Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Metabolic Networks Analysis using Convex Optimization

A. Agung Julius, Marcin Imielinski and George J. Pappas

Abstract— **Metabolic networks map the biochemical reactions in a living cell to the flow of various chemical substances in the cell, which are called metabolites. A standard model of a metabolic network is given as a linear map from the reaction rates to the change in metabolites concentrations. We study two problems related to the analysis of metabolic networks, the minimal network problem and the minimal knockout problem.**

The minimal network problem amounts to finding the smallest set of reactions that can sustain the production of a metabolite. The minimal knockout problem deals with the question of finding the smallest set of knockouts (reactions with zero rates) that renders the production of a metabolite infeasible.

In this paper we present a convex relaxation technique that results in a very fast computation for the solution to both problems. We also demonstrate that the minimal knockout problem is related to the dual of the minimal network problem.

I. Introduction

Biology is undergoing a paradigmatic shift that brings about a quantitative and analytic facet to the field which was primarily qualitative. The approach to biology that highlights the use of quantitative models and reasoning based on systems and control theory leads to the field of *systems biology*. There are many problems in systems biology that are essentially engineering problems, and require engineering mindset to solve.

Experimental data from cellular and molecular biology suggest that entities in cellular systems influence one another and can be thought of as forming a vast and complex network. Network structures in systems biology appear in many levels, for example, genetic regulatory network, protein-protein interaction, and metabolic networks.

Gene regulatory networks capture interactions between genes and other cell substances, resulting in various models for the fundamental biological process of transcription and translation. In many cases, the product of the genes are enzymes that facilitate various biochemical reactions in the cell.

Metabolic networks map the biochemical reactions in a living cell to the flow of various chemical substances in the cell, which are called metabolites. The metabolic network of an organism can be thought of as production lines in a large scale biochemical plant. They capture the metabolic reactions in which metabolic products are made, and the reactants that are involved. Analysis of a metabolic network is important when we want to engineer the organism to function as a biofactory [1]. This is done by altering the metabolic network (by means of reaction knockout) so as to make it produce certain products and/or not to produce certain other products. Reaction knockouts typically correspond to disabling the genes that produce the necessary enzyme for the reactions involved.

Genome scale metabolic network models enable *in silico* knockout experiment design for the purposes of metabolic engineering, drug discovery, and improving the systems-level understanding of metabolism. Current approaches to the study of such networks employ an analysis of feasible and optimal reaction fluxes through the network at steady state, subject to structural, thermodynamic, and flux capacity constraints [2], [3]. Structural constraints arise from the stoichiometry of the metabolic reactions, thermodynamic constraints are imposed by the irreversibility of certain metabolic reactions, while flux capacity constraints can be derived from the availability of nutrients, the existence of a knockout, and biochemical data on the maximum throughput of enzymes. Finally, the steady state assumption follows from time-scale separation between rapid metabolic reactions and slower environmental and cellular regulatory changes. Given such constraints, the flux configuration through the network is limited to a feasible region, which can be checked for non-emptiness to characterize the production capacity of the network [4], [5], can be analyzed to find points that maximize biomass production [6], [7], [2] or minimize metabolic adjustment [8], [9].

Metabolic networks are characteristic in their high degree of robustness to single reaction knockout [10], [11], [12], [13]. Though this limits the number of simple targets for metabolic engineering or pharmaceutical intervention, it allows for the possibility that multi-pronged perturbations may be effective in compromising the inherent redundancy of this cellular system. Currently, the primary method for genome-scale *in silico* knockout design is flux balance analysis, which uses linear programming (LP) to exhaustively test the metabolic capabilities of all single, double, triple, etc. knockout combinations. This "brute force" linear programming approach is computationally limited to the testing of small (less than 3-5 reactions) knockout combinations [13]. The minimal cut set (MCS) algorithm of Klamt and Gilles is an "rational" network-based approach for knockout experiment design; however, it is intractable for large (i.e. genome-scale) metabolic networks due to the computational complexity of elementary mode computation [14], [15]. Imielinski and Belta have recently introduced $NetKO$, a relaxed implementation of Klamt and Gilles' method that scales to arbitrarily large metabolic networks; however, this method does not guarantee the discovery of a knockout combination with a given desired property [16]. Finally, a related work employs mixed integer linear programming (MILP) to uncover minimal knockout combinations that

Agung Julius and George Pappas are with the Dept. Electrical and Systems Engineering, University of Pennsylvania, 3330 Walnut Street, Philadelphia, PA 19104, U.S.A. agung,pappasg@seas.upenn.edu

Marcin Imielinski is with the Center for Applied Genomics, Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Philadelphia, PA, 19104, U.S.A. imielins@mail.med.upenn.edu

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

achieve a certain network objective [17], [18], [19]. This approach is limited primarily by the computational difficulty associated with solving high-dimensional MILPs.

In this paper, we introduce a fast convex-programming approach for analyzing genome-scale metabolic networks. Specifically, our method solves either of two following problems: 1) minimal network problem: finding minimal sets of reactions that can sustain a desired function, 2) minimal knockout problem: finding minimal sets of reactions that can be disabled to abolish an undesired function. Like MILP-based approaches, we formulate the task of knockout generation using an integer-valued objective function and integer-valued constraints. However, rather than solving the resulting MILP, we employ a convex relaxation of the problem to generate an approximate solution that we subsequently verify and refine via LP. As we show below, our method provides an efficient and scalable approach for knockout computation in genome-scale metabolic networks.

II. Mathematical model

Consider a metabolic network with n metabolites and n_r reactions. The k-th reaction can be written as

$$\alpha_{1,k}A_1 + \ldots + \alpha_{n,k}A_n \rightarrow \beta_{1,k}A_1 + \ldots + \beta_{n,k}A_n, \quad (1)$$

where A_i denotes the i-th metabolite, and $\alpha_{\bullet,k}$ and $\beta_{\bullet,k}$ are nonnegative integers that denote the stoichiometric coefficients of the k-th reaction. Obviously, if A_i is not involved as a reactant in the k-th reaction, $\alpha_{i,k} = 0$. Similarly, if A_i is not involved as a product in the k-th reaction, $\beta_{i,k} = 0$. In **regular reactions**, we have

$$\alpha_{\bullet,k} \neq 0^n, \quad (2a)$$

$$\beta_{\bullet,k} \neq 0^n, \quad (2b)$$

which means that there is always some reactant and product associated with the reaction. Here we assume that all reactions are irreversible. Notice that that this is done without any loss of generality since reversible reactions can be written as two opposite irreversible reactions.

In addition to the regular reactions, we also have **uptake reactions**. These are reactions that can be written as

$$* \rightarrow A_i, \quad (3)$$

which models the uptake of a metabolite A_i from the environment. Notice that uptake reactions can also be expressed as in (1), without the restriction of (2a).

If we denote the concentration of the i-th metabolite as x_i and the rate of the k-th reaction as ω_k, then it is easy to show that x and ω are related through

$$\frac{dx}{dt} = (\beta - \alpha)\omega, \omega \succeq 0, \quad (4)$$

where β and α are $n \times n_r$ matrices formed by the coefficients of (1). The symbol $\succeq$ denotes elementwise inequality.

In microbes, the transient dynamics of the metabolic network is faster than both cellular growth rates and the dynamic changes in the organism's environment [20]. In analysing the network, thus, it is assumed that it is in its steady state. In steady state condition, the rate $\frac{dx}{dt}$ must be elementwise nonnegative. This is because the cell can act as a perpetual sink, but not as a perpetual source (without any uptake). Thus, in steady state condition the following relations hold.

$$(\beta - \alpha)\omega - \frac{dx}{dt} = 0, \quad (5a)$$

$$\omega \succeq 0, \frac{dx}{dt} \succeq 0. \quad (5b)$$

We can write (5) in a more compact manner by introducing **pseudoreactions** as sinks. These are reactions that can be written as

$$A_i \rightarrow *. \quad (6)$$

We associate a sink for every metabolite. Thus, there are n pseudoreactions. Equation (5) can therefore be written compactly as

$$Sv = 0, v \succeq 0, \quad (7)$$

where

$$S := \begin{bmatrix} \beta - \alpha & -I \end{bmatrix} \in \mathbb{Z}_+^{n \times m}, v := \begin{bmatrix} \omega \\ \frac{dx}{dt} \end{bmatrix} \in \mathbb{R}^m, \quad (8)$$

with $m := n + n_r$. Such a model for metabolic networks constitutes an analysis method known as the **flux balance analysis** [20], [21].

III. Convex optimization in metabolic network analysis

A. Minimal network problem

In this paper, we are interested in two types of problems in metabolic network analysis. The first problem is called the *minimal network problem*. Mathematically, this problem is related to finding the minimal set of reactions that can sustain growth under certain environmental nutrient conditions.

Definition 3.1 (Minimal metabolic network problem): Given a model of a metabolic network as in (7), a set of target products P, and a set of uptake reactions U, determine the smallest set of reactions (regular and uptake) that can sustain the production of P without involving any reaction in U.

The set U defines the nutrient metabolites that are absent in the environment. In this problem, thus, we are interested in finding out the smallest set of reactions that can lead to the production of products in P without using any U uptake. An answer to this question leads to knowledge about the *minimum genome* that is required to sustain growth and replication of the organisms [22].

It is well known that set inclusion defines a partial ordering. Thus, generally there is not a unique solution for such a problem. We are interested in enumerating the solutions. The problem can be cast as an optimization problem as follows.

$$\begin{aligned} &\text{minimize } \text{card}(v), \\ &\text{subject to } Sv = 0, v \succeq 0, \\ &\qquad v_{i \in P} \succ 0, v_{i \in U} = 0. \end{aligned} \quad (9)$$

The symbol $\text{card}(v)$ denotes the cardinality of the vector v, which is the number of nonzero entries of v. The same operation is sometime called the ℓ_0 norm of a vector, denoted by $\|\cdot\|_0$.

The optimization problem defined in (9) involves a non-convex objective function and convex constraints. This type of problem is known as *convex cardinality problem [23]*. Finding an exact solution to this kind of problems involves mixed integer linear programming (MILP) [22], [24], and it is known to have NP-Hard complexity.

There exist convex relaxation methods for this type of problems [23]. We have succesfully applied the relaxation technique in another systems biology problem, the identification of sparse biomolecular networks [25], [26]. Similar program has been carried out independently by Papachristodoulou and Recht [27] and Han *et al* [28].

The algorithm for the relaxation of (9) is given as an iteration as follows:

Algorithm 1 Convex relaxation for minimal network problem

Require: A stoichiometric matrix S, a set of target products P, and a set of uptake reactions U

1: Initialize the weight $w^{(1)} = [1, \ldots, 1] \in \mathbb{R}^m$, and the counter $k = 1$.

2: **repeat**

3: Solve the Linear Programming problem for some small number $\varepsilon > 0$:

$$\begin{aligned} \text{minimize } & \sum_{i=1}^{m} w_i^{(k)} v_i^{(k)}, \\ \text{subject to } & Sv^{(k)} = 0, v^{(k)} \succeq 0, \\ & v_{i\in P}^{(k)} \succ \varepsilon, v_{i\in U}^{(k)} = 0, \end{aligned} \tag{10}$$

to find the flow for the $k-$th iteration, $v^{(k)}$.

4: Update weight according to

$$w_i^{(k+1)} = \phi(v_i^{(k)}), \tag{11}$$

where

$$\phi(x) := \frac{\delta^n}{\delta^n + x^n}, \tag{12}$$

5: **until** $\left\|v^{(k)} - v^{(k-1)}\right\| \leq \delta$, where $\delta > 0$ is a small number that indicates the convergence of the iteration.

The update function $\phi(\cdot)$ in (12) is parameterized by δ and n. See Figure 1 for different shapes of the function for different parameters.

The intuition behind the algorithm can be explained as follows. We replace the ℓ_0 norm objective function with a weighted ℓ_1 objective function. The weight w is adjusted at every iteration so that elements of the flow vector v that are small are given more weight than large elements (see Figure 1). Thus, we put more emphasis on driving small elements to zero.

Example 3.2: Consider a small metabolic network with 8 metabolites as follows.

$$\begin{aligned} &\mathrm{A} + \mathrm{B} \xrightarrow{(1)} \mathrm{E}, \mathrm{A} + \mathrm{C} \xrightarrow{(2)} \mathrm{F}, \mathrm{C} + \mathrm{D} \xrightarrow{(3)} \mathrm{G}, \\ &\mathrm{E} + \mathrm{A} \xrightarrow{(4)} \mathrm{H}, \mathrm{F} + \mathrm{C} \xrightarrow{(5)} \mathrm{H}, \mathrm{G} + \mathrm{C} \xrightarrow{(6)} \mathrm{H}, \\ &* \xrightarrow{(7)} \mathrm{A}, * \xrightarrow{(8)} \mathrm{B}, * \xrightarrow{(9)} \mathrm{C}, * \xrightarrow{(10)} \mathrm{D}. \end{aligned}$$

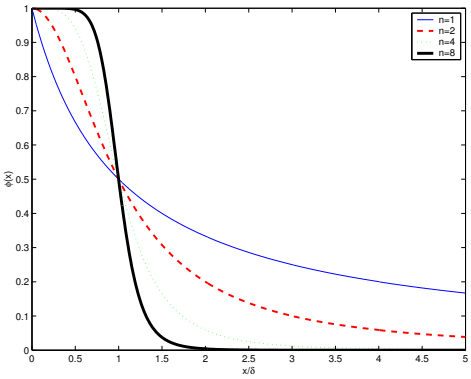

Fig. 1. The shapes of the update function $\phi(x)$ for different parameters.

Reactions (7-10) represent the uptake of nutrients from the environment. We append this list of reactions with 8 pseudoreactions corresponding to the 8 metabolites. We want to find the smallest network that can sustain the production of metabolite H in the absence of metabolite D. Following (9), this problem can be cast as

$$\begin{aligned} &\text{minimize } \mathrm{card}(v), \\ &\text{subject to } Sv = 0, v \succeq 0, \\ &\qquad v_{18} \succ \varepsilon, v_{10} = 0, \end{aligned} \tag{13}$$

where S is the stoichiometric matrix of the network. We compute a solution to this problem with Algorithm 1, and in less than a few seconds obtain $\{1, 4, 7, 8\}$ as a result. We can easily (manually) verify that it is indeed impossible to produce H with fewer than 4 reactions.

As mentioned earlier, the minimal network problem generally does not admit a unique solution. We are interested in generating multiple minimal solutions. This is done by adding a constraint in the optimization problem that makes sure that it does not generate a solution that is already known. Consider the problem shown in (9). Suppose that C is known minimal set of reactions that we wish to exclude from the search. We can add a constraint

$$\sum_{i\in C^*} v_i \geq \varepsilon \tag{14}$$

into the optimization problem. Here $C^* = (C \cup U)^c$ and ε is some positive number. The idea behind the constraint in (14) is intuitive. We basically force the nullified flows not to be all zero. Notice that this constraint is actually convex, so it can be readily incorporated into the convex relaxation (10).

There is a possibility that the optimization solver will just add a reaction to the minimal set C so that (14) is satisfied. We therefore need to check if the new solution is indeed a minimal set. Suppose that the new solution nullifies a set of reaction N, then checking the minimality of the new solution can be done by solving the following optimization problem.

$$\begin{aligned} &\text{minimize } \mathrm{card}(v), \\ &\text{subject to } Sv = 0, v \succeq 0, \\ &\qquad v_{i\in P} \succ 0, v_{i\in(U\cup N)} = 0. \end{aligned} \tag{15}$$

If the argument that minimizes the problem in (15) coincides with the new solution, we can conclude that it is indeed minimal. As before, we can solve (15) using the relaxation algorithm in Algorithm 1.

We apply this algorithm to find other solution(s) for the problem given in Example 3.2. We found $\{2, 5, 7, 9\}$ as a new solution. Again, we can manually verify that $\{1, 4, 7, 8\}$ and $\{2, 5, 7, 9\}$ are the only minimal solutions to this problem. Therefore, in this example, the proposed algorithm is able to provide us with all the minimal solutions.

B. Minimal knockout problem

The second type of problems that we are interested in, is the minimal knockout problem. This type of problems is related to finding the minimal set of reactions to knockout, in order to shutoff the production of certain metabolites. Knocking out a reaction means, mathematically, constraining the flow to be zero. In metabolic engineering, reaction knockout is carried out by deactivating/deleting the gene(s) that produces the enzymes that facilitate the reaction [29].

We want to find the minimal number of reaction knockouts that will shutoff a certain set of product metabolites represented by $P \subset \{1, \ldots, n\}$. We assume that we can only disable regular reactions and uptake reactions (by starving the organism of the corresponding nutrients), and not pseudoreactions. More specifically, we assume that the set of reactions that can be knocked out is given by M. A knockout can then be parameterized by a set $K \subset M$ of reactions whose flows are constrained to zero.

For simplicity, let us consider the special case where P is a singleton. That is, we assume that we want to shutoff only one product. The minimal knockout problem can then be formulated as follows.

Definition 3.3 (Minimal knockout problem): Minimize $\mathrm{card}(K)$ subject to $K \subset M$, such that the following linear optimization problem is not feasible.

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & Sv = 0, v \succeq 0, v_{i\in K} = 0, v_{i\in P} \succeq \epsilon, \end{array} \tag{16}$$

for any $\epsilon > 0$.

We can write down (16) in a standard LP form

$$\begin{array}{ll} \text{minimize} & 0^T v \\ \text{subject to} & Av \succeq b, v \succeq 0, \end{array} \tag{17}$$

where

$$A = \begin{bmatrix} S \\ -S \\ I_K \\ -I_K \\ I_P \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \epsilon \end{bmatrix},$$

and I_K and I_P are the matrices formed by the rows of an identity matrix, corresponding to the set K and P respectively.

The dual problem of (17) is given by

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y \preceq 0, y \succeq 0. \end{array} \tag{18}$$

Let us partition y according to A^T,

$$A^T y =: S^T y_1 - S^T y_2 + I_K^T y_3 - I_K^T y_4 + I_P^T y_5.$$

We can then rewrite (18) as

$$\begin{array}{ll} \text{maximize} & \epsilon y_5 \\ \text{subject to} & \begin{cases} S^T y_1 - S^T y_2 + I_K^T y_3 - I_K^T y_4 + I_P^T y_5 \preceq 0, \\ y \succeq 0. \end{cases} \end{array} \tag{19}$$

The duality theorem of linear programming states that the infeasibility of the primal problem (16) is equivalent to the optimal value of the dual problem (19) being $+\infty$. Using this result, we can obtain a geometric necessary and sufficient condition for the infeasibility of the primal problem, as stated in the following theorem.

Theorem 3.4: The primal problem (16) is infeasible if and only if

$$I_P^T \in \varkappa + \mathrm{im}\, S^T + \mathrm{im}\, I_K^T, \tag{20}$$

where $\varkappa$ is the nonpositive cone and im denotes the image of a matrix.

$$\varkappa = \{y \mid y \preceq 0\}.$$

Proof: (if) Suppose that (20) holds. For any $y_5 \geq 0$, we can always find γ_1, γ_2 and $\gamma_3 \preceq 0$ such that

$$I_P^T y_5 = S^T \gamma_1 + I_K^T \gamma_2 + \gamma_3. \tag{21}$$

Since γ_1 and γ_2 can always be written as

$$\gamma_1 = y_2 - y_1, \ \gamma_2 = y_4 - y_3,$$

where $y_i \geq 0, i = 1, \ldots 4$, it follows that the optimal value of the dual problem (19) is $+\infty$.

(only if) Suppose that (20) is false. From the same argument as above, y is feasible in the dual problem only if $y_5 = 0$. It follows that the optimal value of the dual problem is finite and coincides with that of the primal problem. ■

We introduce the following notation.

Notation 3.5: We denote the vectors obtained by collecting the rows corresponding to M from I_P^T as p_1. The remaining rows form another vector p_2. Similarly, S^T is decomposed into S_1^T and S_2^T.

With this notation, (20) can be written as

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \in \varkappa + \mathrm{im} \begin{bmatrix} S_1^T \\ S_2^T \end{bmatrix} + \mathrm{im} \begin{bmatrix} I_K^T \\ 0 \end{bmatrix}. \tag{22}$$

The problem of finding the smallest knockout set that renders the production of a product P infeasible can then be cast as finding a space of the form $\mathrm{im} \begin{bmatrix} I_K^T \\ 0 \end{bmatrix}$ with the smallest dimension, such that (22) holds. We can then write down the minimal knockout problem as the following convex cardinality optimization problem.

$$\begin{array}{ll} \text{minimize} & \mathrm{card}\,(\max(0, \eta)) \\ \text{subject to} & S_2^T y - p_2 \succeq 0, \ S_1^T y - p_1 + \eta \succeq 0, \end{array} \tag{23}$$

with η and y as the variables. The nonzero entries of the optimal η values represent the reactions that are knocked out. We can rewrite (23) without the $\max$ operator as follows.

$$\begin{array}{ll} \text{minimize} & \mathrm{card}\,(\xi) \\ \text{subject to} & \begin{cases} S_2^T y - p_2 \succeq 0, \ S_1^T y - p_1 + \eta \succeq 0, \\ \xi - \eta \succeq 0, \ \xi \succeq 0, \end{cases} \end{array} \tag{24}$$

with ξ, η, and y as the variables. Notice that (24) is amenable to the ℓ_1 relaxation technique that we have discusses previously (see Algorithm 1). Therefore, a convex relaxation algorithm for solving (24) can be given as follows.

Algorithm 2 Convex relaxation for minimal network problem

Require: A stoichiometric matrix S, a set of target products P, and a set of eligible knockouts M

1: Initialize the weight $w^{(1)} = [1, \ldots, 1] \in \mathbb{R}^m$, and the counter $k = 1$.

2: **repeat**

3: Solve the Linear Programming problem:

$$\text{minimize} \sum_{i=1}^{m} w_i^{(k)} \xi_i^{(k)}, \tag{25}$$

$$\text{subject to } S_2^T y^{(k)} - p_2 \succeq 0,\ S_1^T y^{(k)} - p_1 + \eta^{(k)} \succeq 0,$$
$$\xi^{(k)} - \eta^{(k)} \succeq 0,\ \xi^{(k)} \succeq 0,$$

to find the flow for the $k-$th iteration, $v^{(k)}$.

4: Update weight according to

$$w_i^{(k+1)} = \phi(\xi_i^{(k)}), \tag{26}$$

where

$$\phi(x) := \frac{\delta^n}{\delta^n + x^n}, \tag{27}$$

for some parameters $\delta, n > 0$.

5: **until** $\left\|\xi^{(k)} - \xi^{(k-1)}\right\| \leq \delta$, where $\delta > 0$ is a small number that indicates the convergence of the iteration.

Example 3.6: Consider the small network in Example 3.2. We want to find a minimal set of knockouts that can render the production of H infeasible. Moreover, we assume that we can only knockout uptake reactions, which means $M = \{7, 8, 9, 10\}$. We run Algorithm 2 for this problem, and obtained $\{7, 9\}$ which corresponds to starving the system of A and C. We can verify that this is indeed a minimal knockout set for our purpose.

We can find alternative solutions to the one already computed by the algorithm by introducing a linear constraint, as discussed in the Subsection III-A. Applying this technique, we obtain another minimal set $\{2, 3\}$, which corresponds to starvation of B and C. We can verify that these are indeed two sets of minimal reactions.

IV. Metabolome scale applications

This section is devoted to the application of the algorithms presented in the previous section in metabolome scale problems. We use the metabolic network model constructed by Palsson and coworkers [3]. The network consists of 762 metabolites, 148 uptake reactions, and 1177 regular reactions. We append the list of reactions with 762 pseudoreactions/sinks for each of the metabolites. The connectivity of the network is shown in Figure 2.

All the computation in this paper was implemented in MATLAB using the `cvx` toolbox [30]. The program runs on an Intel Xeon 2.8 GHz processor, and each of the computations below takes less than 10 seconds to complete.

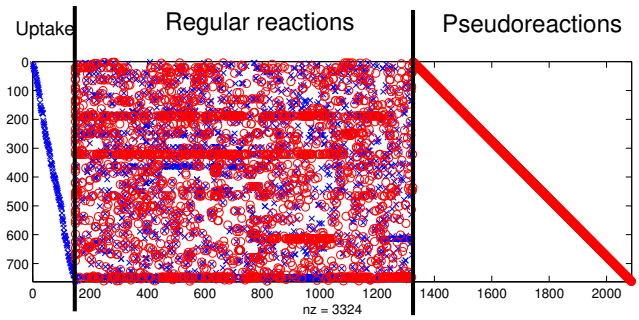

Fig. 2. The elements of the metabolic network of *E. coli* that we use in this paper. Each row represents a metabolite, and each column represents a reaction. The $\times$ signs indicate positive coefficients, $\circ$ signs indicate negative coefficients. Notice that we have three sections in the network, uptake reactions, regular reactions, and pseudoreactions/sinks.

A. *Minimal network of* Escherichia coli

We are interested in finding the minimal metabolic network in *E. coli* that can sustain the production of biomass on a glucose-only medium [22]. Biomass production is an abstract way to model cell growth. To model this medium, we assume that the set of metabolite uptakes consists of only carbon dioxide, hydrogen, potassium, sodium, ammonium, oxygen, phosphate, sulphate, glucose, and water. The feasibility of biomass production is then modeled as the feasibility of a sink containing various components of *E. coli* biomass.

Applying Algorithm 1 to the problem, we obtain a network consisting of 247 metabolic reactions. This is smaller than 25% of the number of regular reactions, which indicates a huge redundancy in the metabolic network of *E. coli* in a glucose-only growth medium. The flows of the minimal network is shown in Figure 3. The largest flows in the plot are annotated. For example, we can see that there is a huge uptake of hydrogen, as well hydrogen as byproduct. Another major byproduct is water. Two main reactions are marked with R1 and R2 in Figure 3. These are reactions that convert ADP (*Adenosine diphosphate*) into ATP (*Adenosine triphosphate*). ATP is the main energy storage in the cell. which is required in most anabolic reactions in the cell. The creation of ATP can thus be seen as energy uptake by the cell. R1 is the ATP synthase reaction, the main synthesis reaction of ATP. R2 is a reaction in the Nucleotide Salvage Pathway that recovers bases and nucleotides that are formed during degradation of RNA and DNA.

B. *Minimal knockout for* Escherichia coli

In this problem, we are interested in finding the minimal knockout set that disable the biomass production in the metabolic network of *E. coli*. We use the same network as in the previous subsection. By applying Algorithm 2 to the problem, we obtain a minimal set consisting of 7 reactions. The reactions are listed below.

Reaction: udpg $\rightarrow$ udpgal Pathway: Alternate Carbon Metabolism
Reaction: akg + coa + nad $\rightarrow$ CO_2 + nadh + succoa Pathway: Citrate Cycle

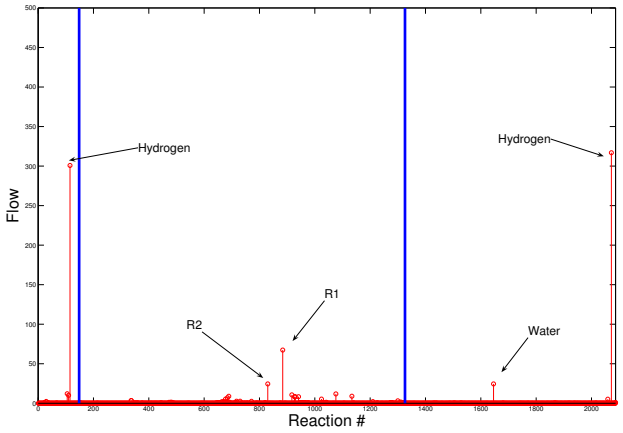

Fig. 3. The flows of the minimal metabolic network of *E. coli* on a glucose-only medium. The left section consists of uptake reactions, the middle section consists of regular reactions, and the right section consists of pseudoreactions. The dominant flows are marked and annotated.

Reaction: g3p + nad + pi → 13dpg + H + nadh Pathway: Glycolysis/Gluconeogenesis
Reaction: ADP + 4 H + pi → ATP + 3 H + H_2O Pathway: Oxidative Phosporylation
Reaction: accoa + pi → actp + coa Pathway: Pyruvate Metabolism
Reaction: thr-L → 2obut + NH_4 Pathway: Valine, leucine, and isoleucine metabolism
Reaction: ala-L + btn + 2 H → cys-L + dtbt Pathway: Cofactor and Prosthetic Group Biosynthesis

It is interesting to contrast this result with that of the minimal network problem. On the one hand, the network is *robust*, in the sense that growth can be sustained even if more than 75% of the network is knocked out. However, we also see *fragility* in the network, where knocking out 7 *certain* reactions in the network can disable growth. This minimal set of reactions is not unique. In fact, using the method discussed in the previous section, we can find other minimal knockout set with 7 reactions that disable the production of biomass in *E. coli.*

Acknowledgement. The authors would like to thank Michael Zavlanos, Calin Belta and Stephen Boyd for valuable discussion during the preparation of this paper.

REFERENCES

[1] D. F. Savage, J. Way, and P. A. Silver, "Defossiling fuel: How synthetic biology can transform biofuel production," *ACS Chemical Biology*, vol. 3, no. 1, pp. 13–16, 2008.

[2] N. D. Price, J. L. Reed, and B. O. Palsson, "Genome-scale models of microbial cells: evaluating the consequences of constraints.," *Nat Rev Microbiol*, vol. 2, pp. 886–97, Nov 2004.

[3] J. L. Reed, T. D. Vo, C. H. Schilling, and B. O. Palsson, "An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR).," *Genome Biology*, vol. 4, no. 9, p. R54, 2003.

[4] M. Imielinski, C. Belta, A. Halász, and H. Rubin, "Investigating metabolite essentiality through genome-scale analysis of Escherichia coli production capabilities.," *Bioinformatics*, vol. 21, pp. 2008–16, May 2005.

[5] M. Imielinski, C. Belta, H. Rubin, and A. Halsz, "Systematic analysis of conservation relations in Escherichia coli genome-scale metabolic network reveals novel growth media.," *Biophys J*, vol. 90, pp. 2659–2672, Apr 2006.

[6] A. Varma, B. W. Boesch, and B. O. Palsson, "Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates.," *Appl Environ Microbiol*, vol. 59, pp. 2465–73, Aug 1993.

[7] 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*, vol. 60, pp. 3724–31, Oct 1994.

[8] D. Segr, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks.," *Proc Natl Acad Sci U S A*, vol. 99, pp. 15112–15117, Nov 2002.

[9] T. Shlomi, O. Berkman, and E. Ruppin, "Regulatory on/off minimization of metabolic flux changes after genetic perturbations.," *Proc Natl Acad Sci U S A*, vol. 102, pp. 7695–7700, May 2005.

[10] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson, "Integrating high-throughput and computational data elucidates bacterial networks.," *Nature*, vol. 429, pp. 92–6, May 2004.

[11] D. Segr, A. Deluna, G. M. Church, and R. Kishony, "Modular epistasis in yeast metabolism.," *Nat Genet*, vol. 37, pp. 77–83, Jan 2005.

[12] I. Thiele, T. D. Vo, N. D. Price, and B. . Palsson, "Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants.," *J Bacteriol*, vol. 187, pp. 5818–5830, Aug 2005.

[13] D. Deutscher, I. Meilijson, M. Kupiec, and E. Ruppin, "Multiple knockout analysis of genetic robustness in the yeast metabolic network.," *Nat Genet*, vol. 38, pp. 993–998, Sep 2006.

[14] S. Klamt and E. D. Gilles, "Minimal cut sets in biochemical reaction networks.," *Bioinformatics*, vol. 20, pp. 226–34, Jan 2004.

[15] S. Klamt, "Generalized concept of minimal cut sets in biochemical networks.," *Biosystems*, Nov 2005.

[16] M. Imielinski and C. Belta, "Exploiting the pathway structure of metabolism to reveal high-order epistasis," *BMC Systems Biology (in press)*, 2008.

[17] A. P. Burgard and C. D. Maranas, "Probing the performance limits of the Escherichia coli metabolic network subject to gene additions or deletions.," *Biotechnol Bioeng*, vol. 74, pp. 364–75, Sep 2001.

[18] A. P. Burgard, P. Pharkya, and C. D. Maranas, "Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization.," *Biotechnol Bioeng*, vol. 84, pp. 647–657, Dec 2003.

[19] S. S. Fong, A. P. Burgard, C. D. Herring, E. M. Knight, F. R. Blattner, C. D. Maranas, and B. O. Palsson, "In silico design and adaptive evolution of Escherichia coli for production of lactic acid.," *Biotechnol Bioeng*, vol. 91, pp. 643–648, Sep 2005.

[20] A. Varma and B. O. Palsson, "Metabolic flux balancing: Basic concepts, scientific and practical use," *Nature Bio/Technology*, vol. 12, pp. 994–998, 1994.

[21] J. S. Edwards, M. W. Covert, and B. O. Palsson, "Metabolic modelling of microbes: the flux balance approach," *Environmental Microbiology*, vol. 4, no. 3, pp. 133–140, 2002.

[22] A. P. Burgard, S. Vaidyaraman, and C. D. Maranas, "Minimal reaction sets for Escherichia coli metabolism under different growth requirements and uptake environments," *Biotechnology Progress*, vol. 17, pp. 791–797, 2001.

[23] S. Boyd, "ℓ_1-norm norm methods for convex cardinality problems." Lecture Notes for EE364b, Stanford University, 2007. Available online at www.stanford.edu/class/ee364b/.

[24] A. Brooke, D. Kendrick, A. Meeraus, and R.Raman, *GAMS: The Solver Manuals*. Washington DC: GAMS Development Corp., 1998.

[25] A. A. Julius, M. Zavlanos, S. Boyd, and G. J. Pappas, "Genetic network identification using convex programming." poster and abstract at the *8th Int. Conf. Systems Biology*, also submitted for publication., 2007.

[26] M. M. Zavlanos, A. A. Julius, S. Boyd, and G. J. Pappas, "Identification of stable genetic networks using convex programming," in *Proc. American Control Conference*, (Seattle, USA), 2008.

[27] A. Papachristodoulou and B. Recht, "Determining interconnections in chemical reaction networks," in *Proc. American Control Conference*, (New York, USA), pp. 4872 – 4877, 2007.

[28] S. Han, Y. Yoon, and K. H. Cho, "Inferring biomolecular interaction networks based on convex optimization," *Computational Biology and Chemistry*, vol. 31, no. 5-6, pp. 347–354, 2007.

[29] H. Alper, Y. S. Jin, J. F. Moxley, and G. Stephanopoulos, "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*," *Metabolic Engineering*, vol. 7, no. 3, pp. 155–64, 2005.

[30] S. Boyd and M. C. Grant, "cvx – MATLAB software for disciplined convex programming," 2005. http://www.stanford.edu/∼boyd/cvx/.