Atomic Realizations of Chemical Reaction Networks

Joseph Horan
University of Waterloo
jahoran@uwaterloo.ca

June 16, 2013

Abstract: When considering chemical reaction networks (CRNs), we often ignore the atomic compositions of the species involved (atom-free stoichiometry). Thus, a reasonable question is, "When can a CRN support an atomic realization?" That is, when can we assign each species an atomic structure which is distinct from that of the other species? After briefly covering definitions, we consider the main tool which we use, an algorithm from Schuster and Höfer [SH91], and draw out a proof that it obtains all extreme vectors. We then give necessary and sufficient conditions for atomic realizations, and discuss the implications, with reference to Famili and Palsson [FP03].

1 Introduction

In science, when explaining the world around us, we encounter systems involving different types of objects, species, elements, amongst other things. In chemistry, one such general system is what we call a *chemical reaction network*, or CRN; this is a set of individual chemical reactions. This system tells us about the participants in the reactions, or *species*, and how they react with each other over time, usually by way of differential equations. This is the field of *stoichiometry*. This is of interest primarily because it allows us to understand real-world processes. It turns out that while these systems tend to be extraordinarily complicated (in particular, highly non-linear), they are much more tractable than general systems, hence we wish to study them in a more in-depth fashion. First, let us briefly refresh ourselves on some of the terminology which we encounter; it is assumed that the reader has a grasp of basic linear algebra and differential equations, though some convex analysis will be explained.

A CRN is specified by four components: the species which participate in the set of reactions; the *stoichiometric coefficients*, usually given in matrix form as Γ, which describe the extent to which a species participates in a reaction; the *kinetics*, which describe the laws (usually differential equations) under which the species react; and the *rate constants*, which describe how quickly reactions take place. Ignoring the rates is reasonable for our purposes, as they are quite finicky and will not provide extra detail in the analysis to come. So, we can represent a chemical reaction network by

$$\sum_{i=1}^{n} \alpha_{ij} S_i \longrightarrow \sum_{i=1}^{n} \beta_{ij} S_i \ , \ j = 1, \dots, r, \tag{1.1}$$

where each S_i is a chemical species, and the α_{ij}'s and β_{ij}'s are stoichiometric coefficients, all of which are non-negative. Notice that we merely denote a species by S_i, as opposed to an atomic formula. This is partially due to notation, but is also due to the fact that we tend to consider *atom-free* stoichiometries; that is, it is common to discuss the kinetics of the system without reference to the exact chemical formulae involved. Indeed, in some cases, this is the only option, as we may not know what formula a species actually

has. Moreover, if two species have the same chemical formula, but different structure, we call them *isomers*. These make our job slightly more challenging.

Now, as per [HJ72], a *conservative* system is one for which there exists a positive vector in the nullspace of the transpose of the stoichiometric matrix. That is, $\exists z \in \text{null}(\Gamma^T) \cap \mathbb{R}^n_{>0}$. This definition encodes the idea that the system maintains the total amount of species involved; in some sense, it is a first integral, to use DEs terminology. A consequence of this is that in conservative systems, the reactions must obey the *law of atomic balance*; that is, the total amount of a particular element must be conserved in the system, and on both sides of the reactions. So considering atom-free stoichiometries as opposed to ones with atomic structure leads to a certain loss of information here. We shall develop some mathematics to deal with atomic structure, and retrieve that information.

Treating the CRN as a dynamical system, one common and reasonable model is

$$\dot{x} = \Gamma R(x), \tag{1.2}$$

where $x \in \mathbb{R}^n_{\geq 0}$ is the species vector containing the concentrations of each species, Γ is the stoichiometric matrix defined by $\Gamma_{ij} = \gamma_{ij} = \beta_{ij} - \alpha_{ij}$, and $R(x)$ is the reaction rate vector, or kinetics of the system. For our purposes, we actually do not care what the kinetics are, but we do imagine them to be decently well-behaved (probably C^1).

Now, let $z \in \text{null}(\Gamma^T)$. Then we have:

$$z^T\dot{x} = z^T\Gamma R(x) = (\Gamma^T z)^T R(x) = 0^T R(x) = 0.$$

Integrating gives us:

$$z^T x = z^T x_0.$$

Then, if z is a positive vector and x_0 is non-negative, we see that all the species are bounded, and thus, by repeated application of Existence-Uniqueness, any solution of (1.2) is defined for all $t \geq 0$.

This is meaningful because in a way, each such vector z encodes a conservation relation for the species. The physical meaning of these vectors is not set in stone, but one interpretation is that they represent the number of a type of atom in each species. It then becomes important to know how to find these non-negative, non-zero (or semi-positive) left nullspace vectors.

2 Finding Semi-Positive Left Nullspace Vectors

We present an algorithm adapted by Schuster and Höfer in 1991 [SH91] from a Russian source [Che68] and a German source [FNB74], both of which are exceedingly difficult to find. We also verify that it does what it says it does.

We work in a finite-dimensional real vector space, to fit the situation. Recall that a *convex cone* is a set C which is closed under non-negative linear combinations, i.e. if $x, y \in C, \alpha, \beta \in \mathbb{R}_{\geq 0}$, then $\alpha x + \beta y \in C$. We hereby refer to convex cones as simply cones. One example of a cone is what is called an *orthant*, which is simply the generalization of a quadrant in the two-dimensional real plane, i.e. one of the 2^n segments of $\mathbb{R}^n$ where all the vectors in the orthant maintain their sign pattern as they move through the orthant. Clearly this property is preserved under non-negative linear combinations, so the orthant is a cone.

A cone then can be expressed in terms of a non-negative linear span of specific vectors in the cone; these vectors are called *extreme rays*, and they *generate* the cone, similar to the way a basis generates a subspace. We define a set of vectors to be *conically independent* in exactly the same way that a set of vectors is linearly independent, but with non-negative scalars. Considering cones is natural, because we desire that our solutions to remain within restrictions, such as the fact that all concentrations of species are non-negative, yet still possibly evolving over time in an unbounded fashion.

Finally, for a matrix A, denote $A_{|,j}$ and $A_{i,-}$ as the j^{th} column of A and the i^{th} row of A. For $n \in \mathbb{N}$, let $[n] = \{1, 2, ..., n\}$. For two vectors $x, y \in \mathbb{R}^n$, let $x \bullet y$ be the standard inner product of x and y, and for scalars a, b, let $a \cdot b$ be usual multiplication. This notation is used to be more clear.

Let $\Gamma = \Gamma^{(0)}$ be the stoichiometric matrix of a CRN. We define $T^{(0)}$ to be the *tableau* given by

$$T^{(0)} = \left[\, \Gamma^{(0)} \mid I_n \, \right].$$

The idea here is to take the non-negative orthant, which is a cone, and intersect it successively with each of the subspaces $\text{null}((\Gamma_{|,l})^T)$, each time obtaining a new cone. We construct vectors in each intersection from pairs of generating vectors from the previous cone, which will appear on the right-hand side of the tableau.

The inductive process goes as follows: Denote the i^{th} tableau as

$$T^{(i)} = \left[\, \Gamma^{(i)} \mid Z^{(i)} \, \right]. \tag{2.1}$$

For each tableau, let S_m be the set

$$S_m = \{a \in [n] \mid Z^{(i)}_{m,a} = 0\}. \tag{2.2}$$

If the j^{th} row of $T^{(i)}$ satisfies

$$\Gamma^{(i)}_{j,i+1} = 0, \tag{2.3}$$

then let

$$T^{(i+1)}_{l,-} = T^{(i)}_{j,-}.$$

If the j^{th} and k^{th} rows of $T^{(i)}$ satisfy

$$\Gamma^{(i)}_{j,i+1} \cdot \Gamma^{(i)}_{k,i+1} < 0, \tag{2.4}$$

$$S_j \cup S_k \nsubseteq S_m, \ \forall m \neq i, j, \tag{2.5}$$

then let a row in $T^{(i+1)}$ be given by

$$T^{(i+1)}_{l,-} = |\Gamma^{(i)}_{j,i+1}| \cdot T^{(i)}_{i,-} + |\Gamma^{(i)}_{i,i+1}| \cdot T^{(i)}_{j,-}.$$

Note that $T^{(i+1)}_{k,i+1} = 0$. This row creation is done as many times as is possible, and the resulting tableau is $T^{(i+1)}$. Condition (2.5) yields conically independent vectors on the right-hand side of the tableau, in the rows of $Z^{(i)}$. In many cases, there will be fewer rows in the new tableau than in the old tableau. If the new tableau is empty because there are no possible combinations, then the algorithm terminates and there are no conservation relations for the system. Finally, the algorithm has completed when $\Gamma^{(r)}$ is reached, which will be an p-by-r matrix of zeroes. Then, the generators/extreme rays of the cone are the p rows of $Z^{(r)}$.

We now spend some time to actually prove that this is the case.

Claim 2.1. The resulting vectors (the p rows of $Z^{(r)}$) are contained in $\text{null}(\Gamma^T)$. Furthermore, they generate the cone $\text{null}(\Gamma^T) \cap \mathbb{R}^n_{\geq 0}$.

Proof. We proceed by induction, as alluded to above.

Base case: The rows of I_n are $e_j, j \in [n]$, and we note that any possible non-negative combination remains in the non-negative orthant. If $\Gamma^{(0)}_{j,1} = 0$, $e_j \in \text{null}(\Gamma^T_{|,1})$, and moreover, since e_j only has one non-zero coordinate and will independent from any other vectors obtained, it is clearly a generator for the cone $\text{null}(\Gamma^T_{|,1}) \cap \mathbb{R}^n_{\geq 0}$. Then, if for two indices j, k conditions (2.4), (2.5) are satisfied, the resulting vector $Z^{(1)}_{l,-}$ from above is indeed in $\text{null}(\Gamma^T_{|,1})$, since we have

$$\begin{aligned} Z^{(1)}_{l,-} \bullet \Gamma_{|,1} &= (|\Gamma^{(i)}_{j,1}| \cdot \Gamma^{(i)}_{k,-} + |\Gamma^{(i)}_{k,1}| \cdot \Gamma^{(i)}_{j,-}) \bullet \Gamma_{|,1} \\ &= |\Gamma^{(i)}_{j,1}| \cdot \Gamma^{(i)}_{k,1} + |\Gamma^{(i)}_{k,1}| \cdot \Gamma^{(i)}_{j,1} \\ &= 0, \end{aligned}$$

seeing as we chose the row combination explicitly for this. As well, the vector will be a generator for the cone, since the only vectors with the same sign pattern also in the cone are multiples. Else, by taking differences, we could obtain a vector with only one non-zero component in the nullspace, which in the current case is impossible.

In general: We have the i^{th} tableau $T^{(i)}$, non-empty, and that the vectors $Z^{(i)}_{j,-}$ generate the cone $\cap_{s=1}^{i}\text{null}((\Gamma_{|,s})^T) \cap \mathbb{R}^n_{\geq 0}$. We construct $T^{(i+1)}_{k,-}$ by either one of the methods in the algorithm, and note that, for example,

$$\begin{aligned}
0 = T^{(i+1)}_{k,i+1} &= \Gamma^{(i+1)}_{k,i+1} \\
&= |\Gamma^{(i)}_{j,i+1}| \cdot \Gamma^{(i)}_{k,i+1} + |\Gamma^{(i)}_{k,1}| \cdot \Gamma^{(i)}_{j,i+1} \\
&= |\Gamma^{(i)}_{j,i+1}| \cdot (|\Gamma^{(i-1)}_{j_1,i+1}| \cdot \Gamma^{(i-1)}_{k_1,i+1} + |\Gamma^{(i-1)}_{k_1,1}| \cdot \Gamma^{(i-1)}_{j_1,i+1}) + \dots \\
&\vdots \\
&= \sum_{s=1}^{n} Z^{(i+1)}_{k,s} \Gamma_{s,i+1} = Z^{(i+1)}_{k,-} \bullet \Gamma_{|,i+1},
\end{aligned}$$

since the entries in $Z^{(i+1)}$ are exactly the factors involved in the row scaling and combinations. In the other case, it can trace back to a situation like above, or the vector is an e_j. Hence in all cases, $Z^{(i+1)}_{k,-} \in \text{null}((\Gamma_{|,i+1})^T)$, as desired.

Now, suppose those vectors do not generate the whole cone. Let v be one of the other generators. Since v must be in the previous cone, we can write it

$$v = \sum_k v_k Z^{(i)}_{k,-},\ v_k \geq 0\ \forall\ k.$$

Then, note that we can find a non-negative combination of $Z^{(i+1)}_{j,-}$'s such that it can be written as a non-negative combination of the same extreme rays as v, with potentially different coefficients. As before, taking scaled differences can now take us out of the nullspace, which is a contradiction. Thus v cannot be a generator for the cone. Therefore, by induction, our claim is proven. □

It should be noted that any resulting conservation vectors will have integer entries, since $\Gamma \in M_{n\times r}(\mathbb{Z})$, and we are performing non-negative combinations with integer scalings. This is important in the following section. As well, taking a combination of all the vectors, say just adding all of them together, yields a strictly positive vector in the cone, which gives us the situation as described at the end of section 1. This happens because for each coordinate index, at least one generating vector will have a non-zero entry in that coordinate, because we formed vectors by taking non-negative combinations of the standard basis vectors.

3 Atomic Realizations

With the algorithm under our belts now, we consider the atomic implications of having conservation relations for the system. As stated in the introduction, usually we deal with atom-free stoichiometries. [ET89] is an occurrence in the literature where atomic stoichiometry is considered, though it's only for a couple of sections in the book; they then move on to atom-free stoichiometry. However, we can use the convex basis of conservation vectors to identify atomic balances between the species, and thus identify the structure of said species.

Specifically, let $Z^{(r)}$ be the matrix consisting of the conservation vectors as rows, satisfying the condition that all columns are distinct. Then, to each row assign a distinct element, and to each column

assign a species. Reading the column then yields the possible atomic structures of that species. This is an *atomic realization* of the system. The stated condition means that no two species have the same atomic structure. Hence, we are excluding isomers from this discussion, since we want to look at atomic structure and composition only, and isomers fog the picture.

We note that we can pick and choose of the conservation relations to get a realization, and in particular, there exist realizations with a minimum number of distinct elements. This is important, because we can then hypothesize, for a given reaction system with species of unknown compositions, what those compositions might be, and decide on a reasonable atomic realization for the model. One thing to note is that it could be that we would have a matrix with distinct columns, but we could take a subset of the rows and end up with non-distinct columns. Thus systems can have minimal requirements as to the complexity of the atomic compositions of the species.

Keeping in mind that the cone has non-empty interior, we could add those two vectors together to get a positive vector, and presumably the one-element atomic relation would be the simplest realization possible, if the vector had distinct entries. This is correct, but ultimately less than satisfactory. Here, this is better interpreted as conservation of atoms/etc. Good judgement should be used when performing analysis, as always.

This discussion differs from the discussion found in Famili and Palsson [FP03], for instance, because they already have structure to their species, and they can tailor their interpretations to the structure. Here, we have minimal structure knowledge, and thus we cannot guess. Given information, of course, we can interpret the relations differently.

We provide an example to illustrate. Consider the system

$$\begin{aligned} A + 2B &\longrightarrow C + D \\ B + C &\longrightarrow A + E. \end{aligned}$$

Performing the algorithm yields, in order:

$$T^{(0)} = \left[\begin{array}{cc|ccccc} -1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$T^{(1)} = \left[\begin{array}{cc|ccccc} 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & -3 & 0 & 1 & 2 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$T^{(2)} = \left[\begin{array}{cc|ccccc} 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 & 0 & 3 \end{array}\right].$$

We observe that the four conservation relations do have distinct columns, so the system has an atomic realization involving four different elements, namely:

$$A = WX, \quad B = XYZ, \quad C = WZ_2, \tag{3.1}$$

$$D = X_3Y_2, \quad E = YZ_3. \tag{3.2}$$

We also note that any three of the four relations taken together have distinct columns as well, but that the majority of ways to take two of the four together either yield a column of zeroes (that is, a species has no

atomic structure at all), or have non-distinct columns. However, taking rows two and four give a possible two-element realization:

$$A = X, \quad B = XY, \quad C = Y_2, \tag{3.3}$$

$$D = X_3, \quad E = Y_3. \tag{3.4}$$

This is a minimal atomic realization, and provides the simplest structure hypothesis. Finally, note that convex integer combinations could possibly be an atomic realization, but taking the generating vectors of the cone capture the idea that atoms are indivisible, as opposed to functional groups.

4 Conclusion

We reviewed the idea of conservation relations in CRN theory, and proved that a given algorithm for finding them works effectively. We then introduced the idea of atomic realizations, in the hopes that this provides yet another tool in the analysis of CRNs, by restricting the possible atomic structures available for each species.

5 Acknowledgements

This note was partially funded by an NSERC Undergraduate Student Research Award. Much thanks also goes to my supervisor, Professor David Siegel, who proposed this topic and asked important questions about what was actually happening.

References

[Che68] S. Chernikov, *Linear inequalities (in russian)*, Nauka, Moscow, 1968.

[ET89] P. Érdi and J. Tóth, *Mathematical models of chemical reactions*, Manchester University Press, Manchester, 1989.

[FNB74] J. Guddat F. Nožička, H. Hollatz and B. Bank, *Theorie der linearen parametrischen optimierung*, Akademie-Verlag, Berlin, 1974.

[FP03] I. Famili and B. Palsson, *The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools*, Biophysical Journal (2003).

[HJ72] F. Horn and R. Jackson, *General mass action kinetics*, Arch. Rational Mech. Anal. (1972).

[SH91] S. Schuster and T. Höfer, *Determining all extreme semi-positive conservation relations in chemical reaction systems: A test criterion for conservativity*, J. Chem. Soc. Faraday Trans (1991).