Physica A 392 (2013) 1122–1132

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Inference of analytical thermodynamic models for biological networks

Eliodoro Chiavazzo, Matteo Fasano, Pietro Asinari *

Energy Department, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

ARTICLE INFO

Article history:
Received 3 June 2012
Received in revised form 13 November 2012
Available online 23 November 2012

Keywords:
Kinetic theory
Non-equilibrium thermodynamics
Dynamical systems

ABSTRACT

We present an automated algorithm for inferring analytical models of closed reactive biochemical mixtures, on the basis of standard approaches borrowed from thermodynamics and kinetic theory of gases. As input, the method requires a number of steady states (i.e. an equilibria cloud in phase–space), and at least one time series of measurements for each species. Validations are discussed for both the Michaelis–Menten mechanism (four species, two conservation laws) and the mitogen-activated protein kinase–MAPK mechanism (eleven species, three conservation laws).

1. Introduction

The reverse engineering of biological networks from experimental observations has recently gained increasing attention, owing to remarkable advancements in modern high-throughput techniques for the generation of time series data on metabolites, genes and other components of biological relevance [1]. However, due to high dimensionality, the latter still remains a demanding task that often requires *a priori* knowledge on the system structure. Predictive mathematical models are highly desirable, for instance, for the external control of cellular functions, and this has motivated an intense effort in such a direction [2,3]. In this work, we intend to investigate the ability of some classical thermodynamic approaches for the automatic prediction of equilibria, dynamical behavior and system structure. The main advantage of such an approach is that it does not require prior knowledge on the underlying biochemical mechanism, and it is solely based on measurements of species concentrations in closed systems.

We focus on biological systems formed by several species interacting according to a web of (bio)chemical reactions in closed systems under fixed temperature T and volume V. We further assume that dissipation is ensured by the existence of a global Lyapunov function G, which is typically linked to a thermodynamic potential, and a unique steady state (equilibrium) is reached after a sufficiently long time. Let the concentration of n species evolve in time according to an autonomous system of ordinary differential equations (ODEs):

$$\dot{x} = \frac{\mathrm{d}x}{\mathrm{d}t} = f(x), \tag{1}$$

with $x = [x_1, \ldots, x_n]^T$ defining the system state (e.g. in terms of molar concentrations x_i). Let x^{eq} and $G(x)$ be the unique equilibrium state of the ODEs (1) and its global Lyapunov function, respectively. Hence, at all instant t, the time derivative of G is non-positive, $\dot{G} = \nabla G f \leq 0$, and it vanishes at steady state: $\dot{G}(T, V, x^{\mathrm{eq}}) = 0$. Time dynamics (1) is often characterized by linear constraints (e.g. due to conservation of the mole number of elements forming the chemical species). Thus, assuming the presence of r conserved quantities, there exists a fixed $(r \times n)$ matrix M such that, at all time instants t:

$$Mx(t) = C, \tag{2}$$

with C being an r-component column of fixed quantities (conserved moieties).

* Corresponding author. Tel.: +39 0110904520; fax: +39 0110904499.
E-mail addresses: eliodoro.chiavazzo@polito.it (E. Chiavazzo), pietro.asinari@polito.it (P. Asinari).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2012.11.030

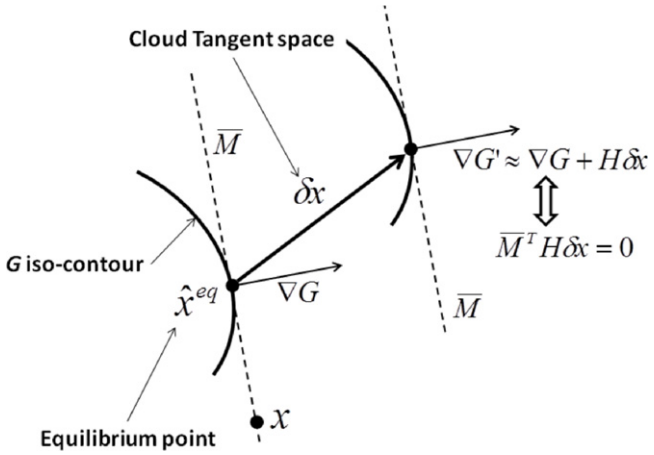

Fig. 1. Geometry underpinning the constrained minimization problems (8) and (14): Solutions are located where the affine hyperplane $x + \bar{M}$ is tangent to the G function iso-lines. Vectors δx spanning the local tangent space to the equilibrium cloud are thus linked to both the null space of the matrix M ($\overline{M} = \ker M$) and the second derivative matrix H of the Lyapunov function G. As a result, orthogonality between the columns of $\overline{M}$ and the gradient of G (∇G) implies: $\overline{M}^T H\delta x = 0$.

2. Searching for conservation laws

Neither the number nor the expressions of the conservation laws (2) are typically known when investigating on a new biological phenomenon, unless pre-existing knowledge on the reaction stoichiometry is available. For addressing the above issues, the suggested approach is based upon the analysis of a collection of scattered steady states (experimental equilibrium cloud), and at least one time series of species concentrations evolving from an arbitrary initial state. In this work, we perform inspection of the equilibrium cloud by means of *principal component analysis* – PCA – [4] in order to estimate the cloud dimension which, as discussed below, indicates the number of conservation laws. For the sake of completeness, it is worth stressing that more recent non-linear techniques, such as *diffusion maps* [5], may also be adopted for estimating the dimension r.

We notice that, in a perfectly closed system, thermodynamics and conservation laws rule the geometry of the manifold collecting all the equilibrium states. As a result, the matrix M is fully determined upon computation of the local tangent space to such a manifold. According to the pictorial representation in Fig. 1, we notice that the following relationship holds: $\overline{M}^T H \Delta X = 0$. Here, the columns of the $n \times (n - r)$ matrix $\overline{M} = \ker(M)$ span the null space of M, H is the second derivative matrix of the Lyapunov thermodynamic function G while the columns of the $(n \times r)$ matrix ΔX form a basis of the local tangent space of the equilibrium cloud. As a result, the following equation holds:

$$M' = \left[\ker\left[\left(\ker \Delta X^T\right)^T H^{-1}\right]\right]^T, \tag{3}$$

where the superscript ^{T} and the prime symbol $'$ denote transposition and the orthonormal basis respectively. For numerical purposes, a generic column δx of the matrix ΔX can be conveniently approximated by local interpolation or finite differences. Nevertheless, we stress that adoption of (3) may lead to inaccurate results due to a poor estimate of the vectors δx, or even to a lack of knowledge on the function G such as the activity coefficients for non-ideal mixtures (see below). Therefore, below we describe a stochastic method, based on the Metropolis algorithm [6], enabling an accurate computation of M by processing the time series of species concentrations. First, we initialize the k-th row, M_k, of the conservation law matrix: This can be made either stochastically or even on the basis of the estimate (3). We assume that all species concentrations are recorded at a discrete set of time instants (t_j) between t_1 and t_m: The latter data is thus available in the form of a time series, stored in the $(n \times m)$ data array $\hat{X} = \{\hat{x}_{ij}\}$, with the generic element $\hat{x}_{ij}$ denoting the concentration of the i-th species at the time instant t_j. Next, we compute the following deviation quantity:

$$d_k = \sum_{j=1}^{m} \left|\hat{C}_{kj} - \bar{C}_k\right|, \tag{4}$$

where, for the k-th conservation law, the j-th term of the time series $\{\hat{C}_{kj}\}$ and its time-averaged value $\bar{C}_k$ are defined as:

$$\hat{C}_{kj} = \sum_{i=1}^{n} \hat{x}_{ij} M_k(i), \qquad \bar{C}_k = \left(\sum_{j=1}^{m} \hat{C}_{kj} \mathrm{d}t_j\right) \Big/ (t_m - t_1),$$

with $\mathrm{d}t_j$ the duration of the time interval corresponding to the term $\hat{C}_{kj}$.