bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Some remarks on moments for stochastic chemical kinetics

Eduardo D. Sontag*and Abhyudai Singh†

July 10, 2015

Abstract

We analyze a class of chemical reaction networks for which all moments can be computed by finite-dimensional linear differential equations. This class allows second and higher order reactions, but only under special assumptions on structure and/or conservation laws.

1 Preliminaries on stochastic chemical kinetics

We start by reviewing standard concepts regarding master equations for biochemical networks, see for instance [3].

Chemical systems are inherently stochastic, as reactions depend on random (thermal) motion. Deterministic models represent an aggregate behavior of the system. They are accurate in much of classical chemistry, where the numbers of molecules are usually expressed in multiples of Avogadro's number, which is $\approx 6 \times 10^{23}$. In such cases, basically by the law of large numbers, the mean behavior is a good description of the system. The main advantage of deterministic models is that they are comparatively easier to study than probabilistic ones. However, they may be inadequate when the "copy numbers" of species, i.e. the numbers of units (ions, atoms, molecules, individuals) are very small, as is often the case in molecular biology when looking at single cells: copy numbers are small for genes (usually one or a few copies), mRNA's (in the tens), ribosomes and RNA polymerases (up to hundreds) and certain proteins may be at low abundances as well. Analogous situations arise in other areas, such as the modeling of epidemics (where the "species" are individuals in various classes), if populations are small. This motivates the study of stochastic models. We assume that temperature and volume Ω are constant, and the system is well-mixed.

We consider a chemical reaction network consisting of m reactions which involve the n species

$$S_i,\ i \in \{1, 2, \ldots n\}\,.$$

The reactions $\mathcal{R}_j$, $j \in \{1, 2, \ldots, m\}$ are specified by combinations of reactants and products:

$$\mathcal{R}_j : \quad \sum_{i=1}^{n} a_{ij} S_i \to \sum_{i=1}^{n} b_{ij} S_i \tag{1}$$

*Rutgers University Mathematics Department and Center for Quantitative Biology. Supported in part by grants AFOSR FA9550-14-1-0060 and ONR 5710003367

†University of Delaware Department of Electrical and Computer Engineering

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where the a_{ij} and b_{ij} are non-negative integers, the *stoichiometry coefficients*, and the sums are understood informally, indicating combinations of elements.

The data in (1) serves to specify the stoichiometry of the network. The $n \times m$ *stoichiometry matrix* $\Gamma = \{\gamma_{ij}\}$ has entries:

$$\gamma_{ij} \;=\; b_{ij} - a_{ij}\,, \quad i = 1, \ldots, n\,, \quad j = 1, \ldots, m\,. \tag{2}$$

Thus, γ_{ij} counts the net change in the number of units of species S_i each time that reaction $\mathcal{R}_j$ takes place. We will denote by γ_j the jth column of Γ:

$$\gamma_j = b_j - a_j$$

where

$$a_j = (a_{1j}, \ldots, a_{nj})' \quad \text{and} \quad b_j = (b_{1j}, \ldots, b_{nj})'$$

(prime indicates transpose) and assume that no $\gamma_j = 0$ (that is, every reaction changes at least one species).

In general, for every $v \in \mathbb{Z}^n_{\geq 0}$, we denote $\oplus v = \oplus(v_1, \ldots, v_n) := v_1 + \ldots + v_n$. In particular, for each $j \in \{1, \ldots, m\}$, we define the *order* of reaction $\mathcal{R}_j$ as

$$A_j \;=\; \oplus a_j \;=\; \sum_{i=1}^{n} a_{ij}$$

(the total number of units of all species participating in the reaction). One allows the possibility of zero order, that is, for some reactions j, $a_{ij} = 0$ for all i. This is the case when there is "birth" of species out of the blue, or more precisely, a species is created by what biologists call a "constitutive" process, such as the production of an mRNA molecule by a gene that is always active. Zeroth order reactions may also be used to represent inflows to a system from its environment. Similarly, also allowed is the possibility that, for some reactions j, $b_{ij} = 0$ for all i. This is the case for reactions that involve degradation, dilution, decay, or outflows.

Stoichiometry information is not sufficient, by itself, to completely characterize the behavior of the network: one must also specify the *rates* at which the various reactions take place. This can be done by specifying "propensity" or "intensity" functions.

1.1 Stochastic models of chemical reactions

Stochastic models of chemical reaction networks are described by a column-vector Markov stochastic process $X = (X_1, \ldots, X_n)'$ which is indexed by time $t \geq 0$ and takes values in $\mathbb{Z}^n_{\geq 0}$. Thus, $X(t)$ is a $\mathbb{Z}^n_{\geq 0}$-valued random variable, for each $t \geq 0$. Abusing notation, we also write $X(t)$ to represent an outcome of this random variable on a realization of the process. The interpretation is:

$$X_i(t) \;=\; \text{number of units of species } i \text{ at time } t\,.$$

One is interested in computing the probability that, at time t, there are k_1 units of species 1, k_2 units of species 2, k_3 units of species 3, and so forth:

$$p_k(t) = \mathbb{P}\left[X(t) = k\right]$$

for each $k \in \mathbb{Z}^n_{\geq 0}$. We call the vector k the *state* of the process at time t.

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Arranging the collection of all the $p_k(t)$'s into an infinite-dimensional vector, after an arbitrary order has been imposed on the integer lattice $\mathbb{Z}^n_{\geq 0}$, we have that $p(t) = (p_k)_{k\in\mathbb{Z}^n_{\geq 0}}$ is the discrete probability density (also called the "probability mass function") of $X(t)$.

This note is concerned with the computation of moments, such as the expectation or mean (i.e, the average over all possible random outcomes) of the numbers of units of species at time t:

$$\mathbb{E}\left[X(t)\right] = \sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)k$$

which is a column vector whose entries are the means

$$\mathbb{E}\left[X_i(t)\right] \;=\; \sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)k_i \;=\; \sum_{\ell=0}^{\infty}\ell \sum_{\{k\in\mathbb{Z}^n_{\geq 0},k_i=\ell\}} p_k(t) \;=\; \sum_{\ell=0}^{\infty}\ell\, p_\ell^{(i)}(t)$$

of the $X_i(t)$'s, where the vector $(p_0^{(i)}(t), p_1^{(i)}(t), p_2^{(i)}(t), \ldots)$ is the marginal density of $X_i(t)$. Also of interest, to understand variability, are the matrix of second moments at time t:

$$\mathbb{E}\left[X(t)X(t)'\right]$$

whose (i,j)th entry is $\mathbb{E}\left[X_i(t)X_j(t)\right]$ and the (co)variance matrix at time t:

$$\mathbb{V}\mathrm{ar}\left[X(t)\right] \;=\; \mathbb{E}\left[\left(X(t)-\mathbb{E}\left[X(t)\right]\right)\left(X(t)-\mathbb{E}\left[X(t)\right]\right)'\right] \;=\; \mathbb{E}\left[X(t)X(t)'\right] - \mathbb{E}\left[X(t)\right]\mathbb{E}\left[X(t)\right]'$$

whose (i,j)th entry is the covariance of $X_i(t)$ and $X_j(t)$, $\mathbb{E}\left[X_i(t)X_j(t)\right] - \mathbb{E}\left[X_i(t)\right]\mathbb{E}\left[X_j(t)\right]$.

1.2 The Chemical Master Equation

A *Chemical Master Equation (CME)* (also known as a *Kolmogorov forward equation*) is a system of linear differential equations for the p_k's, of the following form. Suppose given m functions

$$\rho_j : \mathbb{Z}^n_{\geq 0} \to \mathbb{R}_{\geq 0}\,, \quad j = 1,\ldots,m\,, \quad \text{with } \rho_j(0) = 0\,.$$

These are the *propensity functions* for the respective reactions $\mathcal{R}_j$. An intuitive interpretation is that $\rho_j(k)dt$ is the probability that reaction $\mathcal{R}_j$ takes place, in a short interval of length dt, provided that the state was k at the beginning of the interval. The CME is:

$$\boxed{\frac{dp_k}{dt} \;=\; \sum_{j=1}^{m}\rho_j(k-\gamma_j)\,p_{k-\gamma_j} \;-\; \sum_{j=1}^{m}\rho_j(k)\,p_k\,, \quad k\in\mathbb{Z}^n_{\geq 0}} \tag{3}$$

where, for notational simplicity, we omitted the time argument "t" from p, and where we make the convention that $\rho_j(k-\gamma_j) = 0$ unless $k \geq \gamma_j$ (coordinatewise inequality). There is one equation for each $k \in \mathbb{Z}^n_{\geq 0}$, so this is an infinite system of linked equations. When discussing the CME, we will assume that an initial probability vector $p(0)$ has been specified, and that there is a unique solution of (3) defined for all $t \geq 0$. A different CME results for each choice of propensity functions, a choice that is dictated by physical chemistry considerations. Here we will restrict attention to the most standard model, mass-action kinetics propensities.

We will also introduce the n-column vector:

$$f(k) \;:=\; \sum_{j=1}^{m}\rho_j(k)\,\gamma_j \;=\; \Gamma R(k) \qquad\qquad k\in\mathbb{Z}^n_{\geq 0}$$

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where $R(k) = (\rho_1(k), \ldots, \rho_m(k))'$. One may interpret $f(k)dt$ as the expected change of state during an interval (since γ_j quantifies the size of the jump if the reaction is $\mathcal{R}_j$). Thus, $f(k)$ may be thought of as the rate of change of the state, if the state is k.

1.3 Propensity functions for mass-action kinetics

For each $k = (k_1, \ldots, k_n)' \in \mathbb{Z}^n_{\geq 0}$, we let (recall that a_j denotes the vector $(a_{1j}, \ldots, a_{nj})'$):

$$\binom{k}{a_j} = \prod_{i=1}^{n} \binom{k_i}{a_{ij}}$$

where $\binom{k_i}{a_{ij}}$ is the usual combinatorial number $k_i!/(k_i - a_{ij})!a_{ij}!$, which we define to be zero if $k_i < a_{ij}$.

The most commonly used propensity functions, and the ones best-justified from elementary physical principles, are *ideal mass action kinetics* propensities, defined as follows:

$$\rho_j(k) \;=\; \kappa_{a_j} \binom{k}{a_j}, \;\; j = 1, \ldots, m\,.$$

The m non-negative constants $\kappa_1, \ldots, \kappa_m$ are arbitrary, and they represent quantities related to the volume, shapes of the reactants, chemical and physical information, and temperature. Notice that $\rho_j(k)$ can be expanded into a polynomial in which each variable k_i has an exponent less or equal to a_{ij}. In other words,

$$\rho_j(k) \;=\; \sum_{c_j \leq a_j} \kappa_{c_j}\, k^{c_j} \tag{4}$$

("$\leq$" is understood coordinatewise, and by definition $k^{c_j} = k_1^{c_{1j}} \ldots k_n^{c_{nj}}$ and $r^0 = 1$ for all integers), for suitably redefined coefficients κ_{c_j}'s. Often one uses the simplification

$$\rho_j(k) \;=\; \kappa_j\, k^{a_j}, \;\; j = 1, \ldots, m\,. \tag{5}$$

This simplification means that we approximate $x(x-1)\ldots(x-r+1) \approx x^r$, which introduces a small error if $r > 1$ and the integer x is very small. We provide results for both forms of propensity.

1.4 Moment equations

Suppose given a function $M : \mathbb{Z}^n_{\geq 0} \to \mathbb{R}$ (to be taken as a monomial when computing moments). The definition of expectation gives:

$$\mathbb{E}\left[M(X(t))\right] \;=\; \sum_{k \in \mathbb{Z}^n_{\geq 0}} p_k(t)\, M(k)$$

because $\mathbb{P}\left[X(t) = k\right] = p_k(t)$. We have:

$$\frac{d}{dt}\mathbb{E}\left[M(X(t))\right] \;=\; \sum_{k \in \mathbb{Z}^n_{\geq 0}} \frac{dp_k}{dt}(t) M(k) \;=\; \sum_{k \in \mathbb{Z}^n_{\geq 0}} \left(\sum_{j=1}^{m} \rho_j(k - \gamma_j)\, p_{k-\gamma_j} - \sum_{j=1}^{m} \rho_j(k)\, p_k \right) M(k)\,.$$

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Note this equality, for each fixed j:

$$\sum_{k\in\mathbb{Z}^n_{\geq 0}} p_{k-\gamma_j}(t)\rho_j(k-\gamma_j)M(k) \;=\; \sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)\rho_j(k)M(k+\gamma_j)$$

(by definition, $\rho_j(k-\gamma_j)=0$ unless $k\geq\gamma_j$, so one may perform a change of variables $\widetilde{k}=k-\gamma_j$). There results:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[M(X(t))\right] &= \sum_{k,j} p_k(t)\rho_j(k)M(k+\gamma_j) \;-\; \sum_{k,j} p_k(t)\rho_j(k)M(k)\\ &= \sum_{k\in\mathbb{Z}^n_{\geq 0}} p_k(t)\sum_{j=1}^m \rho_j(k)\left[M(k+\gamma_j)-M(k)\right]\,.\end{aligned}$$

Let us define, for any $\gamma\in\mathbb{Z}^n_{\geq 0}$, the new function $\Delta_\gamma M$ given by $(\Delta_\gamma M)(k):=M(k+\gamma)-M(k)$. With these notations,

$$\frac{d}{dt}\mathbb{E}\left[M(X(t))\right] \;=\; \sum_{j=1}^m \mathbb{E}\left[\rho_j(X(t))\,\Delta_{\gamma_j}M(X(t))\right]\,. \tag{6}$$

We next specialize to a monomial function:

$$M(k) \;=\; k^u \;=\; k_1^{u_1}k_2^{u_2}\ldots k_n^{u_n}$$

where $u\in\mathbb{Z}^n_{\geq 0}$. In this case,

$$\begin{aligned}(\Delta_{\gamma_j}M)(k) &= M(k+\gamma_j)-M(k)\\ &= (k_1+\gamma_{1j})^{u_1}(k_2+\gamma_{2j})^{u_2}\ldots(k_n+\gamma_{nj})^{u_n}-k_1^{u_1}k_2^{u_2}\ldots k_n^{u_n}\,.\end{aligned}$$

Expanding each binomial $(k_i+\gamma_{ij})^{u_i}$ for which $\gamma_{ij}\neq 0$, distributing, and canceling out the leading term, results in

$$(\Delta_{\gamma_j}M)(k) \;=\; \sum_{\nu\in\mathcal{S}(u,j)} d_\nu k^\nu$$

for appropriate coefficients d_ν, where

$$\mathcal{S}(u,j) \;:=\; \text{set of indices } \nu=u-\mu\in\mathbb{Z}^n_{\geq 0}$$

where $\mu\in\mathbb{Z}^n_{\geq 0}$ satisfies:

$$u \;\geq\; \mu \;\neq\; 0 \tag{7}$$

(inequalities "$\geq$" in $\mathbb{Z}^n_{\geq 0}$ are understood coordinatewise), and

$$\mu_i=0 \text{ for each } i \text{ such that } \gamma_{ij}=0\,. \tag{8}$$

Thus, if $\rho_j(k)=\kappa_j\,k^{a_j}$ as in (5), then:

$$\frac{d}{dt}\mathbb{E}\left[X(t)^u\right] \;=\; \sum_{j=1}^m\sum_{\nu\in\mathcal{S}(u,j)} d_\nu\kappa_j\,\mathbb{E}\left[X(t)^{\nu+a_j}\right]\,. \tag{9}$$

More generally, if

$$\rho_j(k) \;=\; \sum_{c_j\leq a_j}\kappa_{c_j}\,k^{c_j}$$

as in (4) then:

$$\frac{d}{dt}\mathbb{E}\left[X(t)^u\right] \;=\; \sum_{j=1}^m\sum_{c_j\leq a_j}\sum_{\nu\in\mathcal{S}(u,j)} d_\nu\kappa_{c_j}\,\mathbb{E}\left[X(t)^{\nu+c_j}\right]\,. \tag{10}$$

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2 A set of transformations

For each multi-index $u \in \mathbb{Z}^n_{\geq 0}$, we define $\mathcal{R}^0(u) = \{u\}$,

$$\mathcal{R}^1(u) \;:=\; \{\nu + a_j\,,\; 1 \leq j \leq m,\, \nu \in \mathcal{S}(u,j)\}$$

if using definition (5), or

$$\mathcal{R}^1(u) \;:=\; \{\nu + c_j\,,\; 1 \leq j \leq m, c_j \leq a_j,\, \nu \in \mathcal{S}(u,j)\}$$

if using definition (4), and, more generally, for any $\ell \geq 1$,

$$\mathcal{R}^{\ell+1}(u) \;:=\; \mathcal{R}^1(\mathcal{R}^\ell(u))$$

where, for any set U, $R^\ell(U) := \bigcup_{u \in U} \mathcal{R}^\ell(u)$. Finally, we set

$$\mathcal{R}(u) \;:=\; \bigcup_{i=0}^{\infty} \mathcal{R}^i(u)\,.$$

Each set $\mathcal{R}^\ell(u)$ is finite, but the cardinality $\#(\mathcal{R}(u))$ may be infinite. It is finite if and only if there is some $L \geq 0$ such that $\mathcal{R}(u) = \bigcup_{i=0}^{L} \mathcal{R}^i(u)$, or equivalently $\mathcal{R}^{L+1}(u) \subseteq \bigcup_{i=0}^{L} \mathcal{R}^i(u)$.

Equation (9) (or (10)) says that the derivative of the u-th moment can be expressed as a linear combination of the moments in the set $\mathcal{R}^1(u)$. The derivatives of these moments, in turn, can be expressed in terms of the moments in the set $\mathcal{R}^1(u')$, for each $u' \in R^1(u)$, i.e., in terms of moments in the set $\mathcal{R}^2(u)$. Iterating, we have the following:

Main Lemma. Suppose that $N := \#(\mathcal{R}(u)) < \infty$, and write $\mathcal{R}(u) = \{u_1, \ldots, u_N\}$, with $u_1 = u$. Then, there exists a matrix $A \in \mathbb{R}^{N \times N}$ such that, for any sample path $X(\cdot)$, and letting

$$x(t) := \left(\mathbb{E}\left[X^{u_1}(t)\right], \ldots, \mathbb{E}\left[X^{u_N}(t)\right]\right)'\,,$$

it holds that $\dot{x}(t) = Ax(t)$ for all $t \geq 0$.

This motivates the following problem: *characterize those chemical reaction networks for which* $\#(\mathcal{R}(u)) < \infty$ *for all* $u \in \mathbb{Z}^n_{\geq 0}$.

One trivial sufficient condition is that all reactions be or order 0 or 1, which means that $\oplus a_j \in \{0, 1\}$. In that case, since $\mu > 0$ in the definition of $\mathcal{S}(u,j)$, it follows that $\oplus a_j \leq \oplus \mu$ for every index j. Therefore, $\oplus(\nu + a_j) = \oplus u + \oplus a_j - \oplus \mu \leq \oplus u$ for all u, and the same holds for $\nu + c_j$ if $c_j \leq a_j$. In summary, the degree of all elements in $\mathcal{R}(u)$ is $\leq \oplus u$, so indeed $\#(\mathcal{R}(u)) < \infty$. A generalization to "weighted L^1 norms" where one uses instead $\oplus v = \beta_1 v_1 + \ldots + \beta_n v_n$ for non-unity coefficients will be described next.

2.1 Lyapunov-like functions

Definition. A function $V : \mathbb{Z}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ will be called a Lyapunov-like function with respect to a given chemical network if the following two properties hold:

1. for each u, v: $v \in \mathcal{R}^1(u) \;\Rightarrow\; V(v) \leq V(u)$ [nondecreasing property],
2. for each $\alpha \geq 0$: $V_\alpha := \{v \,|\, V(v) \leq \alpha\}$ is finite [properness].

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Theorem. For every chemical network, the following two statements are equivalent:

- There exists a Lyapunov-like function.
- $\#(\mathcal{R}(u)) < \infty$ for all $u \in \mathbb{Z}^n_{\geq 0}$.

Proof. Sufficiency is clear: pick any u, and let $\alpha := V(u)$; iterating on the nondecreasing property, $V(v) \leq \alpha$ for all $v \in \mathcal{R}(u)$, meaning that $\mathcal{R}(u) \subseteq V_\alpha$, and thus $\#(\mathcal{R}(u)) < \infty$.

To prove the converse, assume that $\#(\mathcal{R}(u)) < \infty$ for all u. Define $V(u) := \max_{w \in \mathcal{R}(u)} \oplus w$. Since $\#(\mathcal{R}(u)) < \infty$, it follows that $V(u) < \infty$. As $u \in \mathcal{R}(u)$, it follows from the definition of V that $\oplus u \leq V(u)$. Now pick any u, v so that $v \in \mathcal{R}^1(u)$. Since $\mathcal{R}(v) \subseteq \mathcal{R}(u)$, it follows that $\{\oplus w, w \in \mathcal{R}(v)\} \subseteq \{\oplus w, w \in \mathcal{R}(u)\}$. Therefore $V(v) \leq V(u)$ (nondecreasing property). Now pick any $\alpha \geq 0$, which we may take without loss of generality to be a nonnegative integer, and any element $v \in V_\alpha$. Since $\oplus v \leq V(v)$, it follows that $\oplus v \leq \alpha$. So V_α is a subset of the set of all nonnegative vectors v such that $\oplus v \leq \alpha$, which has $\binom{\alpha+n}{n}$ elements. ∎

2.2 Linear V's

The nonincrease requirement means, using the definition of $\mathcal{R}^1(u)$, that

$$V(u - \mu + c_j) \leq V(u) \tag{11}$$

for all $c_j \leq a_j$ $(1 \leq j \leq m)$ under definition (4) for propensities, or just for $c_j = a_j$ if propensities have the simplified form (5), and every μ for which (7) and (8) hold, i.e., every μ for which $u \geq \mu \neq 0$ and $\mu_i = 0$ for every i such that $\gamma_{ij} = 0$. Pick any reaction index j and for this index pick any species index i such that the species S_i changes, that is, $\gamma_{ij} \neq 0$. Now pick $u = \mu = e_i$, the canonical unit vector with a "1" in the ith position (this choice of μ is allowed, since it is false that $\gamma_{ij} = 0$) and apply (11). that a necessary condition for decrease is that

$$\gamma_{ij} \neq 0 \;\Rightarrow\; V(c_j) \leq V(e_i) \tag{12}$$

for all $c_j \leq a_j$, or simply for $c_j = a_j$ in the simplified case (5).

We now consider the special case of Lyapunov-like functions which can be extended to an additive map $V : \mathbb{Z}^n \to \mathbb{R}$. In this case, (12) equivalent to (11). Indeed, to see that (12) implies (11), pick any u, any reaction index j, and any μ such that (7) and (8) hold. Since $\mu \neq 0$ and $\mu \leq u$, there is some species index i such that $\gamma_{ij} \neq 0$ and $\mu_i \neq 0$, i.e., $\mu \geq e_i$. Applying (12) with this choice of i:

$$V(u - \mu + c_j) \;\leq\; V(u) + V(e_i - \mu) + V(c_j - e_i) \;\leq\; V(u)\,,$$

where we used that $V(\mu - e_i) \geq 0$.

A map $V : \mathbb{Z}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ that extends to an additive function $V : \mathbb{Z}^n \to \mathbb{R}$ is necessarily of the form

$$V(u) \;=\; \beta u \;=\; \beta_1 u_1 + \ldots + \beta_n u_n \tag{13}$$

for some $\beta = (\beta_1, \ldots, \beta_n) \in \mathbb{Z}^n_{\geq 0}$ and it automatically satisfies the properness property provided that all $\beta_i \neq 0$, which we assume from now on. Thus, $\#(\mathcal{R}(u)) < \infty$ will be satisfied for all u if V has this form and satisfies (12). This condition can be made a little more explicit in the linear case. Let $\Delta_j := \{i \,|\, \gamma_{ij} \neq 0\}$. Then a linear Lyapunov-like function amounts to picking a β such that

$$i \in \Delta_j \;\Rightarrow\; \beta a_j \leq \beta_i\,. \tag{14}$$

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.3 Special case: multi-layer feedforward networks

A general class for which there is a linear Lyapunov-like function, and hence $\#(\mathcal{R}(u)) < \infty$ for all u, is that of multi-layer feedforward networks with linear reactions in the first layer. These are defined as follows. We find it convenient to separate degradations from more general reactions. So we will assume that there are reactions $\mathcal{R}_j$, $j \in \{1, 2, \ldots, m\}$, which are partitioned into $p \geq 1$ layers: $\mathbf{R}_1, \ldots, \mathbf{R}_p$. Species S_i, $i \in \{1, 2, \ldots, n\}$, are also partitioned into p layers $\mathbf{S}_1, \ldots, \mathbf{S}_p$. In addition, we allow additional "pure degradation" reactions $D_j : S_{i_j} \to 0$, $j \in \{1, \ldots, d\}$ (so the total number of reactions is $m' = m + d$).

We assume that the reactions $\mathcal{R}_j$ that belong to the first layer $\mathbf{R}_1$ are all of order zero or one, i.e. they have $\oplus a_j \in \{0, 1\}$. (This first layer might model several independent separate chemical subnetworks; we collect them all as one larger network.) More generally, for reactions at any given layer π, the only species that appear as reactants in nonlinear reactions are those in layers $< \pi$ and the only ones that can change are those in layer π, that is:

$$\text{if } \mathcal{R}_j \in \mathbf{R}_\pi \ : \ \begin{cases} a_{ij} \neq 0 \text{ and } \oplus a_j > 1 & \Rightarrow \ S_i \in \bigcup_{1 \leq s < \pi} \mathbf{S}_\pi \\ \gamma_{ij} \neq 0 & \Rightarrow \ S_i \in \mathbf{S}_\pi \,. \end{cases} \tag{15}$$

This means that except for order zero or one reactions, every reaction $\mathcal{R}_j$ at layer $1 < \pi \leq p$ has the form:

$$a_{i_1 j} S_{i_1} + \ldots a_{i_q j} S_{i_q} \ \to \ a_{i_1 j} S_{i_1} + \ldots a_{i_q j} S_{i_q} + b_{i_{q+1} j} S_{i_{q+1}} + \ldots b_{i_{q+q'} j} S_{i_{q+q'}}$$

with $S_{i_1}, \ldots, S_{i_q}$ in layers $< \pi$ and $S_{i_{q+1}}, \ldots, S_{i_{q+q'}}$ in layer π.

We claim that there is a linear Lyapunov-like function for any such network, Note that, for a degradation reaction $D_j : S_{i_j} \to 0$, the entry γ_{ij} of the stoichiometry vector is nonzero (and equal to -1) only when $i = i_j$, and for this index we have $a_{ij} = 1$. Thus condition (14) simply requires $\beta_i \leq \beta_i$ and hence is automatically satisfied no matter what is the choice of β. Thus we may ignore degradations and assume from now on that only the reactions $\mathcal{R}_j$ are present. We prove the claim by induction on the number of layers p. If $p = 1$, all reactions have order 0 or 1, so we can take $\beta_i = 1$ for all i. Arrange the species indices so that $S_{r+1}, \ldots, S_n$ are the species in $\mathbf{S}_p$; these do not appear any reactions belonging to $\mathbf{R}_\pi$ for $\pi < p$. So layers $\mathbf{R}_\pi$ for $\pi < p$ and species in $\mathbf{S}_\pi$ for $\pi < p$ define a network with $p - 1$ layers, and we may assume by induction that a linear V_0 has been defined for that network. This means that we have a vector of positive numbers $\beta^0 = (\beta_1, \ldots, \beta_{n-r})$ such that (14) holds for this subnetwork, which means, for any extension to a vector $\beta = (\beta^0, \star)$ with n components (since the coefficients of a_j are zero for indices $r + 1, \ldots, n$) that $\beta a_j \leq \beta_i$ whenever $i \in \Delta_j$, when j, i index reactions and species in the first $p - 1$ layers.

So all that is needed is to define the additional coefficients β_i, $i \in \{r + 1, \ldots, n\}$, such that the inequality $\beta a_j \leq \beta_i$ holds for all pairs (i, j) such that (1) $\mathcal{R}_j \in \mathbf{R}_p$ or $S_i \in \mathbf{S}_p$ and (2) $\gamma_{ij} \neq 0$. We show that it suffices to pick all these β_i equal to a common value $\bar{\beta} :=$ maximum of $\beta^0 a_j^0$ over all reactions $\mathcal{R}_j \in \mathbf{R}_p$, where a_j^0 is the restriction of the vector a_j to its first r components.

If $\mathcal{R}_j \in \mathbf{R}_p$ and $S_i \notin \mathbf{S}_p$, the second condition in (15) (with $\pi = p$) says that (2) is not satisfied. Thus, we only need to consider $S_i \in \mathbf{S}_p$, i.e. $i \in \{r + 1, \ldots, n\}$. Suppose first that $\oplus a_j > 1$. The first condition in (15) (with $\pi = p$) insures that $a_{ij} = 0$ for all such i. Thus, $a_j = (a_j^0, 0)$ where the vector 0 has length $n - r$. it follows that $\beta a_j = \beta^0 a_j^0 \leq \bar{\beta}$. Next, suppose that $\oplus a_j \leq 1$. If $\oplus a_j = 0$, then $a_j = 0 \leq \bar{\beta}$. So assume $\oplus a_j = 1$ and pick the unique index i' such that $a_{i'j} = 1$. If $S_{i'} \in \mathbf{S}_\pi$, with $\pi < p$, then once again $a_{ij} = 0$ for all $i \in \{r + 1, \ldots, n\}$ and $\beta a_j = \beta^0 a_j^0 \leq \bar{\beta}$. Finally, assume that $a_j = e_i$ with $i \in \{r + 1, \ldots, n\}$. Now $\bar{\beta} = \beta a_j \leq \bar{\beta}$ is trivially satisfied. ∎

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.4 Examples

Let us start with the system shown in [1] to have moment closure:

$$\mathcal{R}_1 : 0 \to S_1\,, \;\; \mathcal{R}_2 : S_1 \to S_1 + S_2\,, \;\; \mathcal{R}_3 : S_1 + S_2 \to S_1 + S_2 + S_3\,,$$

$$D_1 : S_1 \to 0\,, \;\; D_2 : S_2 \to 0\,, \;\; D_3 : S_3 \to 0\,.$$

This is a three-layer system with one reaction in each layer, plus degradations. As we said, we may ignore degradations, so we consider:

$$a_1 = (0,0,0)',\; a_2 = (1,0,0)',\; a_3 = (1,1,0)',$$

and we have $\Delta_1 = \{1\}$, $\Delta_2 = \{2\}$, $\Delta_3 = \{3\}$. We must find a positive vector $\beta = (\beta_1, \beta_2, \beta_3)$ such that $\beta a_i \leq \beta_i$, $i \in \{1,2,3\}$, i.e., $\beta_1 \leq \beta_2$ and $\beta_1 + \beta_2 \leq \beta_3$. We may pick $\beta = (1,1,2)$.

Here is a more complicated example involving several reversible first order reactions as well as some dimeric and trimeric reactions:

$$\mathcal{R}_1 : S_1 \to S_2\,, \;\; \mathcal{R}_2 : S_2 \to S_1\,, \;\; \mathcal{R}_3 : S_3 \to S_4\,, \;\; \mathcal{R}_4 : S_4 \to S_3\,,$$

$$\mathcal{R}_5 : 2S_1 + S_2 + S_3 \to 2S_1 + S_2 + S_3 + S_5\,,$$

$$\mathcal{R}_6 : S_1 + 3S_5 \to S_1 + 3S_2 + S_6\,.$$

We have three layers: $\mathbf{R}_1 = \{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4\}$, $\mathbf{R}_2 = \{\mathcal{R}_5\}$, $\mathbf{R}_3 = \{\mathcal{R}_6\}$, and $\mathbf{S}_1 = \{S_1, S_2, S_3, S_4\}$, $\mathbf{S}_2 = \{S_5\}$, $\mathbf{S}_3 = \{S_6\}$. Using e_i to denote canonical unit vectors:

$$a_j = e_j,\; j \in \{1,2,3,4\}, a_5 = (2,1,1,0,0,0)',\; a_6 = (1,0,0,0,3,0)',$$

and

$$\Delta_1 = \Delta_2 = \{1,2\},\; \Delta_3 = \Delta_4 = \{3,4\},\; \Delta_5 = \{5\},\; \Delta_6 = \{6\}.$$

We must find a positive vector β such that $\beta a_1 \leq \beta_1$, $\beta a_1 \leq \beta_2$, $\beta a_2 \leq \beta_1$, $\beta a_2 \leq \beta_2$, $\beta a_3 \leq \beta_3$, $\beta a_3 \leq \beta_4$, $\beta a_4 \leq \beta_3$, $\beta a_4 \leq \beta_4$, $\beta a_5 \leq \beta_5$, $\beta a_6 \leq \beta_6$, i.e. so that $\beta_1 = \beta_2$, $\beta_3 = \beta_4$, $2\beta_1 + \beta_2 + \beta_3 \leq \beta_5$, and $\beta_1 + 3\beta_5 \leq \beta_6$. These constraints can be satisfied with $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$, $\beta_5 = 4$, $\beta_6 = 13$.

2.5 Special case: conserved variables

In some applications, one is interested in computing the moments $\mathbb{E}\left[X(t)^u\right]$ only for trajectories $X(t)$ which remain in some specified subset $\mathcal{C} \subseteq \mathbb{Z}^n_{\geq 0}$. When this subset has the form

$$\mathcal{C} \;=\; \mathcal{C}_1 \times \mathbb{Z}^{n_2}_{\geq 0} \;\subset\; \mathbb{Z}^{n_1}_{\geq 0} \times \mathbb{Z}^{n_2}_{\geq 0} \tag{16}$$

and the subset $\mathcal{C}_1$ is finite, the right-hand side of equation (10) (or (9)) can be simplified.

For example, suppose that the first two species S_1 and S_2 indicate the activity of a specified gene (inactive and active, respectively), with S_1 and S_2 reacting according to $S_1 \to S_2$, $S_2 \to S_1$, and no other reactions involve a change in S_1 and S_2. (This does not rule out reactions such as $S_1 \to S_1 + S_3$ which would model transcription from the active conformation, since such a reaction does not change S_1 nor S_2.) It is the case that $X_1(t) + X_2(t)$ remains constant in time, so $X_1(t) + X_2(t) = X_1(0) + X_2(0)$ for all t. Moreover, given the biological motivation for these

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

equations, we are only interested in the cases where $(X_1(0), X_2(0)) = (1,0)$ or $= (0,1)$. Thus, we have that $X_1(t) + X_2(t) = 1$ for all t. This restricts the components $(X_1(t), X_2(t))$ of $X(t)$ to take values in the finite set $C_1 = \{(1,0),(0,1)\}$, and hence all moments can be assumed to have the first two exponents equal to one:

$$\mathbb{E}\left[X_1(t)^{u_1} X_2(t)^{u_2} X_3^{u_3} \dots X_n^{u_n}\right] \;=\; \mathbb{E}\left[X_1(t) X_2(t) X_3^{u_3} \dots X_n^{u_n}\right] .$$

More generally, for any positive integers r and L, let $\mathcal{L}_{L,r} := \{0,\dots,L\}^r$. Then, for any finite subset $\mathcal{C} \subset \mathbb{Z}^r_{\geq 0}$, there is some integer L with the property:

$$\forall\, v \in \mathbb{Z}^r_{\geq 0} \;\; \exists \{p_d \in \mathbb{R}, d \in \mathcal{L}_{L,r}\} \quad \text{so that} \;\; k^v = \sum_{d \in \mathcal{L}_{L,r}} p_d k^d \text{ for all } k \in \mathcal{C} . \tag{17}$$

In other words, every monomial can be expressed as a linear combination of monomials with exponents $\leq L$.

To prove (17), observe that the set $\mathcal{F}$ of functions $\mathcal{C} \to \mathbb{R}$ is a finite-dimensional vector space (canonically identified with $\mathbb{R}^{\#(C)}$, where $\#(\mathcal{C})$ is the cardinality of $\mathcal{C}$). Introduce for each i the subspace $\mathcal{F}_{i,r}$ of $\mathcal{F}$ spanned by the monomial functions $k \mapsto k^d$, $d \in \mathcal{L}_{i,r}$. Since $\mathcal{F}_{0,r} \subseteq \mathcal{F}_{1,r} \subseteq \mathcal{F}_{2,r} \subseteq \dots$ is a nondecreasing sequence of subspaces, there is some L such that $F_{L',r} = F_{L,r}$ for all $L' > L$ (in fact, one may take $L = \#(C) - 1$) and this proves the result.

A lower exponent may suffice for a proper subset of $\{0,\dots,L\}^r$. For example, consider the set $\mathcal{C}_1 = \{(2,0),(1,1),(0,2)\} \subset \{0,1,2\}^2$. Then every monomial function on $\mathcal{C}_1$ is a linear combination of $f_1(x,y) = x$, $f_2(x,y) = xy$, and $f_3(x,y) = y$. For example, $g(x,y) = x^3$ can be written as $g = 4f_1 - 3f_2$, as can be verified by plugging-in the three elements of $\mathcal{C}_1$. More generally, for $r = 2$ and two species satisfying $X_1 + X_2 \equiv L$, we may write $X_1 = L - X_2$, and, using the inverse of the Vandermonde matrix of order L and evaluated at $\{0,\dots,L\}$, we may express X_2^i in terms of $1,\dots,X_2^L$, for any $i > L$.

Now given any set as in (16) with $\#(\mathcal{C}_1) < \infty$, we may apply the above observation to $\mathcal{C}_1$, and this means that all moments $X_1(t)^{u_1} X_2(t)^{u_2} X_3^{u_3} \dots X_n^{u_n}$ can be written as a linear combination of moments for which the first n_1 exponents are $\leq L$. The remaining reactions could be a feedforward network, and now moments are all determined by a finite set of linear differential equations, so long as we only care about initial conditions in a finite invariant set.

A simple example is as follows. We consider the following set of chemical reactions:

$$\mathcal{R}_1 : S_1 \xrightarrow{u} S_2 \,, \;\; \mathcal{R}_2 : S_2 \xrightarrow{k_2} S_1 \,, \;\; \mathcal{R}_3 : 0 \xrightarrow{k_3 u} S_3 \,, \;\; \mathcal{R}_4 : S_2 + S_3 \xrightarrow{k_4} S_2$$

where we think of "u" as an external input. This is basically the incoherent feedforward loop considered in [2] to study adaptation and the fold-change detection property in stochastic systems. The only difference is that in that paper we used separate creation and degradation reactions $0 \to S_2 \to 0$ (the first with rate u), but here, in order to impose a conservation law, we think of S_2 as being an active form of a kinase (the input controlling the change to active form), which can be constitutively de-activated by a reverse reaction. The effect of u on S_3 is incoherent, in the sense that u promotes formation of S_3, as well as degradation, because the larger u, the larger the active concentration of S_2, which degrades S_3. Observe that we have

$$\rho_1 = uS_1,\; \rho_2 = k_2 S_2,\; \rho_3 = k_3 u,\; \rho_4 = k_4 S_2 S_3,$$

and

$$a_1 = (1,0,0),\; a_2 = (0,1,0),\; a_3 = (0,0,0),\; a_4 = (0,1,1),$$

bioRxiv preprint doi: https://doi.org/10.1101/021410; this version posted July 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$\gamma_1 = (-1,1,0),\ \gamma_2 = (1,-1,0),\ \gamma_3 = (0,0,1),\ \gamma_4 = (0,0,-1)\,.$$

Note that $X_1(t) + X_2(t)$ is constant along all solutions. Suppose that $X_1(0) + X_2(0) = 2$ (any number will work, but the formulas are more involved). Let us obtain a linear differential equation for the mean of $X_3(t)$. We introduce these notations:

$$x_i = \mathbb{E}\left[X_2^i X_3\right],\ i = 0,1,2\,,\quad y_i = \mathbb{E}\left[X_2^i\right],\ i = 1,2\,.$$

We are interested in $x_0(t)$. In general (omitting arguments t),

$$\dot{x}_i = -k_4\mathbb{E}\left[X_2^{1+i}X_3\right] + k_3\mathbb{E}\left[X_2^i\right]u + k_2\mathbb{E}\left[X_2X_3((X_2-1)^i - X_2^i)\right] + \mathbb{E}\left[X_1X_3((X_2+1)^i - X_2^i)\right]u$$

and

$$\dot{y}_i = -k_2\mathbb{E}\left[X_2^{1+i}\right] + k_2\mathbb{E}\left[X_2(X_2-1)^i\right] + \mathbb{E}\left[X_1((X_2+1)^i - X_2^i)\right]u\,.$$

Using that $X_1 = 2 - X_2$ and $S_2^3 = -2S_2 + 3S_2^2$ (which is valid whenever $S_2 \in \{0,1,2\}$), we conclude that:

$$\begin{aligned}
\dot{y}_1 &= 2u - k_2y_1 - y_1u \\
\dot{y}_2 &= k_2y_1 + 2u + 3y_1u - 2y_2u - 2k_2y_2 \\
\dot{x}_0 &= k_3u - k_4x_1 \\
\dot{x}_1 &= k_3y_1u - k_4x_2 - k_2x_1 + 2x_0u - x_1u \\
\dot{x}_2 &= (k_2+2k_4)x_1 + 2x_0u - (3k_4+2k_2)x_2 + k_3y_2u + 3x_1u - 2x_2u
\end{aligned}$$

which may be written in the "bilinear" form $\dot{X} = (A + uB)X$, where $X = (y_0, y_1, y_2, x_0, x_1, x_2)'$ with the convention that $y_0 \equiv 1$ and appropriate 5×5 matrices A, B. Similar (but larger) systems may be written for the second and larger moments of X_3 as well as mixed moments.

More abstractly, given any finite continuous-time Markov chain with n_1 states q_i and transition rates λ_{ij}, we may introduce n_1 species S_i and reactions $S_i \to S_j$ with rate λ_{ij}. The stoichiometric matrix consists of columns with exactly one entry equal to 1 and one entry equal to -1, so the sum $X_1(t) + \ldots + X_{n_1}(t)$ is conserved (see e.g. Section 4.8 in [3]). Thus, starting from an initial condition with $X_1(0) + \ldots + X_{n_1}(0) = 1$ we have that at all times we have precisely one $X_i(t) = 1$. This provides an embedding of the Markov Chain: state is q_i at time t if $X_i(t) = 1$. This construction is of interest when reaction parameters κ_i in a network are described by functions of finite Markov chains (Hidden Markov Models) and the network is of a feedforward type, to conclude that finite-dimensional ODE's exist for moments.

References

[1] A. Singh and M. Soltani. Quantifying intrinsic and extrinsic variability in stochastic gene expression models. *PLoS ONE*, 8(12):e84301, 2013.

[2] M. Skataric, E.V. Nikolaev, and E.D. Sontag. Scale-invariance in singularly perturbed systems. In *Proc. IEEE Conf. Decision and Control, Los Angeles, Dec. 2014*, pages 3035–3040, 2014.

[3] E.D. Sontag. Lecture notes on mathematical systems biology, 2002-2015. `http://www.math.rutgers.edu/~sontag/FTP_DIR/systems_biology_notes.pdf`.