Testing Feasibility in Metabolic Pathways is as Easy as Matrix Multiplication

R. Nigam
Department of Physics,
Stanford University, Stanford, CA 94305-4085, USA

Abstract

A new algorithm to test thermodynamic infeasibility of reaction fluxes in metabolic pathways is presented. Once a flux has been computed the only computational step involves the multiplication of two matrices. This will be useful in large metabolic networks which give rise to multiple fluxes that need to be tested for feasibility. We applied this test to the *E. coli* metabolic network.

Keywords: Metabolic network; Thermodynamics; Energy balance analysis

1 Introduction

Study of metabolic networks is important in predicting the cellular behavior as a result of interactions between the different reactions occurring within the cell (3,6). For example the flux balance analysis (FBA) has proved to be useful for studying the steady state kinetic rates (or fluxes) inside the cell (8). This approach is especially a valuable tool in the absence of knowledge of detailed kinetic parameters of reactions inside the cell. Recently there have been attempts to further constrain the fluxes that are obtained by FBA, to satisfy the second law of thermodynamics by a method called energy balance analysis (EBA) (1). The EBA eliminates thermodynamically infeasible fluxes that are got by applying FBA alone (4).

In this paper we incorporate the energy feasibility constraint that is akin to Kirchhoff's potential law (7) into the linear FBA theory. This places additional constraints on the FBA solution space and eliminates thermodynamically infeasible fluxes that do not satisfy the loop law. The algorithm presented here is a necessary condition for thermodynamic feasibility, and it can be applied in the context of multiple fluxes. A single computation that involves the multiplication of two matrices, is necessary to check the thermodynamic infeasibility of multiple fluxes in a metabolic network. This model will be useful in predicting the behavior of large scale networks, especially in making predictions of gene regulation and thermodynamic chemical potentials of the different chemical reactions that go on inside the cell.

2 Flux and Energy Balance Analysis

The theory of FBA invokes the law of flux conservation, and has been formulated as a linear programming problem (8), where there is a linear objective function to be optimized, which could for example be to maximize growth, to maximize ATP production, or to minimize glucose intake. The theory maximizes an objective function that is a linear combination of the fluxes.

$$\max_{\boldsymbol{f}\in Z_f} \boldsymbol{d}^T\boldsymbol{f} \tag{1}$$

where, $Z_f = \{\, \boldsymbol{f} \in \mathcal{R}^n \mid S\boldsymbol{f} = \boldsymbol{0}, \boldsymbol{l} \le \boldsymbol{f} \le \boldsymbol{u} \,\}$ is a set of n-component flux vectors satisfying FBA constraints of mass conservation, $S\boldsymbol{f} = \boldsymbol{0}$, $S \in \mathcal{R}^{m\times n}$ is a stoichiometric matrix, m is the number of reactants

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 inequalities in the set Z_f are componentwise for the vectors. For example, for a substrate uptake flux f_j, one can set l_j and u_j equal to the corresponding experimentally measured value. Also these lower and upper bound constraints can be used to distinguish reversible and irreversible reactions, where $l_j > 0$ for the latter. The vector of objective function coefficients have to be determined experimentally. For a realistic stoichiometric matrix, the number of fluxes n exceeds the number of metabolites m. So a convenient way to solve the above system of underdetermined equations is to resort to linear programming.

Energy balance analysis on the other hand invokes the second law of thermodynamics, as a consequence of which, reaction fluxes are always in the direction of decreasing chemical potential (5). This is the direction in which the entropy of the reaction increases. Since the FBA analysis gives rise to an infinite number of fluxes, many of these flux distributions violate the second law and hence are infeasible. Hence, an algorithm is developed to test for thermodynamic infeasibility.

One can partition the columns of the stoichiometric matrix S into non-redundant columns $G \in \mathcal{R}^{m \times n_i}$ corresponding to the internal flux vector $\boldsymbol{x} \in \mathcal{R}^{n_i}$ and columns H corresponding to the boundary flux vector $\boldsymbol{y}$. The internal fluxes are between two metabolites both of which are represented as rows in the stoichiometric matrix S. From the reduced row echelon form (7) of G, one can find the null space matrix N of G, $GN = 0$. The matrix $N \in \mathcal{R}^{n_i \times n_l}$ consists of n_l basis vectors of the null space of G, corresponding to the number of simple loops n_l in the network. Associated with each internal flux x_i (ith component of $\boldsymbol{x}$), for reaction i, there is a chemical potential difference $\Delta\mu_i$. These potential differences belong to the set Z_e, satisfying a law similar to the Kirchhoff's loop law in electrical circuits (1), namely
$Z_e = \{\, \boldsymbol{\Delta\mu} \in \mathcal{R}^{n_i} \mid K\boldsymbol{\Delta\mu} = \mathbf{0}, \boldsymbol{\beta} \leq \boldsymbol{\Delta\mu} \leq \boldsymbol{\alpha} \,\}$. Where, $K = N^T$, $\mathbf{0}$ is a zero vector of size n_l, $\boldsymbol{\beta}$ 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 the constraints in the sets Z_f, Z_e and equation (2) can be scaled by a positive constant without changing the linear programming solution.

In addition, for $\boldsymbol{x}$ and $\boldsymbol{\Delta\mu}$ to be thermodynamically feasible, they must satisfy the following nonlinear constraint:

$$\begin{cases} x_i \Delta\mu_i < 0 & \text{for } x_i \neq 0 \text{ and } \Delta\mu_i \neq 0, \\ x_i = 0, \Delta\mu_i = 0 & \text{otherwise.} \end{cases} \quad (2)$$

3 Thermodynamic feasibility constraint

The number of rows n_l of the K matrix gives the number of loops or cycles in the network (Strang, 2003). If in any row j of the K matrix, $K_{j*} = [k_{j1}, \ldots, k_{jn_i}]$ (here, k_{jn_i} is the (j, n_i)th entry of matrix K) all the entries (there should be more than one non-zero entry) are of the same sign, corresponding to the set of positive fluxes for the jth cycle, then the flux distribution is thermodynamically infeasible, otherwise the flux distribution is cycle feasible.

It can be seen that to satisfy $K\boldsymbol{\Delta\mu} = \mathbf{0}$, for a single row of the K matrix corresponding to a single cycle, atleast one of the $\Delta\mu_i$ should be of a different sign from the rest of the other components in the $\boldsymbol{\Delta\mu}$ vector. This when combined with equation (2), prevents the formation of loops along the flux direction in the network, and hence satisfies the second law of thermodynamics. This idea will be used in the infeasibility testing algorithm presented in the next section. The input to the algorithm is the matrix of internal fluxes X that satisfy the FBA constraints, and the metabolic pathway, which is specified by the sgn(K) matrix. The sgn(x) function returns 1 for positive x, -1 for negative x and 0 for $x = 0$. (For the matrix K, the sgn function is taken elementwise). The internal flux vectors for which thermodynamic infeasibility is tested are columns of the matrix X. Two matrices that keep track of the zeros in each cycle of the K matrix and the zero fluxes in each flux vector of the X matrix are maintained. This is a bookkeeping step

which tells us the fluxes that are absent from a cycle. From this information the number of non-zero fluxes in a cycle can be computed. This gives rise to a matrix I, whose rows correspond to the cycles in the network. The jth column of I, I_{*j} corresponds to the jth flux vector, which is the jth column of matrix X. The column vector I_{*j} stores information on the number of non-zero fluxes in each cycle. For example, the ith flux component f_i is zero in the jth cycle , if $k_{ji}f_i = 0$, and is non-zero otherwise.

4 Feasibility Algorithm

i) Compute $J = \text{sgn}(K)\text{sgn}(X)$, $L = J \bmod I$. Here, the mod operation is done componentwise.
ii) If the (i, j)th element of the matrix I, $I_{ij} = 1$, then the jth flux vector in the ith cycle is cycle feasible.
iii) If the (i, j)th element of the matrix L, $L_{ij} \neq 0$ and $I_{ij} \neq 1$, then the jth flux vector corresponding to the jth column of the X matrix, X_{*j} is cycle feasible for the ith cycle of the metabolic network corresponding to the i row of the matrix K, K_{i*}.
iv) If $J_{ij} = 0$ then cycle feasible, else if $J_{ij} \neq 0$ and $L_{ij} = 0$ then thermodynamically infeasible.
v) Partition the columns of L into cycle feasible and thermodynamically infeasible fluxes. Since this is a necessary condition for thermodynamic feasibility, the cycle feasible fluxes have to be further investigated to see if they are thermodynamically feasible or infeasible.

Two lemmas are proved regarding the correctness and complexity of the algorithm.

Lemma 1: (correctness lemma)
The algorithm correctly tests a flux vector for thermodynamic infeasibility.
Proof: In step (ii), $I_{ij} = 1$ means that the jth flux vector has one non-zero flux in the ith cycle, hence it is cycle feasible, by setting the corresponding change in chemical potential to zero. In step (iii), $L_{ij} \neq 0$ iff $J_{ij} \neq I_{ij}$ or $-I_{ij}$ or $J_{ij} = 0$, hence the ith row of K has entries with different signs, hence it is cycle feasible for the jth flux vector. In step (iv), $J_{ij} = 0$ iff the ith row of K has entries with different signs which cancel each other, and hence is cycle feasible. But, $L_{ij} = 0$ iff $J_{ij} = I_{ij}$ or $-I_{ij}$, hence the ith row of K have all positive or negative sign entries and hence is thermodynamically infeasible for the jth flux vector.

Lemma 2: (complexity lemma) The algorithm has polynomial complexity.
Proof: Matrix multiplication is polynomial.

The output of the algorithm is the L matrix. Its entries tell, which cycles are thermodynamically infeasible, as a result of some fluxes. These fluxes are then said to be thermodynamically infeasible. The remaining fluxes are cycle feasible that have to be tested for thermodynamic feasibility or infeasibility, since this algorithm is only a necessary condition for testing thermodynamic feasibility. The algorithm can be used as a building block to larger programs for metabolic networks.

5 Application to *E. coli*

We use the stoichiometric matrix S of the model *E. coli* system from Table 1 of (3) for our FBA/EBA analysis. The reaction network contains 19 metabolites linked by 23 reactions (Figure 1 in (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 C02 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 Delgado and Liao (1997):
$\boldsymbol{x} = [J_{pgi}, J_3, J_{pep}, J_{pyk}, J_{pdh}, J_{ace},$
$J_8, J_{ict}, J_{11}, J_{12}, J_{ppc},$
$J_{14}, J_{15}, J_{16}, J_{tkt}, J_{tal}, J_{resp}, J_{atp}, J_{biomass}, J_{glyox}]$
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: $[J_{pyk}, J_{pdh}, J_{ict}, J_{11}, J_{12}, J_{ppc},$
$J_{resp}, J_{atp}, J_{glyox}]$.

Applying the algorithm one finds that the internal flux vector $\boldsymbol{x} = [0.87, 0.85, 1.58, 1.92, 2.71, 0.27, 0.56, -1.03, -1.11,$
$-1.11, -1.38, 0.12, 0.09, 0.03, 0.03, 0.03, 1.80, 2.95,$
$0.0001, 1.59]^T$ is cycle feasible, infact it is is thermodynamically feasible, but the vector $\boldsymbol{x} = [0.87, 0.85, 1.58, 48.51, 49.31, 0.27, 0.56, -47.63,$
$-47.71, -47.71, -47.98, 0.12, 0.09, 0.03, 0.03, 0.03,$
$-44.80, -136.84, 0.0001, 48.19]^T$ is thermodynamically infeasible.

6 Conclusions

In this paper we give a simple algorithm to test the thermodynamic infeasibility of metabolic pathways. It can test the infeasibility of many flux distributions at once, and hence can be incorporated into algorithms for computing fluxes in large metabolic pathways. This algorithm has been tested on the *E. coli* pathway.

7 Acknowledgments

I would like to thank professors Richard Karp, Robert Laughlin and Shoudan Liang for helpful discussions and their support on the project. This research was supported by a NASA cooperative agreement NCC2-5475, between NASA Ames and Stanford University.

8 References

[1] D. A. Beard, S. Liang and H. Qian, "Energy balance for analysis of complex metabolic networks," *Biophys. J.* **83**: pp. 79-86, 2002.
[2] D. A. Beard, E. Babson, E. Curtis and H. Qian, "Thermodynamic Constraints for Biochemical Networks," *J. Theor. Biol.* **228**: pp. 327-333, 2004.
[3] J. Delgado and J. C. Liao, "Inverse Flux Analysis for Reduction of Acetate Excretion in *Escherichia coli*," *Biotechnol. Prog.* **13**: pp. 361-367, 1997.
[4] N. D. Price, I. Famili, D.A. Beard and B.O. Palsson, "Extreme Pathways and Kirchhoff's second law," *Biophys. J.* **83**: pp. 2879-2882, 2002.
[5] H. Qian, D. A. Beard and S. Liang, "Stoichiometric network theory for nonequilibrium biochemical systems," *Eur. J. Biochem.* **270**: pp. 415-421, 2002.
[6] G. Stephanopoulos, A. Aristidou and J. Nielsen, "Metabolic Engineering: Principles and Applications," 1998.
[7] G. Strang, "Introduction to Linear Algebra," Wellesley-Cambridge Press, Wellesley, MA, 2003.
[8] A. Varma and B. O. Palsson, "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110," *Appl. Environ. Microbiol.* **60**: pp. 3724-3731, 1994.