is by setting the type III fluxes to zero [5]. This turns out to be overly restrictive in some situations and not strictly correct in others. The algorithm

proposed here offers a general method of detecting and removing thermodynamically offending loops.

To test if a particular flux pattern is thermodynamically feasible, [2] proposed a method based on matroid theory. Their method has exponential complexity. Also, in [1] a nonlinear optimization method was presented for this problem. In [4] a sign test was proposed which is a necessary condition for testing thermodynamic infeasibility, which when combined with linear programming (LP) is both necessary and sufficient. This led to a polynomial algorithm to detect thermodynamically infeasible fluxes and compute chemical potential changes for any given flux vector. In this paper we propose a method that relies on the sign test to constructively produce thermodynamically feasible solutions by modifying the infeasible fluxes of the FBA in polynomial time. This method is different from the previous approaches, as it is constructive and is based on linear programming. We also give a different proof for the sign test in lemma 1. We illustrate our algorithm by applying it to a part of the metabolic network of *Escherichia coli* [3], and detect the presence of thermodynamically infeasible fluxes. We modify the infeasible flux vector by our algorithm so that it becomes thermodynamically feasible. The amount by which the flux space is reduced can be got by the sign test introduced in [4], and this is related to the number of thermodynamically infeasible loops present in the network.

2. Flux Balance Analysis: Law of Flux Conservation

In this section, we outline mathematically the law of flux conservation which is the starting point of our analysis. It starts with the metabolic reaction diagram of an organism constructed from its genome by identifying the genes that code for the enzymes that catalyze typically several hundred reactions in the cell. In FBA, the law of mass balance is applied to each metabolite which in steady state implies that the incoming fluxes or rates should balance the outgoing fluxes. The FBA has been formulated as a LP by many authors [8], in which there is a linear objective function $Z = \boldsymbol{d}^T \boldsymbol{f}$ to be maximized or minimized. Usually written as a linear combination of the fluxes, the objective function could for example be growth rate, ATP production, or glucose intake. The optimization of the objective function is subject to the mass balance constraints $S\boldsymbol{f} = \mathbf{0}$ and $\boldsymbol{l} \leq \boldsymbol{f} \leq \boldsymbol{u}$, where, $\boldsymbol{f} \in \mathcal{R}^n$ is the vector of n fluxes, $S \in \mathcal{R}^{m \times n}$ is a stoichiometric matrix, m is the number of reactants or metabolites in the network. All vectors by default will be column vectors. Also, $\boldsymbol{d}$, $\boldsymbol{l}$ and $\boldsymbol{u}$ are vectors $\in \mathcal{R}^n$ of objective function coefficients, lower and upper bound constraints on the fluxes respectively, and $\mathbf{0}$ is a zero vector of size m. The upper and lower bounds on the flux vector are taken componentwise. This constraint is got experimentally. The lower bound constraint in most cases is either zero or negative infinity. In this paper we will assume these two lower bounds for the fluxes. The upper bound constraint for most of the fluxes is infinity, but has a finite value that is experimentally determined for some boundary fluxes. Also, the vector of objective function coefficients has to be determined experimentally. Usually the objective function depends only on the boundary or exchange fluxes. For a reversible reaction the flux is unrestricted and can take on both positive and negative values, but for an irreversible reaction the flux is constrained to be non-negative. This is achieved by setting appropriate lower and upper bound constraints on the reaction fluxes.

3 Energy Balance Analysis: Second Law of Thermodynamics

According to the second law, fluxes must flow from reactants of higher chemical potential to ones of lower chemical potential. This is required by the second law of thermodynamics since the entropy of the reaction is always non-decreasing [6]. The FBA analysis is underconstrained, and gives rise to an infinite number of flux vectors. Many of these fluxes violate the second law and hence are infeasible. From a network topology point of view, it is the presence of cycles in the flux direction, that is Type III pathways, that violate the second law. Applying Kirchhoff's loop law, one gets rid of these entropy violating cycles.

From S we remove the columns corresponding to boundary fluxes and keep only the columns of non-redundant internal fluxes, which is defined as those between metabolites. The resulting matrix $G \in \mathcal{R}^{m \times n_i}$, where, n_i is the number of internal fluxes in the network. Using the reduced row echelon form [7] one can find the null space matrix N of G. The matrix $N \in \mathcal{R}^{n_i \times n_l}$ consists of n_l basis vectors of $\mathcal{N}(G)$, the null space of G. The dimension of $\mathcal{N}(G)$, denoted by $\mathcal{D}(\mathcal{N}(G))$ gives the number of independent loops n_l in the network [7]. By this we mean that a single basis loop cannot be decomposed into smaller loops. By taking linear combinations of these basis loops we can generate bigger and compound cycles (see Strang, 2003, page 363). Moreover, there are many loops which can not be decomposed into smaller ones (called elementary modes or extreme pathways) but they are all linearly dependent. This is why constructing a cycle space is computationally hard, while finding a set of basis vectors for the null space of a matrix is computationally trivial. This basis is unique since the reduced echelon form of G is unique, and the uniqueness is only true for a given form of the G matrix, but not for a given network.

Associated with each internal flux f_i is a chemical potential difference $\Delta\mu_i$. These potential differences satisfy a law similar to the Kirchhoff's loop law in electrical circuits,

namely [1]

$$K\boldsymbol{\Delta\mu} = \mathbf{0} \tag{1}$$

where, $K = N^T \in \mathcal{R}^{n_l \times n_i}$ is a matrix whose rows are the basis vectors of the null space of G, and $\boldsymbol{\Delta\mu} \in \mathcal{R}^{n_i}$ is a column vector of chemical potential differences for the internal fluxes in the cell, and $\mathbf{0}$ is a zero vector of size n_l.

The second law ensures that the entropy increases in each internal reaction i and hence the direction of internal flux f_i is from metabolites of higher chemical potential to one of lower chemical potential, and can be expressed as

$$\begin{cases} f_i \Delta\mu_i < 0 & \text{for } f_i \neq 0 \text{ and } \Delta\mu_i \neq 0, \\ f_i = 0, \Delta\mu_i = 0 & \text{otherwise.} \end{cases} \tag{2}$$

This is as nonlinear constraint which when incorporated into the FBA makes it a nonlinear programming problem. These inequalities are mathematical statements about heat dissipation in a chemical reaction, that must be positive. In non-equilibrium steady state, it is equivalent to the entropy production rate being non-negative [6]. Moreover, in non-equilibrium steady state, the chemical potential can be defined for each metabolite [6], that must satisfy equation (2). Not all flux vectors are consistent with this requirement (see figure 1 for example). Therefore the existence of the chemical potential imposes a constraint, via equation (2), on the flux vectors.

In addition to the above constraints one imposes upper and lower bound constraints on $\boldsymbol{\Delta\mu}$

$$\boldsymbol{\beta} \leq \boldsymbol{\Delta\mu} \leq \boldsymbol{\alpha} \tag{3}$$

where, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} \in \mathcal{R}^{n_i}$ represent the lower and upper bounds on the change in chemical potential $\boldsymbol{\Delta\mu}$, and the inequality is componentwise. The absolute values of the components in $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ mean nothing since equations (1), (2) and (3) can be scaled by a positive constant without changing the linear programming solution. In this model the role of the change in chemical potential is just to set the direction of the flux. Hence, it is the sign of $\Delta\mu_i$ that matters for the flux f_i. We need to refine the EBA model such that the absolute value of the chemical potential change is important. By studying several pathways it should then be possible to put more tight bounds on the fluxes and chemical potential changes for different reactions in the network.

Applying the second law of thermodynamics we restrict the space of metabolic fluxes even further by eliminating flux vectors that violate this law, since the corresponding reactions cannot occur in nature. In the EBA model we remove the boundary fluxes and apply the second law to the remaining internal fluxes in the metabolic network, which is now a closed system. Hence, the entropy of a closed system of reactions can only increase.

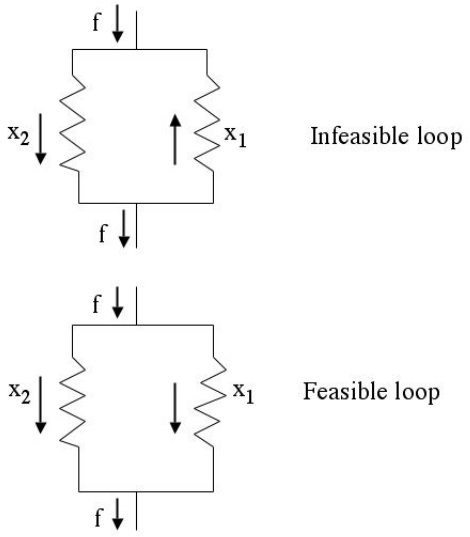

Figure 1. Electric circuit analogy for EBA. Thermodynamically infeasible and feasible loops. The current f is the sum of the currents x_1 and x_2. Currents in the opposite direction are infeasible whereas those in the same direction are feasible.

We introduce some notation and definition here that will be used in the following sections. We will use uppercase indices to denote sets for example, let F be the set of all fluxes in the network, R be the set of unrestricted fluxes, $F^{\geq 0}$ be the set of non-negative fluxes, $F^{<0}, F^{=0}$ and $F^{>0}$ be the set of negative, zero and positive fluxes respectively. Denote the ith flux component $f_i \in F$, r_i is an unrestricted flux, $f_i^{\geq 0}$ is a non-negative flux etc. The matrix N can be written in terms of its column vectors as $N = [N_{*1}, \ldots, N_{*k}, \ldots, N_{*i}, \ldots, N_{*n_l}]$, where N_{*k} is the kth column vector of N. Also $N_{*k} = [n_{1k}, n_{2k}, \ldots, n_{ik}, \ldots, n_{n_i k}]^T$, where n_{ik} is the (i,k)th entry of the matrix N. An internal flux x_i is called non-overlapping if it belongs to only one basis cycle.

4 No Cycle Feasibility Constraint

We introduce a simple test [4] to detect the presence of loops in a metabolic network that violate the second law of thermodynamics. To do so we take advantage of the directionality of the flow of fluxes in a cycle. In this paper we give a different proof of lemma 1 that is used in the test.

Lemma 1: (scaling lemma)
Transforming the internal flux ar_i, for any $a \in \{-1, 0, 1\}$ to r_i scales the ith column K_{*i} of matrix K by a, and hence

scales $\Delta\mu_i$ by a.

Proof: From the flux conservation equation $S\boldsymbol{f} = \mathbf{0}$, one can partition the flux vector $\boldsymbol{f}$ into an internal flux vector $\boldsymbol{x}$ and a boundary flux vector $\boldsymbol{y}$. The columns of the stoichiometric matrix S can likewise be partitioned into columns G corresponding to the internal fluxes and columns H corresponding to the boundary fluxes. That is, $S = [G\ H]$. The flux conservation equation can be rewritten as $G\boldsymbol{x} = -H\boldsymbol{y}$.
Since N is the null space matrix of G, we have $GN = 0$, where 0 is a $(m \times n_l)$ matrix of zeros. Consider the ith component of the unrestricted internal flux ar_i, which is a component of $\boldsymbol{x}$. In the matrix vector product $G\boldsymbol{x}$, ar_i multiplies the ith column G_{*i} of matrix G. The flux ar_i is transformed to r_i by transferring a to all the elements in the column of G_{*i}. By this process the value of the matrix vector product $G\boldsymbol{x}$ is unaffected. Hence G_{*i} becomes aG_{*i}. From the equation $GN = 0$ we have $GN = [G_{*1}, \ldots, G_{*i}, \ldots, G_{*n_i}]\,[N_{*1}, \ldots, N_{*i}, \ldots, N_{*n_l}]$, where G_{*1}, N_{*1} are the first columns of the G and N matrix respectively. This matrix-matrix product can be written compactly as $\left[\sum_{j=1}^{n_i} G_{*j}n_{j1}, \ldots, \sum_{j=1}^{n_i} G_{*j}n_{jn_l}\right]$. In this summation consider the terms involving $j = i$, from which it is clear that transforming G_{*i} to aG_{*i} transforms the ith row of matrix N, that is entries n_{i1},..., n_{in_l} get multiplied by a. Since by construction $K = N^T$, the ith row of matrix N corresponds to the ith column K_{*i} of matrix K, which then is multiplied by a. Equation (1) can be written as $\sum_{j=1}^{n_i} K_{*j}\Delta\mu_j = \mathbf{0}$, where if we consider the term $j = i$, it is clear that $\Delta\mu_i$ is scaled by a whenever K_{*i} gets multiplied by a to keep the equation invariant.

By use of scaling lemma 1 we transform all the unrestricted negative internal fluxes to positive fluxes by putting $a = -1$. Also if $a = 0$ then the ith column K_{*i} of matrix K becomes zero.

After application of lemma 1, if any row of K has the same sign, then the metabolic network is thermodynamically infeasible, and the cycle corresponding to that row violates the second law, and hence must be removed. This is only a necessary condition for infeasibility. It can be seen that the FBA problem with constraints (1) and (2) prevents the formation of such cycles. Therefore, the nonlinear constraint in equation (2) can be replaced by the sign test of lemma 1, which then makes the whole problem linear.

Lemma 2: (pivoting lemma)
Subtracting a multiple of the null space basis vector of the matrix G from the corresponding internal fluxes in a cycle does not change the optimal value of the objective function, but in some cases can impose additional constraints on the fluxes that make the EBA solution sub-optimal.

Proof: The columns of the stoichiometric matrix S can be partitioned into columns G corresponding to the internal fluxes $\boldsymbol{x}$ and columns H corresponding to the boundary fluxes $\boldsymbol{y}$. The objective function consists of only the boundary fluxes and hence can be written as $\boldsymbol{d}_y^T\boldsymbol{y}$, where $\boldsymbol{d}_y$ is a coefficient which in most cases has one non-zero component corresponding to the boundary flux one is trying to optimize. The flux conservation equation $S\boldsymbol{f} = \mathbf{0}$ can be written as $G\boldsymbol{x} + H\boldsymbol{y} = \mathbf{0}$ Let $\boldsymbol{b}_c$ be a basis vector in the null space of matrix G corresponding to cycle c. Subtracting a multiple γ of this vector from $\boldsymbol{x}$ leads to a new flux vector $\boldsymbol{x}' = \boldsymbol{x} - \gamma\boldsymbol{b}_c^T$ of internal fluxes, that still satisfies the flux conservation constraint. The vector $\boldsymbol{b}_c$ is the row $r(c)$ of the K matrix, $K_{r(c)*}$ corresponding to cycle c. We choose the multiplier γ such that the constraints on the internal fluxes are not violated.

Since this pivoting step does not change the objective function and the constraints on the LP, the application of this step will not change the optimal value of an objective function, such as production of biomass, that is composed of throughput fluxes. The examples discussed in [5] are very restrictive and their flux zeroing method does not always preserve the optimal value of the objective function, since they use $\boldsymbol{b}_c$ to be a vector of all ones, which may not always lie in $\mathcal{N}(G)$, the null space of G. Moreover, as a result of this pivoting step some internal fluxes become zero, and this results in the corresponding $\Delta\mu$'s to be zero. From equation (1), some *additional* $\Delta\mu$'s are inferred to be zero. These *additional* $\Delta\mu$'s enforce the corresponding internal fluxes to be zero, in order to maintain thermodynamic feasibility. It is these *additional* constraints on the fluxes that make the EBA solution sub-optimal. In this paper we use the pivoting lemma 2 to transform thermodynamically infeasible loops, to thermodynamically feasible loops. We pivot on a particular flux by making it either zero or changing its sign.

Definition: An internal flux x_i is limiting if by pivoting on it, the constraints on the other internal fluxes are not violated.

5 Algorithm

(i) Solve the FBA for the flux vector $\boldsymbol{f}$. If the FBA cannot find a solution then the problem has no solution.
(ii) Compute the K matrix from the reduced row echelon basis. For the fluxes in $F^{<0}$ and $F^{=0}$ apply the scaling lemma to scale the entries of the columns of the K matrix corresponding to these negative fluxes. Put the fluxes of

$F^{<0}$ in $F^{>0}$.
(iii) While the K matrix is infeasible, identify thermodynamically infeasible cycles, and apply the pivoting lemma to zero out the limiting flux or change its sign in that cycle. Update the sets $F^{<0}$, $F^{=0}$ and $F^{>0}$. (It is best to apply the pivoting lemma to non-overlapping limiting fluxes, if they are available). Update K matrix by use of scaling lemma 1. Repeat step (iii).
(iv) The components of the $\mathbf{\Delta\mu}$ vector corresponding to the zero fluxes in $F^{=0}$ are constrained to be zero. The $\mathbf{\Delta\mu}$ vector satisfying constraints (1), (2) and (3) is solved along with the LP. Due to this some more $\mathbf{\Delta\mu}$ vector components are inferred to be zero, so the corresponding internal fluxes are set to zero to preserve thermodynamic feasibility. Repeat step (iv) until no more additional $(f_i, \mathbf{\Delta\mu}_i)$ pair become zero. In this step the components of the $\mathbf{\Delta\mu}$ vector corresponding to the non-zero fluxes are constrained to be negative and the combined LP is solved. If the network is still thermodynamically infeasible then the reactions corresponding to the zero fluxes are deleted and step (iv) is repeated. (Reaction deletion alters the metabolic network. It makes the flux zero without requiring chemical potential difference to be zero.)
(v) If the LP in step (iv) is unable to find a solution, then report thermodynamic infeasibility.

The main complexity of the algorithm is the computation of the null space matrix K and the linear programming step. Both these run in polynomial time.

Lemma 3: (correctness lemma)
The EBA algorithm finds a thermodynamically feasible flux vector if the loop fluxes are permitted to be unrestricted or can take on zero values and the FBA solution can be found, otherwise it reports that the network is thermodynamically infeasible.

Proof: Existence of a FBA solution means that step (i) of the algorithm is successful. Operations in step (iii), namely the pivoting step can take place only if the limiting loop fluxes are permitted to be unrestricted or can take on zero values. By lemma 2 the pivoting step preserves the optimum value of the FBA problem, if no additional constraints on the fluxes are generated as a result of pivoting. However, one may be able to find a thermodynamically feasible flux vector with non-zero components that satisfy the sign test. In that case the pivoting procedure in step (iii) is not done. Also if step (iii) cannot be executed due to the non-zero constraint on the limiting internal fluxes, one could find other thermodynamically feasible flux vectors with non-zero components by starting the linear program at a different initial flux vector. For practical problems, the loop fluxes are permitted to have zero values, and the algorithm always proceeds by finding a series of optimal flux vectors by applying the pivoting step (iii) to remove all the thermodynamically infeasible loops in the network by zeroing out or changing sign of the limiting flux. The algorithm terminates with a thermodynamically feasible solution. In this case the original infeasible K matrix is transformed into a feasible one, that satisfies the sign test, by the repeated application of step (iii) of the algorithm finitely many times, till a certain number of fluxes have been transformed. If the network is still thermodynamically infeasible then in step (iv) the reactions corresponding to zero internal fluxes are deleted and this ensures that the algorithm finds a thermodynamically feasible solution. If after this the algorithm in step (v) is unable to find a solution, then the solution generated by the FBA is thermodynamically infeasible. The sign test introduced in [4] is a necessary condition for thermodynamic infeasibility.

Lemma 4: (basis lemma)
If the network is thermodynamically infeasible or feasible in one basis then it is thermodynamically infeasible or feasible in all basis.

Proof: Let $\mathcal{B}$ be a basis of the null space of G, and $K_{\mathcal{B}}$ be the matrix whose rows contain the basis vectors $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n_l})$. Consider another basis $\mathcal{C}$ with matrix $K_{\mathcal{C}}$. The basis vectors of $\mathcal{C}$ are linear combinations of basis vectors of $\mathcal{B}$. It is therefore possible to go back between the two basis since the linear operations can be undone. So the map is invertible. Hence,

$$K_{\mathcal{C}} = T_{\mathcal{C}\mathcal{B}} K_{\mathcal{B}} \tag{4}$$

where, $T_{\mathcal{C}\mathcal{B}}$ is an invertible transformation matrix. So the equations, $K_{\mathcal{B}}\mathbf{\Delta\mu}_{\mathcal{B}} = \mathbf{0}$ and $K_{\mathcal{C}}\mathbf{\Delta\mu}_{\mathcal{C}} = \mathbf{0}$ have the same set of solutions. Hence if $\mathbf{\Delta\mu}_{\mathcal{B}}$ is infeasible or feasible then so is $\mathbf{\Delta\mu}_{\mathcal{C}}$ and viceversa.

6 Application of the EBA Algorithm

6.1 Illustrative Example

In this section we consider the example discussed in [2]. We discuss feasibility and infeasibility of flux vectors using the simple machinery of linear algebra without resorting to matroid theory.

The set of reactions are shown below (see figure 2).
rxn 1 (x_1): A $--> $ B
rxn 2 (x_2): B $--> $ C
rxn 3 (x_3): C $--> $ A
rxn 4 (x_4): C $--> $ D
rxn 5 (x_5): D $--> $ B

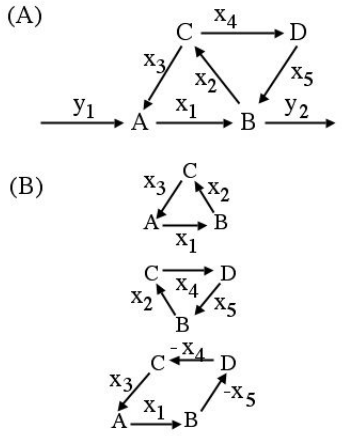

Figure 2. (A) Illustrative example network from [2] with five internal fluxes x_1, x_2, x_3, x_4, x_5 and two boundary fluxes y_1, y_2. (B) Three internal cycles corresponding to type III pathways for the network.

This reaction network has 5 internal fluxes corresponding to the 5 reactions, and associated with each internal flux is a change in chemical potential. In addition to these there are 2 boundary or exchange fluxes: y_1 transports A into the network and y_2 transports B out of the network. All reactions are reversible, but the arrows indicate the positive direction.

For this reaction network we wish to determine the maximum steady state production of reactant B, for a given maximal input flux of reactant A. This problem only assumes that A is the only available input substrate, and that its value is set to the maximum value.

The stoichiometric matrix G for the four metabolites A, B, C and D corresponding to the four rows respectively and the five reactions corresponding to the five columns of G respectively, is

$$G = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Also, the stoichiometric matrix H for the four metabolites as rows and the two exchange fluxes as columns.

$$H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The full stoichiometric matrix is $S = [G \;\; H]$ and the flux distribution $\boldsymbol{f}^T = [\boldsymbol{x}^T, \boldsymbol{y}^T] \in \mathcal{R}^7$ that satisfies the flux conservation constraint $S\boldsymbol{f} = \boldsymbol{0}$. The five internal fluxes corresponding to the five internal reactions are unrestricted, and are represented by the vector $\boldsymbol{x} \in \mathcal{R}^5$, while the two boundary fluxes are positive, and are represented by the vector $\boldsymbol{y} \in \mathcal{R}^2$.

The null space basis vectors from the reduced row echelon form of G are given in [2], and can be easily computed in MATLAB. They are,

$$\boldsymbol{k_1}^T = [1, 1, 1, 0, 0] \tag{5}$$

$$\boldsymbol{k_2}^T = [0, 1, 0, 1, 1] \tag{6}$$

The dimension of the null space of matrix G is 2, hence there are two basis cycles. The K matrix then consists of $\boldsymbol{k_1}^T$ and $\boldsymbol{k_2}^T$ as its rows, corresponding to cycle 1 and cycle 2, which are A $-->$ B $-->$ C $-->$ A and
B $-->$ C $-->$ D $-->$ B respectively (see figure 2). These two basis cycles can be combined into a single larger non-basis cycle 3: A $-->$ C $-->$ D $-->$ B $-->$ A with the corresponding null space vector $[-1, 0, -1, 1, 1]^T$. This is got by taking the difference between equations (6) and (5).

This representation tends to give a clear picture about shared and non-shared fluxes among cycles. Here flux x_2 is shared among the two cycles, whereas x_1 and x_3 belong to cycle 1, corresponding to $\boldsymbol{k_1}^T$, while, x_4 and x_5 belong to cycle 2, corresponding to $\boldsymbol{k_2}^T$.

$$K = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

The flux vector $\boldsymbol{f} = [\boldsymbol{x}^T, \boldsymbol{y}^T]^T \in R^7$ and change in chemical potential vector $\boldsymbol{\Delta\mu} \in R^5$ can be easily computed using MATLAB.

Consider $y_1 = 3$ units of A are provided as input to the network, and we want to maximize the transport flux y_2 of B. This leads to the following optimal flux and change in chemical potential vectors.

$$\boldsymbol{y} = [3, 3]^T \tag{7}$$

$$\boldsymbol{x} = [1, -1, -2, 1, 1]^T \tag{8}$$

$$\boldsymbol{\Delta\mu} = [-3, 2, 1, -1, -1]^T \tag{9}$$

Transforming the second and third columns of the K matrix corresponding to the negative flux components in $\boldsymbol{x}$, yields the transformed basis vectors $\bar{\boldsymbol{k_1}}^T =$

$[1,-1,-1,0,0]$ and $\bar{\boldsymbol{k_2}}^T = [0,-1,0,1,1]$. Since these vectors have components that are not of the same sign, the flux vector is thermodynamically feasible, and it also satisfies the flux balance constraints too. It is interesting to note that this flux vector has all non-zero components.

Now consider another internal flux vector $\boldsymbol{x} = [2,1,-1,2,2]^T$, that is optimal with respect to FBA with non-zero internal flux components that satisfies the mass balance constraint and produces 3 units of B for 3 units of A. Transforming the third column of K corresponding to the third negative component in $\boldsymbol{x}$ yields the transformed basis vectors $\bar{\boldsymbol{k_1}}^T = [1,1,-1,0,0]$ and $\bar{\boldsymbol{k_2}}^T = [0,1,0,1,1]$. Now we observe that all the components of $\bar{\boldsymbol{k_2}}^T$ have the same sign, hence the presence of cycle 2 makes this internal flux vector thermodynamically infeasible, even though it satisfies the flux balance constraints. To make this vector thermodynamically feasible we identify that x_2 is a limiting flux which can be made negative, without making the other fluxes zero or negative by subtracting $1.5 * \boldsymbol{k_1}^T$ from $\boldsymbol{x}$, to give $[0.5,-0.5,-2.5,2,2]^T$, which has a similar sign pattern as the internal flux vector in equation (8), and so is feasible, the $\boldsymbol{\Delta\mu}$ vector from equation (9) is the change in chemical potential that satisfies the nonlinear constaint of the second law in equation (2) for this transformed internal flux vector. By this transformation we retain the optimality of the FBA solution.

If the limiting flux x_2 is restricted, then we have to zero it out and so to make $\boldsymbol{x} = [2,1,-1,2,2]^T$ feasible we subtract $\boldsymbol{k_2}^T = [0,1,0,1,1]$ from $\boldsymbol{x}$, to yield $\bar{\boldsymbol{x}} = [2,0,-1,1,1]^T$, now $\bar{x}_2 = 0$. From the thermodynamic constraint in equation (2), we have $\Delta\mu_2 = 0$, this constraint gets rid of the second column of the K matrix using equation (1), and we transform the third column of the K matrix corresponding to the third negative flux component in $\bar{\boldsymbol{x}}$. The resulting K matrix indicates that cycle 2 is thermodynamically infeasible, as all the entries of the second row are of the same sign. Hence $\bar{\boldsymbol{x}}$ is thermodynamically infeasible. In this case we make the second component $\bar{x}_2 = 0$ and do not remove the reaction corresponding to this flux.

Our criterion for the EBA test is quite robust and we can immediately tell if the flux vector is thermodynamically infeasible, without carrying out any nonlinear optimization.

6.2 Applications of the EBA Algorithm: Analysis of *E. coli* Central Metabolism

We use the stoichiometric matrix S of the model *E. coli* system from Table 1 [3] for our FBA/EBA analysis. The reaction network contains 19 metabolites linked by 23 reactions (see figure 3). Out of these 23 fluxes there are 3 external or boundary fluxes and the rest 20 are internal fluxes. The network considered takes glucose as input and produces acetate and carbon dioxide. The energy and the metabolites involved in this process are used for the synthesis of proteins, DNA, RNA etc. We applied our algorithm to maximize the production of biomass flux, which is a linear combination of the different fluxes with experimentally determined stoichiometric coefficients. These coefficients are for the conversion of key metabolites to biomass. In the FBA optimization the internal fluxes are unrestricted, and only satisfy the flux balance constraint. The CO_2 and acetate fluxes come from the literature (references found in [3]) Since only the relative rates matter, the glucose flux is set to 1, and all other fluxes are normalized with respect to it.

The G matrix is formed by considering the columns of the following internal fluxes from Table 1 of [3]:
$\boldsymbol{x} = [J_{pgi}, J_3, J_{pep}, J_{pyk}, J_{pdh}, J_{ace}, J_8, J_{ict}, J_{11}, J_{12}, J_{ppc}, J_{14}, J_{15}, J_{16}, J_{tkt}, J_{tal}, J_{resp}, J_{atp}, J_{biomass}, J_{glyox}]$ and the H matrix is formed from the columns of the external fluxes $\boldsymbol{y} = [J_{gluc}, q_{CO_2}, q_{ace}]$.
The null space of the G matrix is of dimension 1, hence the K matrix consists of one row, corresponding to a single loop in the network. $K = [0,0,0,1,1,0,0,-1,-1,-1,-1,0,0,0,0,0,-1,-3,0,1]$. From the non-zero entries of the K matrix, we see the following 9 fluxes form a cycle: $[J_{pyk}, J_{pdh}, J_{ict}, J_{11}, J_{12}, J_{ppc}, J_{resp}, J_{atp}, J_{glyox}]$.

The optimal flux vectors and $\boldsymbol{\Delta\mu}$ satisfying the flux balance and thermodynamic constraints computed by our algorithm are rounded to two decimal places, and are:
$\boldsymbol{y} = [1, 2.2, 0.3]^T$
$\boldsymbol{x} = [0.87, 0.85, 1.58, 1.92, 2.71, 0.27, 0.56, -1.03, -1.11, -1.11, -1.38, 0.12, 0.09, 0.03, 0.03, 0.03, 1.80, 2.95, 0.0001, 1.59]^T$
$\boldsymbol{\Delta\mu} = [-8.25, -8.25, -8.25, -6.27, -6.27, -8.25, -8.25, 10.95, 10.95, 10.95, 10.95, -8.25, -8.25, -8.25, -8.25, -8.25, -10.74, -17.30, -8.25, -6.27]^T$

The optimized biomass flux is $J_{biomass} = 7.27 \times 10^{-5} \approx 0.0001$ per unit of glucose consumed. To see if the vector of internal fluxes $\boldsymbol{x}$, is thermodynamically feasible, we transform columns 8, 9, 10 and 11 of the K matrix corresponding to the negative flux components x_8, x_9, x_{10} and x_{11}, and see that the K matrix satisfies the feasibility criteria. We now compute the $\boldsymbol{\Delta\mu}$ vector by solving the combined linear program in step (iv) of the EBA algorithm, after imposing additional constraints that $\Delta\mu_8, \Delta\mu_9, \Delta\mu_{10}$ and $\Delta\mu_{11}$ must be positive and the other components of $\boldsymbol{\Delta\mu}$ are to be negative. This example illustrates that to satisfy the thermodynamic constraints one does not always have to zero out internal fluxes, which are part of a loop.

Now consider the following optimal internal flux vector that satisfies FBA, but is thermodynamically infeasible.
$\boldsymbol{x} = [0.87, 0.85, 1.58, 48.51, 49.31, 0.27, 0.56, -47.63, -47.71, -47.71, -47.98, 0.12, 0.09, 0.03, 0.03, 0.03,$

$-44.80, -136.84, 0.0001, 48.19]^T$
To see that it is thermodynamically infeasible, we transform columns 8, 9, 10, 11, 17 and 18 of the K matrix corresponding to the negative flux components to get a transformed K matrix $[0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 3, 0, 1]$, and see that all components have the same sign, so it is infeasible. To convert it to a feasible flux vector we identify x_{20} as a limiting flux that we want to make negative by subtracting $48.25 * K^T$ from $\boldsymbol{x}$ to give
$[0.87, 0.85, 1.58, 0.26, 1.06, 0.27, 0.56, 0.62, 0.55, 0.55, 0.27, 0.12, 0.09, 0.03, 0.03, 0.03, 3.45, 7.91, 0.0001, -0.0654]^T$
We see that this is feasible after changing the sign of the 20th column in the K matrix, the non-zero components of the K matrix have different signs, so we calculate $\boldsymbol{\Delta\mu}$ satisfying the second law of thermodynamics by constraining all its components except the last one to be negative, and is
$[-80.93, -80.93, -80.93, -138.85, -138.85, -80.93, -80.93, -45.83, -45.83, -45.83, -45.83, -80.93, -80.93, -80.93, -80.93, -80.93, -45.83, -0.10, -80.93, 48.26]^T$

There is another way of transforming the infeasible $\boldsymbol{x}$ to be thermodynamically feasible, by making the limiting flux $x_{20} = 0$, by subtracting $x_{20} * K^T$ from $\boldsymbol{x}$ to get $\bar{\boldsymbol{x}} = [0.87, 0.85, 1.58, 0.33, 1.12, 0.27, 0.56, 0.48, 0.48, 0.21, 0.12, 0.09, 0.03, 3.38, 7.71, 0.0001, 0]^T$
Since $x_{20} = 0$, we can delete the 20th column of the K matrix, and we see that this transformed internal flux vector is thermodynamically feasible since the other components of the K matrix have different signs, and we calculate the $\boldsymbol{\Delta\mu}$ vector by solving the combined linear program after imposing the constraints that all its components are negative, since all the flux components in $\bar{\boldsymbol{x}}$ are positive, in order to satisfy equation (2). The feasible $\boldsymbol{\Delta\mu}$ vector is
$\bar{\boldsymbol{\Delta\mu}} = [-80.23, -80.23, -80.23, -121.70, -121.70, -80.23, -80.23, -47.65, -47.65, -47.65, -47.65, -80.23, -80.23, -80.23, -80.23, -80.23, -47.65, -1.73, -80.23, 0]^T$

We applied our algorithm to the full network of *E. coli* and found several cycles that violated the second law of thermodynamics.

7 Conclusion

In this paper we give a simple linear programming algorithm for flux and energy balance analysis. It uses the sign of the null space to decide if the flux vector computed by flux balance analysis satisfies the second law of thermodynamics. This technique is different from the previous approaches, as it is constructive, and can generate several solutions for the metabolic network. The method terminates when it finds a thermodynamically feasible solution. We applied the method to a part of the metabolic

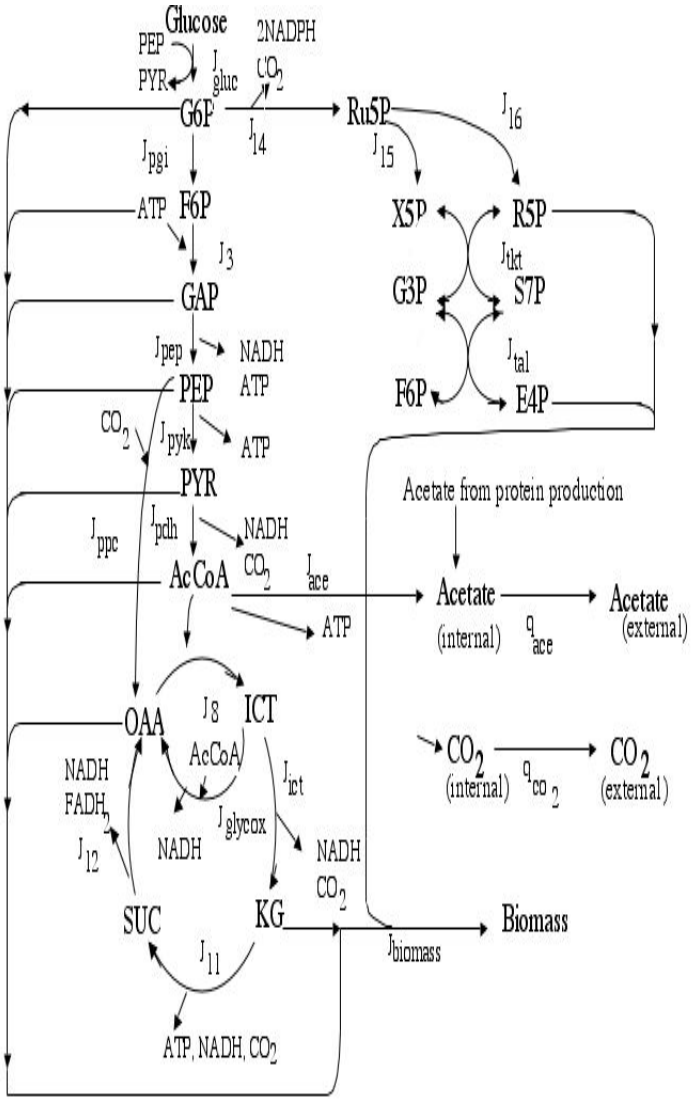

Figure 3. *E. coli* **central metabolism**

network of *E. coli* and computed the fluxes and change in chemical potentials for the internal reactions. It should however be noted that inspite of the additional thermodynamic constraint, the metabolic network is still degenerate and has an infinity of flux and chemical potential difference vectors that satisfy all the constraints. Hence, more realistic bounds on the values of fluxes are required to further constrain the system.

Acknowledgments
The first author would like to thank Prof. Richard Karp for helpful discussions, during visits to the International Computer Science Institute, Berkeley. We also thank Prof. Robert Laughlin of the department of Physics at Stanford for his support on the project. We thank the referees for their comments. This work was supported by a joint NASA-NCI biotechnology research program.

References

[1] Beard, D. A., Liang, S. and Qian, H., "Energy Balance for Analysis of Complex Metabolic Networks", *Biophys. J.* (2002) **83**, 79-83.

[2] Beard, D. A., Babson, E., Curtis, E. and Qian, H., "Thermodynamic Constraints for Biochemical Networks", *J. Theor. Biol.* (2004) **228**, 327-333.

[3] Delgado, J. and Liao, J. C., "Inverse Flux Analysis for Reduction of Acetate Excretion in *Escherichia coli*", *Biotechnol. Prog.* (1997) **13**, 361-367.

[4] Nigam, R. and Liang, S., "Thermodynamic Feasibility of Metabolic Networks", International Conference on Bioinformatics and its Applications, Fort Lauderdale, USA, in press (2004).

[5] Price, N. D., Famili, I., Beard, D. A. and Palsson, B. O., "Extreme Pathways and Kirchhoff's Second Law", *Biophys. J.* (2002) **83**, 2879-2882.

[6] Qian, H., Beard, D. and Liang, S., "Stoichiometric Network Theory for Nonequilibrium Biochemical Systems", *Eur. J. Biochem.* (2003) **270**, 415-421.

[7] Strang, G., "Introduction to Linear Algebra" *Wellesley-Cambridge Press*, Wellesley, MA (2003).

[8] Varma, A., and Palsson, B. O., "Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild Type Escherichia coli W3110", *Appl. Environ. Microbiol.* (1994) **60**, 3724-3731.