Testing Thermodynamic Compliance of Chemical Reaction Networks in Polynomial Time on Average

Meredith McCormack-Mager Carlos Munoz
Zev Woodstock

July 23, 2015

Abstract

Violations of the second law of thermodynamics often occur undetected in chemical reaction networks, leading to inaccuracy of the model. Beard et al. have given necessary and sufficient conditions for determining thermodynamic feasibility of flux vectors based on cycles of an oriented matroid derived from a reaction network's stoichiometric matrix. Their theorem implies an algorithm that requires computing all cycles of the oriented matroid, of which we show there are sometimes exponentially many. Such an algorithm would thus run in exponential time. We extend the ideas of Beard et al.to find a polynomial time algorithm for checking thermodynamic feasibility of a reaction network. Rather than relying on cycle computation, our algorithm instead depends on interior point methods for linear programming, which run in polynomial time in the worst case, and at most sub-linear time on average. As linear programming methods continue to improve, so will the efficiency of our algorithm. Of independent interest, we demonstrate several nice theorems for cycle computation.

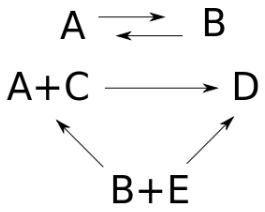
1 Introduction

Chemical reaction networks are a fundamental modelling tool in a variety of scientific disciplines. Though the name suggests a restriction to chemistry, chemical reaction networks can be used to analyze processes as diverse as the body's immune response to HIV[10], analysis of a genome[9], and finding the optimal material with which to build an airplane[4].

In order to understand and model specific biological process, new mathematical tools must be discovered or added onto. Some of the mathematical topics we conducted research on were oriented matroids and computation of their cycles, linear programming, and new definitions and theorems that apply to the thermodynamics of chemical reaction networks.
Chemical reaction networks can consist of several consecutive reactions that may or may not be elementary consecutive reactions that are used to represent biological reactions. Chemical reaction networks can also be applied to synthetic reactions meaning that the paper can be applied to almost any set of chemical reaction networks. The implications of this is that our findings not only apply to purely biochemistry applications, but they are also of some use to the pure chemistry field.
The notion of chemical reaction networks is similar to the mathematical concept of reaction graph. Reaction graphs are used to represent the transformations of reaction species. Each transformation of a specific reaction specie is unique. The reaction graphs can be used to depict a visual representation of the reaction mechanism or a sequence of synthetic reactions. While our focus in this paper is not to create the reaction graphs, but to study the underlying thermodynamics of a given chemical reaction network. More specifically whether or not the system's thermodynamics are feasible or unfeasible.

2 Background

Chemistry Defintions: A simple case that we study is the scenario where there are three species. For example we have species A, B, and C.



For the above picture we consider C to be an intermediate. An intermediate is a species that is produced and then consumed. We consider species A and B to be terminal because they are not intermediates.

One concept of chemistry that plays a key role in the success of determining the feasibility of a chemical reaction network is the law of mass action. The law of mass action is a mathematical model that stemmed from the research of Cato M. Guldberg and Peter Waage. The model allows for the explanation and prediction of phenomena in chemistry. The phenomena that we are interested in studying by law of mass action are the behaviors of homogeneous mixtures that exhibit dynamic equilibrium(s). Homogeneous implies that the solution must be well-stirred. In short in able to study various chemical reaction networks we must know some information about the system that we are studying.
Thermodynamics We are particularly concerned with the 2^{nd} law of thermodyanamics, which states that energy can neither be created, nor can it be destroyed. Two important implications that apply to biochemical reactions are as follows:

1. Spontaneous chemical reactions increase the disorder in the universe. We know that energy is never lost, created or destroyed, only transformed, therefore energy is lost in the form of heat.

2. Spontaneous processes will proceed to a state with the least potential energy.

In essence the impact that these implications have on biochemical processes is that a feasible and closed system will go to a state with the least potential energy without external influences.

Properties of **S** Matrix, or Network Structure:

1. A stoichiometric matrix is an M-by-R matrix, where M equals the total number of species in a model, and R equals the total number of reactions in a model. Each row corresponds to a species, and each column corresponds to a reaction.

2. The matrix indicates which species and reactions are involved as reactants and products.

3. Reactants appear as negative values.

4. Products appear as positive values.

5. All other locations in the matrix contain a 0.

The null space of an $m \times n$ matrix $\mathbf{A}$, denoted Null $\mathbf{A}$, is the set of all solutions to the homogeneous equation $\mathbf{Ax = 0}$. Written in set notation, we have

$$\text{Null } \mathbf{A} \ = \{\mathbf{x} : \mathbf{x} \in R^n \text{and Ax}=\mathbf{0}\}$$

If ξ is a subspace of R^n, a matroid $\mathscr{M}$ is defined as follows. A "vector" of $\mathscr{M}$ is a sign pattern of the form sgn(υ) for some non-zero vector $\upsilon \in \xi$. The set of all vectors of $\mathscr{M}$ is denoted $\mathscr{V}$. A "cycle" of $\mathscr{M}$ is a member of $\mathscr{V}$ which has minimal support.

A sign pattern α is an n-tuple, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ each of whose entries is taken from the set $\{+, -, 0\}$. A signed support of a sign pattern α is the pair (α^+, α^-). Minimal support occurs if $\beta_i \in \mathscr{V}$ and there is no $\beta_j \in \mathscr{V}$, $i \neq j$, such that the signed support of β_j is a proper subset of the signed support of βi.

Circuit Axioms:

Set of cycles of $\mathscr{M}$ is denoted $\mathscr{C}$

1. $\emptyset \notin \mathcal{C}$
2. If $\alpha \in \mathscr{C}$ then its negative $-\alpha \in \mathscr{C}$ or $\mathscr{C} = -\mathscr{C}$
3. For all $\alpha, \beta \in \mathscr{C}$, if $\alpha \subseteq \beta$, then $\alpha = \beta$ or $\alpha = -\beta$
4. Suppose $\alpha, \beta \in \mathscr{C}$, $\alpha \neq \beta$ and $\gamma \in \alpha^+ \cap \beta^-$ there is a $\phi \in \mathscr{C}$ such that

$$\phi^+ \subseteq (\alpha^+ \cup \beta^+) \backslash \{\gamma\}$$

$$\phi^- \subseteq (\alpha^- \cup \beta^-) \backslash \{\gamma\}$$

The next thing is to check that a the collection of signed co-cycles $\mathscr{M}^*$ satisfies exactly the same set of axioms stated above. While we don't exploit this property in our paper, it could well be a potential avenue in the future. Naturally the concept of the dual oriented matroid arises. More specifically the dual matroid $\mathscr{M}^*$ is obtained from the orthogonal complement of the subspace $\xi \in R^n$ which is denoted as $\xi^\perp$. In other words $(\mathscr{M}(\xi))^* = \mathscr{M}(\xi^\perp)$, the previous statement may be hard to picture, since it occurs in a high-dimensional space. In order for this to be obvious we need to state an important orthogonality property for oriented matroids:

Definition 1. ($\perp$) *If $\alpha \in \mathscr{C}$ is a cycle and $\beta \in \mathscr{C}^*$ is a cocycle of an oriented matroid with $(\alpha^+ \cap \beta^+) \cup (\alpha^- \cap \beta^-) \neq \emptyset$, then also $(\alpha^+ \cap \beta^-) \cap (\alpha^- \cap \beta^+) \neq \emptyset$, and conversely.*

A "vector" of an oriented matroid is any composition of cycles. On the other hand a "covector" is a vector of the dual oriented matroid, more specifically any composition of cocycles. The notation would work in the following manner, the "vectors" of $\mathscr{M}^*$ are referred to as the "co-vectors" of $\mathscr{M}$. The set of all "co-vectors" are denoted as $\mathscr{V}^*$. Naturally it follows that the "cycles" of $\mathscr{M}^*$ are referred to as the "co-cycles" of $\mathscr{M}$, and the set of co-cycles are denoted $\mathscr{C}^*$.

Theorem 1. *A flux vector* $\boldsymbol{J}$ *is feasible according to steady state mass balance and the thermodynamic constraint if and only if*

1. $\boldsymbol{S}\cdot\boldsymbol{J} = 0$ *and*
2. $sgn(\tilde{\boldsymbol{J}})$ *is orthogonal to* γ *for each* $\gamma \in \mathscr{C}^*$.
   If $\boldsymbol{J}_j$ *is allowed to be 0, even if* $\Delta\mu_j \neq 0$, *then condition* (2) *is replaced by* $(2')$*:*
2. $|\langle \gamma,\ sgn(\tilde{\boldsymbol{J}})\rangle| < |\langle \gamma, \gamma\rangle|$ *for each* $\gamma \in \mathscr{C}^*$.

If a flux vector $\boldsymbol{J}$ *satisfies conditions (1) and (2), we will say that* $\boldsymbol{J}$ *is strictly feasible, while if it only satisfies (1) and the relaxed condition (2'), we will say that* $\boldsymbol{J}$ *is T-feasible, or simply feasible.*

Definition 2. *Vectors* v *and* w *are* sign orthogonal *if either of the following is true:*

1. *The support of* v *and support of* w *contain no common elements.*
2. *There exists an index at which* v *and* w *have the same nonzero sign, and another index at which they have opposite sign.*

3 Oriented Matroids

3.1 Cycle Computation

1. Obtain Stoichiometric Matrix from Chemical Reaction Network

   Reaction 1: A $\rightarrow B$
   Reaction 2: B $\rightarrow C$
   Reaction 3: C $\rightarrow A$
   Reaction 4: C $\rightarrow D$
   Reaction 5: D $\rightarrow B$
   Reaction 6: 0 $\rightarrow A$
   Reaction 7: B $\rightarrow 0$

   Slightly abusing the notation reaction 6 and reaction 7 denote the cases where there are external fluxes that transport A into the system, and B out of the system.

   The chemical reactions above yield the following Stoichiometric Matrix

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

Now note that the last two columns correspond to the external fluxes of the system. Since we're only interested in the internal reactions we omit the last two columns, and obtain $\widetilde{S}$.

$$\widetilde{S} = \begin{bmatrix} -1 & 0 & +1 & 0 & 0 \\ +1 & -1 & 0 & 0 & +1 \\ 0 & +1 & -1 & -1 & 0 \\ 0 & 0 & 0 & +1 & -1 \end{bmatrix}$$

2. Compute Nullspace of Stoichiometric Matrix
   The nullspace of a matrix is found by setting $\widetilde{S}\mathbf{x} = 0$ where x is a $n \times 1$ vector. We find that a basis for $\widetilde{S}$ is as follows:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} x_2$$

3. Compute Signed Vectors of Basis

$$\begin{aligned} \alpha_1 &= sgn\big(\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \end{bmatrix}\big) = \begin{bmatrix} + & + & + & 0 & 0 \end{bmatrix} \\ \alpha_2 &= sgn\big(\begin{bmatrix} 0 & 1 & 0 & 1 & 1 \end{bmatrix}\big) = \begin{bmatrix} 0 & + & 0 & + & + \end{bmatrix} \\ \text{C} &= \left\{ \begin{bmatrix} + & + & + & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & + & 0 & + & + \end{bmatrix} \right\} \end{aligned}$$

   Now let it be noted that α_1 and α_2 belong to $\mathscr{C}$, which is the set of cycles of the Matroid, $\mathscr{M}$. We are allowed to do this since we know that cycles computed from basis vectors have minimal support.

4. Obtain cycles from the signed Vectors of Basis vectors
   Now taking α_1 and α_2 and we realize that they do not have any sign differences, therefore we can not compute more cycles. However, given the following contrived basis:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} x_2$$

   with the following sign vectors and cycles

$$\beta_1 = sgn(\begin{bmatrix}1 & 1 & 1 & 0 & 0\end{bmatrix}) = \begin{bmatrix}+ & + & + & 0 & 0\end{bmatrix}$$
$$\beta_2 = sgn(\begin{bmatrix}0 & -1 & 0 & 1 & 1\end{bmatrix}) = \begin{bmatrix}0 & - & 0 & + & +\end{bmatrix}$$
$$\text{C}=\left\{\begin{bmatrix}+ & + & + & 0 & 0\end{bmatrix},\begin{bmatrix}0 & - & 0 & + & +\end{bmatrix}\right\}$$

We now have that there is a disagreement between β_1 and β_2 at the 2^{nd} position. A disagreement is defined as a sign difference in the j^{th} element between two vectors of $\mathscr{C}$. In order to resolve a disagreement in the j^{th} position between two vectors, a linear combination of the two vectors must be taken such that $c_1 v_{1,j} + c_2 v_{2,j} = 0$ for the j^{th} element.
Therfore, $c_1 = 1$ and $c_2 = 1$, and

$$v_1 = \begin{bmatrix}1 & 1 & 1 & 0 & 0\end{bmatrix} \quad v_2 = \begin{bmatrix}0 & -1 & 0 & 1 & 1\end{bmatrix}$$
$$c_1 v_1 + c_2 v_2 = \begin{bmatrix}1 & 0 & 1 & 1 & 1\end{bmatrix}$$
$$v_3 = \begin{bmatrix}1 & 0 & 1 & 1 & 1\end{bmatrix}$$
$$\beta_3 = sgn\begin{bmatrix}1 & 0 & 1 & 1 & 1\end{bmatrix} = \begin{bmatrix}+ & 0 & + & + & +\end{bmatrix}$$
$$\mathscr{C} = \left\{\begin{bmatrix}+ & + & + & 0 & 0\end{bmatrix},\begin{bmatrix}0 & - & 0 & + & +\end{bmatrix},\begin{bmatrix}+ & 0 & + & + & +\end{bmatrix}\right\}$$

Note that the set $\mathscr{C}$ is now complete.

5. Repeat while checking that the cycles obtained are minimally supported
Now while the previous two examples terminated before this step, this step would be done for Chemical Reaction Networks who have a stochiometric matrix of internal reactions with a nullspace of dimension greater or equal to 3.

3.2 Improvements to Cycle Computation

There are several manners in which one could implement to improve the aforementioned algorithm. One manner in which the algorithm could be made more efficiently would be by allowing the presence of an all positive signed vector to terminate the algorithm since this would imply that the system is not thermodynamically feasible. Another stopping criteria that could also be implemented is if the all minus vector is observed.

3.3 Row-Reduced Echelon Form Basis Approach to Cycle Computation in Oriented Matroids

We use Gauss-Jordan elimination to find a basis in row-reduced echelon form for the nullspace of the stoichiometric matrix. This basis will assist us both in proving that an oriented matroid can contain exponentially many cycles in the worst case and later as a step in our algorithm. Note that Gauss-Jordan elimination runs in $\text{O}(nk^2)$ time for a k-by-n matrix.

Let nullspace $\xi \subseteq \mathbb{R}^n$ be a k-dimensional subspace and η is the nullspace of the stoichiometric matrix. Then let B = $\{v_1, ..., v_k\}$ be a basis for ξ such that

$$\begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix}$$

is in Reduced Row Echelon form. We will use this basis B in the following three proofs.

Theorem 2. *If $v \in B$, then sgn(v) is a cycle of the oriented matroid corresponding to the nullspace.*

Proof. Let $v_i \in$ B and let $\alpha = \text{sgn}(v_i)$. Assume for a contradiction that α is not a cycle. Then there exists nonzero $\beta \in \mathscr{V}$ such that $\beta^+ \subseteq \alpha_i^+$ and $\beta^- \subseteq \alpha_i^-$ and $\beta \neq \alpha$. By the definition of $\mathscr{V}$, there exists $w \in \xi$ such that $\text{sgn}(w) = \beta$. We have $\text{supp}(w) \subset \text{supp}(v_i)$ and we can write w as a linear combination $c_1v_1 + ... + c_kv_k$ of the basis vectors in B. Assume for a contradiction that there exists $c_j \neq 0$, $j \neq i$. There exists an index ℓ at which v_j has a leading one (in particular, the first non-zero index of v_j). Then for all $m \neq j$, we have $v_{m,\ell} = 0$. So $w_\ell = c_j \neq 0$. But $v_{i,\ell} = 0$. Then $\text{supp}(w) \not\subset \text{supp}(v_i)$, a contradiction. Thus, for all $j \neq i$, we have $c_j = 0$. Therefore $w = c_iv_i$. So $\text{sgn}(w) = \pm \text{sgn}(v_i)$, i.e. $\beta = \pm\alpha$. But β is nonzero and $\beta^+ \subseteq \alpha^+$ and $\beta^- \subseteq \alpha^-$, so $\beta = \alpha$, a contradiction. Hence α is a cycle. □

Proposition 1. *If w is a positive linear combination of two vectors who do not have a disagreement, then sgn(w) is not a cycle.*

Proof. Let nonzero $v_1, v_2 \in \xi$ have no disagreements, and let $\alpha = \text{sgn}(v_1)$, $\beta = \text{sgn}(v_2)$. Let $w = c_1v_1 + c_2v_2$ for $c_1, c_2 \geq 0$, and $\gamma = \text{sgn}(w)$. We will show that the signed support of α is strictly contained in the signed support of γ. For all $i \in \alpha^+$, we have we have $\gamma_i = \alpha_i$, since $\beta_i = 0$ or β_i is positive. Likewise, for all $j \in \alpha^-$, we have we have $\gamma_i = \alpha_i$. Since β is nonzero, we have $\gamma = \alpha + \beta \neq \alpha$. So the signed support of α is strictly contained in the signed support of γ, and hence γ does not have minimal signed support, i.e. γ is not a cycle. □

Theorem 3. *If w is the resulting vector from a positive pairwise linear combination that resolves a disagreement between basis vectors from B, then sgn(w) is a cycle.*

Proof. Consider basis vectors v_i, $v_j \in$ B having a disagreement at index ℓ. Let $c_i = |v_{j,\ell}|$ and $c_j = |v_{i,\ell}|$. Then let $w = c_iv_i + c_jv_j$. So $w_\ell = 0$. Let $\gamma = \text{sgn}(w)$. Assume for a contradiction that γ is not a cycle. Then there exists nonzero $\delta \in \mathscr{V}$ such that $\delta^+ \subseteq \gamma^+$ and $\delta^- \subseteq \gamma^-$ and $\delta \neq \gamma$. By the definition of $\mathscr{V}$, there exists $u \in \xi$ such that $\text{sgn}(u) = \delta$. We have $\text{supp}(u) \subset \text{supp}(w)$ and, since B is a basis for ξ, we have $u = d_1v_1 + ... + d_kv_k$. Assume for a contradiction that there exists $d_p \neq 0$ for $p \neq i, j$. There exists an index r at which v_p has a leading one. Then, for all $m \neq p$, we have $v_{m,r} = 0$. So $u_r = d_p \neq 0$. But

$v_{i,p} = 0$ and $v_{j,p} = 0$. So $w_p = c_i v_{i,p} + c_j v_{j,p} = 0$. Then $\text{supp}(u) \not\subset \text{supp}(w)$, a contradiction. Thus, for all $p \neq i, j$, we have $d_p = 0$. Therefore $u = d_i v_i + d_j v_j$. Since $\text{supp}(u) \subset \text{supp}(w)$, $u_\ell = 0$. Then $\frac{d_i}{d_j} = \frac{c_i}{c_j}$. Thus $u = c'w$. So $\text{sgn}(u) = \pm \text{sgn}(w)$, i.e. $\delta = \pm\gamma$. But δ is nonzero and $\delta^+ \subseteq \gamma^+$ and $\delta^- \subseteq \gamma^-$, so $\delta = \gamma$, a contradiction. Hence γ is a cycle. □

Remark. Note that n-wise linear combinations of reduced row echelon form basis vectors do not always result in vectors whose sign vectors are cycles, even if the combination resolves a series of disagreements. Consider the following reduced row echelon basis B:

$$\text{B} = \{v_1, v_2, v_3\} = \{(1, 0, 0, 1, 1), (0, 1, 0, 0, -1), (0, 0, 1, -2, 1)\}$$

Each pairwise combination of the basis vectors contains a disagreement, and we can resolve these disagreements with the following linear combinations:

$$\begin{aligned} w_1 &= v_1 + v_2 = (1, 1, 0, 1, 0) \\ w_2 &= 2v_1 + v_3 = (2, 0, 1, 0, 3) \\ w_3 &= v_2 + v_3 = (0, 1, 1, -2, 0) \end{aligned}$$

Note that there is a disagreement between v_3 and w_1 at index 4, and a disagreement between v_2 and w_2 at index 5. We resolve these disagreements as follows:

$$\begin{aligned} w_4 &= v_3 + 2w_1 = (2, 2, 1, 0, 1) \\ w_5 &= 3v_2 + w_2 = (2, 3, 1, 0, 0) \end{aligned}$$

We find that $\text{sgn}(w_4)$ is not a cycle, because the signed support of w_5 is a subset of the signed support of w_4. Further, we can discern that $\text{sgn}(w_5)$ is a cycle, since it is impossible to generate a triple-wise combination of basis vectors with smaller signed support (a consequence of the leading ones in each basis vector), and none of the basis vectors or pairwise combinations computed above have a signed support contained in the signed support of w_4. We have not yet discovered an efficient method for determining which n-wise linear combinations of basis vectors from B produce cycles, and we are left checking for minimality by comparison to other signed vectors in the matroid. In many cases, this means comparing exponentially many signed vectors to one another.

3.4 Oriented Matroids Can Have an Exponential Number of Cycles

Even if we were to find a polynomial time algorithm for finding each cycle of the oriented matroid stemming from ξ, the complexity of the algorithm would still depend on the number of cycles. We present a family of nullspaces whose corresponding oriented matroids contain exponentially many cycles, guaranteeing exponential complexity in the worst case of such an algorithm. Here we give examples of nullspaces with dimension $k = 3$ and $k = 4$, which contain

$2(2^3-1)$ and $2(2^4-1)$ cycles respectively. We then give a proof that this family of nullspaces extends to a general dimension k.

$$N_3 = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \left(\begin{array}{c|ccc} I_3 & 1 & 1 & 0 \\ & -1 & 0 & 1 \\ & 0 & -1 & -1 \end{array}\right)$$

$$N_4 = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \left(\begin{array}{c|cccccc} I_4 & 1 & 1 & 1 & 0 & 0 & 0 \\ & -1 & 0 & 0 & 1 & 1 & 0 \\ & 0 & -1 & 0 & -1 & 0 & 1 \\ & 0 & 0 & -1 & 0 & -1 & -1 \end{array}\right)$$

Suppose $\xi \subset \mathbb{R}^n$ is a k-dimensional subspace whose basis vectors $\{v_1, \ldots, v_k\}$ are the rows of a matrix N_k with the following structure, which we denote $(\star)$:

1. N is in row reduced echelon form, with I_k making up the first k columns.
2. Let v_i and v_j be distinct row vectors with $i < j$. There exists a column of n_k whose entries are zero except for a 1 in row i and a -1 in row j.
3. Except for the first k columns, every column has the structure described in criterion 2.

Note that each row vector disagrees with every other row vector once, and at a unique entry.

Remark. In this nullspace, we need not concern ourselves with negative coefficients of basis vectors while computing cycles. If we use some positive and some negative coefficients in a linear combination, the resulting vector will never be a cycle, since the support of the resulting vector will contain the support of an all positive or all negative coefficient linear combination. If we use all negative coefficients, the resulting vector is the negative of an all positive linear combination, and, if it is a cycle, will already have been accounted for.

Lemma 1. *Suppose the conditions of $(\star)$, and let $w = c_1 v_{a_1} + \ldots + c_m v_{a_m}$ for $c_i > 0$. If there exists a $c_i = p \neq 1$, then w is not a cycle or every $c_i = p$.*

Proof. Suppose that there exists a $c_j = p \neq 1$. Suppose further that w is a cycle. Let $u = w - v_j$. Then u and v_j have a disagreement due to Proposition 1. In fact, we see that u and v_j have exactly $m-1$ disagreements, one corresponding to each vector summing to make u, a consequence of the structure of N. For every index ℓ at which u and v_j do not have a disagreement, either $u_\ell = 0$ or $v_{j,\ell} = 0$. So w_ℓ is only zero when $u_\ell = v_{j,\ell} = 0$. Thus we can only send at most m-1 previous nonzero entries of u to zero with the addition of v_j. The binary linear combination x of $v_{a_1}, \ldots, v_{a_m}$ resolves all m-1 of these disagreements. To ensure that $\text{supp}(u) \not\supset \text{supp}(x)$, we must resolve every disagreement between u and v_j. Let ℓ be an entry at which u and v_j have a disagreement. There is only

one basis vector $v_{a_b} \neq v_j$ with a nonzero value at ℓ. Then v_{a_b} is necessarily in $\{v_{a_1}, \ldots v_{a_m}\} \backslash \{v_i\}$. In order to resolve the disagreement between u and v_j at ℓ, we must have $c_b = p$. Since each of the $m-1$ disagreements corresponds to a distinct vector in $\{v_{a_1}, \ldots v_{a_m}\} \backslash \{v_i\}$, every $c_i = p$. □

Proposition 2. *Suppose the conditions of* $(\star)$*. If* $w = c_1 v_1 + \ldots + c_k v_k$ *for* $c_i \in \{0, 1\}$ *and* v_i *a row vector of* N_k*, then sgn(w) is a cycle.*

Proof. We will show this using strong induction on the number of nonzero c_i's. Due to Lemma 1, we need only concern ourselves with binary linear combinations of vectors. Let ζ be the set of binary linear combinations of $v_1, \ldots, v_k$.

Base case: Suppose $w = v_{a_1}$. Since N is in reduced row echelon form, we know that w is a cycle due to Theorem 2.

Inductive case: Suppose every nonzero $u = c_1 v_{a_1} + \ldots + c_\ell v_{a_\ell}$ for $c_i \in \{0, 1\}$ and distinct v_{a_i} is a cycle. We will show that, if $\ell + 1 \leq n$, then $w_{\ell+1} = w_\ell + v_{a_{\ell+1}}$ is a cycle. We do so by checking that no vector $x \in \zeta$ has $\text{supp}(x) \subseteq \text{supp}(w_{\ell+1})$. Suppose $\ell + 1 \leq n$ and let $w_{\ell+1} = w_\ell + v_{a_{\ell+1}}$ for distinct v_{a_i}. Every vector $x \in \zeta$ (with the exception of $w_{\ell+1}$) falls into one of two cases.

1. Suppose that x can be written as a linear combination of the same vectors that make up $w_{\ell+1}$. There exists at least one index m at which $v_{a_{\ell+1}}$ and x have a disagreement. Then $w_{\ell+1,m} = 0$ and $x_m \neq 0$. Thus we have $\text{supp}(w_{\ell+1}) \not\supseteq \text{supp}(x)$.

2. Suppose x can only be expressed as a linear combination of basis vectors that includes at least one vector v_j not used to make up $w_{\ell+1}$. We know that v_j has a leading one at some entry m. Then $x_m = 1$ and $w_{\ell+1,m} = 0$, so $\text{supp}(w_{\ell+1}) \not\supseteq \text{supp}(x)$.

Therefore w has minimal signed support, and is thus a cycle. □

If we restrict to binary linear combinations, there are $2^k - 1$ nonzero w's. Hence, we know that N corresponds to an oriented matroid containing at least $2^k - 1$ cycles (in fact, the matroid contains $2(2^k - 1)$ cycles), which is exponentially many.

We note that these subspaces are not the only ones corresponding to exponentially many cycles in the oriented matroid, rather they are a simple example with which to work. Because such a family of subspaces exists, any algorithm dependent on computing all the cycles will run in at least exponential time in the worst case. Whether or not such an algorithm will run in exponential time on average is an open question. If we want to find an algorithm that runs in polynomial time in the worst case, it is clear that we must avoid cycle computation.

4 Construction of a Non-Orthogonal Vector to Determine Thermodynamic Feasibility

Earlier, we defined what it means for two vectors to be sign orthogonal. If we assume that the flux vector is a positive vector (strongly thermodynamically feasible[8]), then we can narrow the criteria for when a vector is orthogonal to the flux vector. As the support of the flux vector will be $\{1, \ldots, n\}$, we know that the first condition for sign orthogonality will only hold for the zero vector when compared to the flux vector. It is forbidden for the zero vector to be a cycle, so we ignore this case. As such, we focus solely on determining whether every cycle satisfies the second orthogonality criterion (Theorem 1) in comparison with the flux vector.

In fact, every nonzero vector is sign orthogonal to the all positive vector except those that have all nonnegative or all nonpositive entries (and are nonzero). If a nonpositive vector is present in a vectorspace, then its negative, a nonnegative vector, is present. Thus, to determine whether there exists a strongly thermodynamically feasible flux vector for a particular chemical reaction network, we need only check whether or not a nonnegative nonzero vector exists in the nullspace of the network's stoichiometric matrix.

4.1 Deriving a System of Inequalities

To accomplish this, we return to the row reduced echelon form basis for our nullspace ξ. Consider the following example in which the row vectors of N are a basis for ξ:

$$N = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -3 & -2 \\ 0 & 1 & 0 & -2 & 4 \\ 0 & 0 & 1 & 1 & -1 \end{pmatrix}$$

Suppose there exists a nonzero nonnegative vector $w \in N$. Then w is a linear combination $c_1v_1 + c_2v_2 + c_3v_3$ of basis vectors v_1, v_2, and v_3. To ensure that the first three entries of w are nonnegative, we restrict $c_1, c_2, c_3 \geq 0$. To make sure the fourth entry of w is nonnegative, we introduce the constraint:

$$-3c_1 - 2c_2 + c_3 \geq 0.$$

To keep the fifth entry nonnegative, we introduce:

$$-2c_1 + 4c_2 - c_3 \geq 0.$$

One nonzero solution to this system of inequalities is $c_1 = 0, c_2 = 1, c_3 = 3$, indicating that a nonzero nonnegative vector in ξ exists, and that the system corresponding to this nullspace is thermodynamically infeasible.

In general, we can construct a set of constraints based on the columns of the matrix describing the nullspace as we did in the above example. This construction results in k nonnegativity restrictions on the coefficients c_i and $n - k$ constraints of the following form, where the $a_{i,j}$'s are taken from the columns

of N not containing leading 1s:

$$\begin{aligned} a_{1,1}c_1 + \ldots + a_{k,1}c_k &\geq 0 \\ a_{1,2}c_1 + \ldots + a_{k,2}c_k &\geq 0 \\ \vdots \qquad\qquad & \vdots \\ a_{1,n-k}c_1 + \ldots + a_{k,n-k}c_k &\geq 0 \end{aligned}$$

Proposition 3. *A chemical reaction network is strongly thermodynamically feasible if and only if there exists no feasible nonzero solution to this system of inequalities.*

4.2 Linear Programming

Using linear programming, we can determine whether a solution in the positive orthant exists for any set of linear inequalities. First, we must alter our system of constraints slightly and introduce an objective function based on a new nonnegative variable c_0.

Let $Z = -c_0$ and let our system of constraints be as follows, where the $a_{i,j}$'s and c_i's are as before:

$$\begin{aligned} -c_0 - a_{1,1}c_1 - \ldots - a_{k,1}c_k &\leq 0 \\ -c_0 - a_{1,2}c_1 - \ldots - a_{k,2}c_k &\leq 0 \\ \vdots \qquad\qquad & \vdots \\ -c_0 - a_{1,n-k}c_1 - \ldots - a_{k,n-k}c_k &\leq 0 \end{aligned}$$

Our original system of inequalities has a feasible solution if and only if we can maximize Z to 0 [7]. However, note that $c_1 = \ldots = c_k = 0$ is always a feasbile solution to this set of inequalities, causing Z to always maximize to 0, but we require a nonzero solution.

4.3 Exclusion of the Trivial Solution

As stated above, our algorithm returns a useless solution, so we must find a way to exclude the all zero result as a feasible solution. We can model our system as a system of halfplanes in k-dimension, where the set of feasible solutions is the largest polytope contained in the intersection of every halfplane. Each halfplane is bounded by a boundary hyperplane, and each of these hyperplanes passes through the origin. One consequence of this is that the origin (corresponding to the all zero solution) is always in the feasible set, which we have already noted. A related consequence is that, if the feasible set is nontrivial, the polytope is unbounded in some direction in the positive orthant.

Suppose P is a cross-section of the positive orthant intersecting every axis at a nonzero value. Let F be the intersection of P and the feasible polytope. If the feasible set is nontrivial, then P will intersect the feasible polytope nontrivially. Otherwise the intersection will be empty.

Proposition 4. *A chemical reaction network is strongly thermodynamically feasible if and only if F is nonempty.*

Suppose without loss of generality that P is the hyperplane $c_1 + \ldots + c_k = 1$. We solve for $c_1 = 1 - c_2 - \ldots - c_k$, which we use to rewrite our system of inequalities as follows:

$$\begin{gathered}
-c_0 + (a_{1,1} - a_{2,1})c_2 + \ldots + (a_{1,1} - a_{k,1})c_k \leq a_{1,1} \\
-c_0 + (a_{1,1} - a_{2,2})c_2 + \ldots + (a_{1,1} - a_{k,2})c_k \leq a_{1,1} \\
\vdots \qquad\qquad \vdots \\
-c_0 + (a_{1,1} - a_{2,n-k})c_2 + \ldots + (a_{1,1} - a_{k,n-k})c_k \leq a_{1,1} \\
c_2 + \ldots + c_k \leq 1.
\end{gathered}$$

We then use interior point methods to solve the linear programming problem of maximizing $Z = -c_0$.

5 Conclusions

We use an interior point method proposed by Anstreicher [1], which runs in $O(\frac{k^3}{log(k)}n)$ time in the worst case. It is known that interior point methods run in at most $O(\sqrt{k}log(k))$ on average[2]. Computational evidence suggests that this is not a strict upper bound, and we hope to soon see an improved average run time analysis.

We hope to further reduce the upper bound on our algorithm by proposing constraints on which stoichiometric matrices might reasonable arise from real life chemical reaction networks. One way to do so would be to restrict the original chemical reaction network to unimolecular and bimolecular reactions. We suggest that termolecular reactions be excluded because this type of elementary reaction is uncommon. For large reaction networks, this restriction would create a sparse stoichiometric matrix, which we believe would speed up the average run time not only of the linear programming subproblem, but also of the nullspace computation and Gauss-Jordan elimination. In a sparse stoichiometric matrix, we also suspect that the dimension of the nullspace will be much smaller than the number of reactions, which will further serve to reduce the run time of our algorithm.

Once identifying whether there are unique subspaces that correspond to these elementary reactions, a mixed system of elementary reactions could be studied to see if they correspond to specific reactions.

We have found a case in which an oriented matroid contains exponentially many cycles, but we do not know whether oriented matroids contain exponentially many cycles on average. We would like to find conditions on a subspace to quickly determine when its corresponding oriented matroid contains exponentially many cycles or polynomially many cycles. We suspect that any such conditions will depend on the number of disagreements between basis vectors.

6 Acknowledgments

Firstly, we would like to thank our research mentor, Dr. Anne Shiu, for her invaluable advice and encouragement. We also thank Dr. Maya Johnson and Dr. Jay Walton for great discussions that led to great ideas.

Ms. McCormack-Mager and Mr. Woodstock were supported by a summer undergraduate research grant awarded by the Texas A&M University Math Department and sponsored by the National Science Foundation (DMS-1460766). Mr. Munoz was funded under the UBM Grant (DMS-1029401).

References

[1] Anstreicher, Kurt M. "Linear programming in O ([n3/ln n] L) operations." SIAM Journal on Optimization 9.4: 803-812. 1999

[2] Anstreicher, Kurt M., et al. Average performance of a self-dual interior-point algorithm for linear programming. University of Iowa, 1992.

[3] Beard, Daniel A., et al. *Thermodynamic constraints for biochemical networks.* Journal of theoretical biology 228.3: 327-333. 2004.

[4] Birakayala, Narayana, and Edward A. Evans. *A reduced reaction model for carbon CVD/CVI processes.* Carbon 40.5: 675-683. 2002.

[5] Bjorner, A., Las Vergnas, M., Sturmfels, B., White, N., Ziegler, G., 1993. Oriented Matroids. Encyclopedia of Mathematics and its Applications, Vol. 46. Cambridge University Press, Cambridge.

[6] Clarke, B.L., 1988. Stoichiometric network analysis. Cell Biophys. 2, 237253.

[7] Cormen, Thomas H. et al. *Introduction to algorithms*. MIT press, 2009.

[8] Reimers, Arne C. *Metabolic Networks, Thermodynamic Constraints, and Matroid Theory*. Diss. Freie Universitt Berlin, 2014.

[9] Schilling, Christophe H., Jeremy S. Edwards, and Bernhard O. Palsson. *Toward metabolic phenomics: analysis of genomic data using flux balances*. Biotechnology Progress 15.3: 288-295. 1999.

[10] Wodarz, D. and Nowak, M. A. *Specific therapy regimes could lead to long-term immunological control of HIV.* Proc. Natl. Acad. Sci. U.S.A., 96(25):144649, 1999.