Math. Model. Nat. Phenom.
Vol. 7, No. 2, 2012, pp. 32–50
DOI: 10.1051/mmnp/20127102

Algorithms to Study Large Metabolic Network Dynamics

Dima Grigoriev[1] *, Satya Swarup Samal[2], Sergey Vakulenko[3], Andreas Weber[4]

[1] CNRS, Mathématiques, Université de Lille, Villeneuve d'Ascq, 59655, France

[2] Algorithmic Bioinformatics, Bonn-Aachen International Center for Informationtechnology, Dahlmannstraße 2, 53113, Bonn, Germany

[3] Institute for Mechanical Engineering Problems, Bolshoy pr. V. O.61, Saint Petersburg, Russia and ITMO Univerisity, Saint Petersburg

[4] Institut für Informatik II, University of Bonn, Friedrich-Ebert-Allee 144, 53113, Bonn, Germany

Abstract. We consider a class of systems of differential equations with quadratic nonlinearities. This class describes important biochemical models. We show that systems of this class can realize any structurally stable dynamics. Given a low dimensional dynamics, we describe algorithms that allow to realize this dynamics by a large biochemical network. Some concrete biochemical examples are studied. Moreover, we show how a big system with random kinetic rates can simulate a number of low dimensional ones. The proposed method is applied on Calcium oscillations, extracellular signal-regulated kinase (ERK) signaling pathway and multistationary Mitogen-activated protein kinase cascade system (MAPK) models from biochemistry.

Keywords and phrases: Metabolic networks, attractors

Mathematics Subject Classification: 35Q53, 34B20, 35G31

1. Introduction

Biochemical kinetic models can be represented as systems of ordinary differential equations with polynomial—often quadratic—nonlinearities. These polynomial nonlinearities appear as a result of the mass action law, and the corresponding models can involve hundreds of equations. Moreover, kinetic rate values, which are coefficients in these equations, are very often not known precisely. Therefore, the investigation of large time behavior and of the attractors of biochemical systems is a great challenge for biochemistry. Since one can expect that these systems are not robust (structurally stable), variations in kinetic rates can lead to dramatical effects. Note that direct numerical simulations do not allow to study all possible variants of large time dynamical behavior. Indeed, if a model includes 1000 differential equations and 2000 different kinetic coefficients, it is difficult to investigate all possible situations by an exhaustive search. An additional challenge is that kinetic rates of biochemical networks have multiple orders of magnitude. Signaling and metabolic reactions typically occur rapidly whereas regulatory events involved in cellular growth or cell differentiation can take hours. These multiple time-scales lead

*Corresponding author. E-mail: dmitry.grigoryev@math.univ-lille1.fr

to computational difficulties. Biochemical models are "stiff", i.e., they involve "fast" and "slow" reaction dynamics.

In this paper our goal is to state an algorithmic approach, which allows to obtain information about possible kinds of large time dynamical behavior without relying on primitive brute-force methods that do not scale to larger systems. In our approach the difficulties connected with "stiffness" become an advantage, which helps to find different reductions to low dimensional dynamics by algebraico-topological methods. The method (an algorithm) proposed here to describe different low dimensional reductions of dynamics of large metabolic networks uses the following tools: **i**) graph theory and algebraic methods; **ii**) theory of slow invariant manifolds; and **iii**) some special approximations.

Moreover, we shall show for an important class of systems of biochemical kinetics (let us denote this class by $\mathcal{C}$) that the dynamics of systems from $\mathcal{C}$ may be very complicated, even if nonlinearities involve only quadratic terms. For a larger class of quadratic systems such a result was established by M. D. Korzuchin [12], see also [19]. In particular, the Korzuchin theorem explains how the famous Belousov-Zhabotinsky reaction can appear (for an overview on this topic see [19]). We prove a theorem which shows that systems from this class $\mathcal{C}$ of large metabolic networks can exhibit a complicated dynamics. More precisely, this result asserts that such networks can approximate any prescribed dynamics with arbitrary given accuracy $\varepsilon > 0$, when the reagent number N is large enough and the kinetic rates are adjusted in a special way. This result extends the Korzuchin theorem on systems, which involve a smaller number of free parameters. We describe an algorithm of the realization of a given robust dynamics on infinite time intervals. We refer to our method, which is a purely algebraic, as "shock testing", since we make some kinetic rates small and other much larger. Morever, by the shock method we prove a theorem that even a *fixed* large quadratic kinetic system with $N \gg 1$ variables and random kinetic rates can simulate different systems with n variables. In a sense, a random big system from our class $\mathcal{C}$ can simulate a number of low dimensional ones belonging to $\mathcal{C}$. An algorithm producing such simulations is simple and it can be easily performed by genetic regulation mechanisms.

2. Model

We consider the following systems

$$\frac{dx_i}{dt} = -x_i S_i(x) + R_i(x), \tag{2.1}$$

where $x = (x_1, \ldots, x_N)^{tr}$, x_i are unknown reagent concentrations, and R_i are polynomials

$$R_i(x) = \sum_{\mathbf{j}}^{N} R_{i\mathbf{j}} x^{\mathbf{j}}, \quad R_{i\mathbf{j}} \geq 0, \tag{2.2}$$

where $\mathbf{j} = col(j_1, \ldots, j_N)$ are multiindices, and $x^{\mathbf{j}} = x_1^{j_1} \ldots x_N^{j_N}$. We suppose that the term R_i does not involve a factor x_i, i.e., $j_i = 0$. Notice that the fundamental theorem of chemical kinetics [8, 19] states that, if system (2.1) describes a chemical kinetics, then the coefficients $R_{i\mathbf{j}}$ must be non-negative:

$$R_{i\mathbf{j}} \geq 0. \tag{2.3}$$

In real metabolic networks the terms R_i are quadratic and S_i are linear functions having the form

$$S_i(x) = k_i + \sum_{j=1}^{N} K_{ij} x_j, \tag{2.4}$$

where

$$k_i > 0, \quad K_{ij} \geq 0. \tag{2.5}$$

Condition (2.5) helps to obtain estimates for solutions. In the book [19] it is shown that this condition follows from fundamental ideas of chemical kinetics. We set the initial conditions

$$x(0) = \phi. \tag{2.6}$$

Equations (2.1) and (2.6) give us a Cauchy problem.

We say that the vector x is positive if it lies in the positive cone $\mathbb{R}^N_> = \{x : x_i > 0\}$. Let us show that the Cauchy problem (2.1), (2.6) is well posed and that it defines a chemical kinetics.

Lemma 2.1. *For positive ϕ the Cauchy problem (2.1), (2.6) has positive solutions $x(t)$ on some open interval $(0, T(\phi))$. These solutions are unique. Therefore, problem (2.1), (2.6) defines a local semiflow in the positive cone $\mathbb{R}^N_>$.*

As an example, we consider a system of equations describing an interaction of a noncoding RNA with many mRNA's (see [21]). Some ncRNA's and especially miRNA's have many (up to a few hundreds) mRNA targets [1]. Direct experimental validations and reliable theoretical predictions of the targets are difficult problems. The kinetic models including association of ncRNA with two mRNA's and many mRNA's were analyzed in [18] and [21], respectively. In these works ncRNA's have been considered as *global regulators*. For m distinct mRNA's the equations take the following form:

$$\frac{dN_i}{dt} = w_i - k_i N_i - r_i N_i N_*, \tag{2.7}$$

$$\frac{dN_*}{dt} = w - k_* N_* - N_* \sum_{i=1}^{m} r_i N_i, \tag{2.8}$$

where N_i and N_* are the mRNA and ncRNA populations, w_i and u are the transcription rates, $k_i > 0$ and $k_* > 0$ are the degradation rate constants, and $r_i > 0$ are association rate constants. Naturally, we are looking for positive solutions assuming that $N_i, N_*(0) > 0$, since only such solutions have biological meaning.

This system was studied numerically [21]. In the dynamics of this relatively simple system oscillations do not appear. Indeed, it is easy to see that system (2.7), (2.8) defines a global dissipative strongly monotone semiflow in a box $\mathit{\Pi} \subset \mathbf{R}^{m+1}_+$, which follows from the condition $r_i > 0$. So this system has convergent trajectories and to describe its large time behavior one has to find the equilibria only.

3. Decomposition

We seek for an integer $n \ll N$ and a decomposition of variables $x = (z, y)$, where

$$z \in \mathbb{R}^{N-n}_>, \quad y \in \mathbb{R}^n_> \tag{3.1}$$

such that in new variables (z, y) the functions S_i and R_i take a special form. Let us denote by $\mathcal{K}_z$ the set of indices

$$\mathcal{K}_z = \{i_1, i_2, \dots, i_{N-n}\}, \quad z_l = x_{i_l},$$

corresponding to the z-subspace, respectively, corresponding to the y-subspace

$$\mathcal{K}_y = \{j_1, j_2, \dots, j_n\}, \quad y_l = x_{j_l}.$$

Assume that S_i and R_j can be represented in the following form:

$$R_i(z, y) = \sum_{\mathbf{j}}^{N} \bar{R}_{i\mathbf{j}} y^{\mathbf{j}} = T_i(y), \quad i \in \mathcal{K}_z, \tag{3.2}$$

$$R_i(z,y) = \sum_{\mathbf{j}}^{N} \sum_{\mathbf{k}}^{N} \tilde{R}_{i\mathbf{jk}} y^{\mathbf{j}} z^{\mathbf{k}}, \quad i \in \mathcal{K}_y, \tag{3.3}$$

$$S_i(z,y) = m_i + \sum_{j=1}^{n} M_{ij} y_j = P_i(y), \quad i \in \mathcal{K}_z, \tag{3.4}$$

$$S_i(z,y) = \bar{k}_i + \sum_{j=1}^{n} \bar{K}_{ij} y_j + \sum_{j=1}^{N-n} \tilde{K}_{ij} z_j \quad i \in \mathcal{K}_y. \tag{3.5}$$

One can suppose that y-variables are hubs in the networks, i.e., they participate in many reactions and mediate them whereas z-variables are "satellites". This structure is typical for many biochemical network systems [15, 20].

For quadratic systems (2.1) with linear S_i of the form (2.4) and quadratic R_i such that

$$R_i(x) = r_i + \sum_{i \neq j}^{N} \bar{R}_j x_j + \sum_{k \neq i}^{N} \sum_{l \neq i}^{N} \bar{R}_{kl} x_k x_l \tag{3.6}$$

this decomposition can be made by the vertex cover method (see below).

3.1. Systems with conservation laws

In many systems terms $R_i(x)$ do not contain constant terms r_i, but these systems have some linear conservation laws

$$\sum_{sj} e_{sj} x_j = E_s, \quad s = 1, \ldots, N_{CL}, \tag{3.7}$$

where N_{CL} is the number of conservation laws.

One can suggest two variants to handle these systems (a method that we state below, requires non-zero r_i).

a) First we make a separation into fast and slow variables. If the linear conservation laws involve fast variables z_j only, we choose some basis of linear independent variables z_j. Then in new variables our system reduces to a standard form, as above, with r_i depending on E_s.

b) We first choose a basis of linear independent variables x, and then we proceed to a system decomposition on slow and fast components by a procedure connected with a vertex cover finding (see below).

4. Reduction to shortened system by slow manifolds

In this subsection, our goal is to reduce the dynamics of a large system to a shortened y-dynamics. Notice first that if the right hand sides of (2.1) can be transformed to the form (3.2)-(3.5), then we can simplify the task of finding equilibria (note that for a general large system (2.1) it is a formidable problem). Namely, we obtain

$$z_i = \frac{T_i(y)}{P_i(y)} = Z_i(y), \tag{4.1}$$

where T_i and P_i are defined by (3.2) and (3.4), respectively. After we substitute Z_i in equations for y:

$$y_i S_i(y, Z(y)) = R_i(y, Z(y)). \tag{4.2}$$

The reduction of dynamics can be done under the following assumption:

Assumption (S). *Assume that the kinetic coefficients involved in y -equations are small with respect to kinetic coefficients k_i involved in z-equations:*

$$\sup_i |\bar{k}_i| + \sup_{i,j} |\bar{K}_{ij}| + \sup_{i,j} |\tilde{K}_{ij}| < \kappa(m_0), \tag{4.3}$$

where

$$m_0 = \min_j m_j.$$

Biologically this means that the dynamics of the y-variables is slow with respect to dynamics of the remaining z-variables.

Lemma 4.1. *Suppose assumption (S) holds. Consider the equations*

$$\frac{dy_i}{dt} = -y_i S_i(y, Z(y)) + R_i(y, Z(y)) = F_i(y). \tag{4.4}$$

Assume that (4.4) defines a global semiflow in an open bounded connected domain $D \subset \mathcal{R}^n$ with a smooth boundary ∂D.

For sufficiently small κ there exists a locally invariant and locally attracting manifold $\mathcal{Z}_n$ of dimension n defined by

$$z_i = Z_i(y) + \tilde{Z}_i(y, \kappa), \quad y \in D, \tag{4.5}$$

where

$$|\tilde{Z}_i(\cdot, \kappa)|_{C_1(D)} \to 0 \quad \kappa \to 0. \tag{4.6}$$

The dynamics of system (2.1) restricted to $\mathcal{Z}_n$ is defined by the equations

$$\frac{dy_i}{dt} = F_i(y) + \tilde{F}_i(y, \kappa), \tag{4.7}$$

where

$$|\tilde{F}_i(\cdot, \kappa)|_{C_1(D)} \to 0 \quad \kappa \to 0. \tag{4.8}$$

Proof. The proof is standard and follows from Theorem 6.1.7 of [10].

Let us describe an algebraic procedure that allows us to construct more and more accurate approximations to the solutions of system (2.1). Let us write down this system in (z, y) variables as follows:

$$\frac{dz_i}{dt} = -z_i Z_i(y) + G_i(y), \tag{4.9}$$

$$\frac{dy_i}{dt} = \kappa F_i(y, z), \tag{4.10}$$

where κ is a small parameter. According to the slaving principle [7] we can represent solutions of this system in the following form

$$z_i = z_i^{(0)} + \kappa z_i^{(1)} + \cdots,$$

where the main term $z_i^{(0)}$ is defined by $z_i^{(0)} = G_i(y) Z_i(y)^{-1}$. For the first correction $z_i^{(1)}$ one has

$$\frac{dz_i^{(1)}}{dt} = -z_i^{(1)} Z_i(y) + \sum_{j=1}^{n} \frac{\partial z_i^{(0)}}{\partial y_j} F_j(y, z^{(0)}(y)), \tag{4.11}$$

that implies

$$z^{(1)} = \sum_{j=1}^{n} \frac{\partial z_i^{(0)}}{\partial y_j} F_j(y, z^{(0)}(y)).$$

It is clear that this procedure can be continued and it gives us asymptotical series for the solution z. □

This last lemma shows that, under some assumptions, the dynamics of a global network (2.1) can be reduced to n equations. In the coming section we describe the method of a dynamics control for this shortened system.

5. Control of slow dynamics

The main idea beyond the mathematical construction of this control is as follows. We adjust coefficients that define an interaction between slow modes y and fast modes z. The number of these coefficients is much bigger than the number of slow modes n. This fact allows us to apply some special approximations.

To simplify the further statement, let us consider the case when the polynomials T_i are linear functions of y, i.e., $deg(T_i) = 1$. Therefore,

$$T_i(y) = \tau_i + \sum_{j=1}^{n} T_{ij} y_j, \quad i \in \mathcal{K}_z. \tag{5.1}$$

Then Z_i are given by

$$Z_i(y) = \frac{\tau_i + \sum_{j=1}^{n} T_{ij} y_j}{m_i + \sum_{j=1}^{n} M_{ij} y_j}. \tag{5.2}$$

Thus, for $i \in \mathcal{K}_y$ one has

$$S_i(y) = \bar{k}_i + \sum_{j=1}^{n} \bar{K}_{ij} y_j + E_i(y, \tilde{K}, T, \tau), \tag{5.3}$$

where

$$E_i(y, \tilde{K}, T, \tau, m, M) = \sum_{j=1}^{N-n} \tilde{K}_{ij} Z_j(y). \tag{5.4}$$

We consider the field $E(y)$ with components E_i in the domain D. Besides E, we consider more general fields $G(y)$ defined by

$$G_i(y, \tilde{K}, T, \tau, m, M, C) = E_i(y, \tilde{K}, T, \tau, m, M) + C_i \tag{5.5}$$

depending on constants C_i. These constants can have any signs. A possibility to control E and thus S by the parameters $\tilde{K}, T, \tau$ follows from the next assertion.

Lemma 5.1. *Let us fix an integer $n > 0$. Consider the set $\mathcal{G}_n$ of all vector fields $G(y)$ defined by (5.5) on a bounded open subdomain $D \subset \mathbb{R}^n$ for all possible values of N and all possible positive values of the coefficients $\tau_i, \tilde{K}_{ij}, T_{ij}, m_p, M_{pl}, C_i$ where $i, l = 1, \ldots, n$, $j = 1, \ldots, N-n$ and $p = 1, \ldots, N-n$. This set $\mathcal{G}_n$ is dense in the space of all smooth vector fields F enabled by C^1-norm.*

Proof. Step 1. The Fourier approximation and reduction to one dimensional case.

We can assume that $D \subset \Pi$, where $\Pi = [0, 2\pi]^n$ is a box. Then for each $\varepsilon > 0$ and a given smooth field $H(y)$ we can construct an approximation

$$H_i(y) = \sum_{\mathbf{m} \in \mathcal{M} \subset \mathbb{N}^n} H^+_{i,\mathbf{m}} \cos(\mathbf{m} \cdot y) + H^-_{i,\mathbf{m}} \sin(\mathbf{m} \cdot y) \tag{5.6}$$

such that

$$|H_i - G_i|_{C^1(D)} < \varepsilon. \tag{5.7}$$

Here $\mathbf{m}$ is a multiindex, $\mathbf{m} = (m_1, \ldots, m_n)$, where $m_i \in \mathbb{N}$, $\mathcal{M}$ is finite subset of $\mathbb{N}^n$ and $\mathbf{m} \cdot y = m_1 y_1 + m_2 y_2 + \ldots + m_n y_n$. Relations (5.4)- (5.7) show that it is sufficient to approximate fields H of the form

$$H_i(y) = H_i(q_{\mathbf{m}}), \quad q_{\mathbf{m}} = \mathbf{m} \cdot y. \tag{5.8}$$

□

Remark 5.2. For systems, which involve linear and quadratic terms only, we can use the following elementary trick instead the Fourier decomposition. We can present the term $y_i y_j$ with $i \neq j$ as follows:

$$y_i y_j = \frac{1}{2}((y_i + y_j)^2 - y_i^2 - y_j^2). \tag{5.9}$$

Step 2. One-dimensional case.

The problem of approximation of fields (5.8) by G_i can be formulated as follows. Given $\varepsilon > 0$ and the smooth function $h(q)$ on some bounded interval $[0, \beta]$, to find coefficients $k_j, C, r_i, a_j, b_j, \tau_j > 0$ such that the sum

$$g(q) = C + \sum_{i=1}^{N_0} k_i \frac{\tau_i + b_i q}{r_i + a_i q} \tag{5.10}$$

approximates the function h on $[a, b]$:

$$|g(\cdot) - h(\cdot)|_{C^1[0,\beta])} < \varepsilon. \tag{5.11}$$

Let us set $b_i = a_i$. Then we can transform (5.10) as follows:

$$g(q, \gamma, r, a, C') = C' + \tilde{g}(q), \quad \tilde{g} = \sum_{i=1}^{N_0} \frac{\gamma_i}{r_i + a_i q}, \tag{5.12}$$

where

$$C' = C + \sum_{i=1}^{N_0} k_i, \quad \gamma_i = k_i(\tau_i - r_i).$$

There are possible different approaches resolving this approximation problem. We state two methods. The first approach uses the Weierstrass theorem and polynomial approximations, and the second one is a least square procedure.

Method 1. According to the Weierstrass theorem, linear combinations

$$R(q) = C' + \sum_{k=1}^{M} c_k (1 + q)^{-k}, \tag{5.13}$$

where M is an arbitrary integer and c_k are arbitrary coefficients, are dense in the space of all smooth functions on $[0, \beta]$ in the C^1 norm.

Thus, it is sufficient to prove that functions $P_k(q) = (1 + q)^{-k}$ can be approximated, with an arbitrary accuracy, by g defined by (5.12). We can prove it by an inductive procedure. It is clear that this fact holds for $k = 1$. Assume this assertion holds for $k = p$. Let us prove it for $k = p + 1$. We use the estimate

$$|(1 + q)^{-p-1} - c_p h((1 + q)^{-p} - (1 + (q - h))^{-p})| < C_p h,$$

that is valid on $[0, \beta]$ for appropriate constants c_p and $C_p > 0$. For small h this estimate gives the needed approximation. Thus the lemma is proved.

Method 2. We can apply to this approximation problem the classical method of least squares. Let us introduce the matrix $\mathbf{A}$ with the entries

$$A_{ij} = \int_0^1 (a_i + q_j)^{-1} (a_j + q)^{-1} dq. \tag{5.14}$$

Notice that

$$A_{ij} = (a(i) - a(j)^{-1}(\log(a_i + 1)/a_i)) - \log(a_j + 1)/a_j)))$$

for $i \neq j$, and

$$A_{ii} = a_i^{-1} - (a_i + 1)^{-1}.$$

Let us define the vector $B = (B_1, \ldots, B_n)$ by

$$B_i = \int_0^1 (a_i + q)^{-1} h(q) dq. \tag{5.15}$$

We solve the linear algebraic system

$$\mathbf{A}X = B \tag{5.16}$$

and the solution X gives the optimal approximation in the $L_2([0,\beta])$-norm.

Let us now consider the second term in the right hand side of (4.4). We assume that $R_i(y,z)$ are linear in z:

$$R_i(y,z) = \bar{r}_i + \sum_{l=1}^{N-n} \bar{R}_{il} z_l. \tag{5.17}$$

Then there holds a lemma analogous to the previous one. Let us consider the vector fields J with components J_i defined by

$$J_i(y) = \bar{r}_i + \sum_{l=1}^{N-n} \bar{R}_{il} Z_l(y). \tag{5.18}$$

Lemma 5.3. *Let us fix an integer $n > 0$. Consider the set $\mathcal{J}_n$ of all vector fields $J(y)$ defined by (5.18) on a bounded open subdomain $D \subset \mathbb{R}^n$ for all possible values of N and all possible positive values of the coefficients $\bar{r}_i, \bar{R}_{ij}, T_{ij}, m_p, M_{pl}$ where $i,l = 1,\dots,n$, $j = 1,\dots,N-n$ and $p = 1,\dots,N-n$. This set $\mathcal{J}_n$ is dense in the space of all smooth vector fields F on D in the C^1-norm.*

The proof is the same as for Lemma 5.1.

5.1. Theorem on the complexity of the dynamics

Consider the following class of systems of chemical kinetics:

$$\frac{dx_i}{dt} = -x_i(k_i + \sum_{j=1}^{N} K_{ij}x_j) + r_i + \sum_{j=1}^{N} R_{ij}x_j, \tag{5.19}$$

where $x = (x_1,\dots,x_N) \in \mathbb{R}^N$, $i = 1,\dots,N$ and

$$k_i > 0, K_{ij} \geq 0, \quad r_i > 0, \quad R_{ij} \geq 0, \quad R_{ii} = 0. \tag{5.20}$$

We refer this class as $\mathcal{C}$. Under conditions (5.20) it is easy to show that the Cauchy problem for system (5.19) defines a global semiflow in the positive cone $\mathbb{R}^N{}_>$. In fact, the solutions of system (5.19) can be estimated by

$$||x|| \leq C(r,k) + ||x(0)|| \exp(\lambda_{min} t),$$

where $\lambda_{\min}$ is the maximal eigenvalue of the matrix T with entries $T_{ij} = -k_i\delta_{ij} + R_{ij}$. If $\lambda_{\min} < 0$ then system (5.19) generates a dissipative semiflow because then the solutions are globally bounded as $t \to +\infty$:

$$x(t) \in D_\delta, \quad t > T_0(x(0),\delta),$$

where

$$D_\delta = \{x : 0 < x_i < r_i/k_i + \delta\}$$

and $\delta > 0$ is a arbitrary positive number. Furthermore, we use the method of realization of vector fields [3] that implies the following assertion.

Consider vector fields $Q(y)$ defined on a balll $B^n \subset \mathbb{R}^n_>$ such that

$$|Q(y)|_{C^1(B^n)} < 1, \quad Q(y)\cdot n(y) < 0 \quad y \in \partial B^n, \tag{5.21}$$

where n is a outward normal vector at y to the boundary ∂B^n. Therefore, Q is directed inward B^n on the boundary ∂B^n and equation $dy/dt = Q(y)$ defines a global semiflow.

Theorem 5.4. *Consider the family $S^t(P)$ of semiflows defined by system (5.19) where the parameters P are the number N and the non-negative coefficients k_i, K_{ij}, r_i, R_{ij}. Then this family generates dynamics, maximally complex in the following sense. For any $\varepsilon > 0$ and a vector field Q satisfying (5.21) there exists a choice of parameters P such that*

*(***i***) the corresponding semiflow $S^t(P)$ has a n-dimensional locally invariant and locally attracting manifold $\mathcal{Z}_n$ defined by a C^1 smooth map $x = Z(y)$, $y \in B^n$;*

*(***ii***) the semiflow $S^t(P)$ restricted to $\mathcal{Z}_n$ is defined by*

$$\frac{dy}{dt} = Q(y) + \tilde{Q}(y), \quad |\tilde{Q}(y)|_{C^1(B^n)} < \varepsilon. \tag{5.22}$$

The dynamics defined by systems of class $\mathcal{C}$ can generate all possible (up to orbital topological equivalencies) structurally stable dynamics, in particular, all hyperbolic dynamics.

Proof. Estimate (5.22) immediately follows from lemmas 4.1, 5.1 and 5.3, under appropriate choice of P. The second assertion follows from the definition of structural stability and this estimate for sufficiently small ε [16]. The assertion on hyperbolic dynamics follows from the structural stability (persistence) of hyperbolic sets [16].

Let us make a brief comments in connection with fundamental Korzuchin's theorem [12]. This result asserts that any dynamics defined by a system of ordinary differential equations can be realized within a bounded time interval by quadratic systems of chemical kinetics including only monomolecular and bimolecular reactions. These systems are more complicated than systems (5.19) and have the form

$$\frac{dx_i}{dt} = \sum_{j=1}^{N} A_{ij}x_j + \sum_{j=1}^{N} B_{ijl}x_j x_l, \tag{5.23}$$

where A_{ij}, B_{ijl} satisfy some conditions of non-negativity [19]. Theorem 5.4 can be considered as an extension of the Korzuchin theorem on the narrower class $\mathcal{C}$ of models of chemical kinetics. □

6. Checking existence of oscillations, bifurcations, and chaos

The method, described in the previous sections can meet difficulties for large n, since at the very beginning, we can obtain the number of the Fourier coefficients being exponential in n. In fact, the set $\mathcal{M}$ can contain an exponential number of multiindices $\mathbf{m}$. It is a typical effect first described by R. Bellman and called "curse of dimensionalit". How to check then—by a really feasible algorithm—that a large metabolic network can exhibit oscillations or, say, saddle-node bifurcations?

To avoid this curse we propose to use the fundamental idea of "normal forms". We reduce system (4.4) to a system of low dimension, which can exhibit these oscillation and bifurcation effects. For such smaller system the information on equilibria and bifurcations can be obtained by symbolic computations.

For example, we would like to check that a large network exhibits time oscillating solutions. Simple systems, which can demonstrate such a behavior, have the form [13]

$$\frac{dy_1}{dt} = a_0 + a_1y_1 + a_2y_2 + a_{11}y_1^2 + a_{12}y_1y_2 + a_{22}y^2 + \kappa Y_1(y), \tag{6.1}$$

$$\frac{dy_2}{dt} = b_0 + b_1y_1 + b_2y_2 + b_{11}y_1^2 + b_{12}y_1y_2 + b_{22}y_2^2 + \kappa Y_2(y), \tag{6.2}$$

where $a_i, b_i, a_{ij}, b_{ij} \in \mathbb{R}$ are coefficients, $i, j = 1, 2$, $\kappa > 0$ is a small parameter and S_i, Y_1, Y_2 satisfy

$$|Y_i(\cdot)|_{C^1(D)} \le C_1, \quad \kappa > 0 \quad y \in D. \tag{6.3}$$

For $\kappa = 0$ the time behaviour of system (6.1),(6.2) is well studied. These equations can exhibit an Andronov-Hopf bifurcation and limit cycles (for example, coexistence of 4 cycles) [13]. It is clear that

this system also exhibits the saddle-node bifurcations since the normal form of this bifurcation is

$$\frac{dy_1}{dt} = a_0 - a_{11}y_1^2, \quad a_{11} > 0, \tag{6.4}$$

$$\frac{dy_2}{dt} = -b_0 y_2 + O(y_2^2), \quad b_0 > 0. \tag{6.5}$$

This bifurcation occurs, when a_0 passes through 0, whereas b_0, a_{11} are fixed.

To obtain chaos, we can try to reduce (4.4) to the Lorenz system, which demonstrates complicated behavior, including global convergence of trajectories, Andronov-Hopf bifurcations, and chaos. In the next section we describe an algorithm to find reductions to systems like (6.1), (6.2).

6.1. Algorithm

Input is given by a quadratic system

$$\frac{dx_i}{dt} = -x_i(k_i + \sum_{j=1}^{N} K_{ij}x_j) + r_i + \sum_{j\neq i}^{N} R_{ij}x_j, \tag{6.6}$$

where $i = 1, \dots, N$, and N can be large.

The goal of the algorithm is to reduce system (5.19) to a smaller system, which involves rational nonlinearities, and find equilibria and Andronov-Hopf bifurcations for this system.

Below we state the algorithm, which is purely algebraic and use symbolic algebra. The proof of theorem 5.4 also gives an algorithm, however, it uses numerical approximations.

Step 1: Vertex cover. We find a graph (V, E), where substances x_i are vertices, therefore, $V = \{1, 2, \dots, N\}$. The edge $(i, j) \in E$ if and only if the both x_i and x_j involved in a quadratic term in the right hand sides of the system (6.6).

The algorithm works if there is a vertex cover $\mathcal{C} = \{i_1, \dots i_m\}$ for (V, E), which includes $m \geq n$ vertices and $m \ll N$ is not so large (we suppose that $m \in (1, 50)$). The problem of finding a minimal vertex cover is a NP-complete problem (see [2, 5]). However, there exists a simple greedy 2-approximation algorithm (see[2]), which finds a cover that contains $\leq 2m^*$ vertices (if the best cover contains m^* vertices). If such a vertex cover exists, we make a decomposition $x = (y, z)$ described and studied above. We denote

$$y_k = x_{j_k}, \quad k = 1, \dots, m,$$

and z_l, where $l = 1, \dots, N - n$ will be all the rest of x_i such that $i \neq i_k, k = 1, \dots, m$. System (6.6) takes the form

$$\frac{dz_i}{dt} = -z_i(\tilde{k}_i + \sum_{j=1}^{m} \tilde{K}_{ij}y_j) + r_i + \sum_{j\neq i}^{N-m} \tilde{R}_{ij}z_j + \sum_{j=1}^{m} \bar{R}_{ij}y_j. \tag{6.7}$$

where $i = 1, 2, \dots, N - m$,

$$\frac{dy_j}{dt} = -y_i(\bar{k}_i + \sum_{j=1}^{m} \tilde{M}_{ij}y_j + \sum_{l=1}^{N-m} S_{ij}z_j) + r_i + \sum_{l\neq j}^{N-m} T_{ij}z_j + \sum_{l\neq j}^{m} P_{il}y_l. \tag{6.8}$$

where $j = 1, \dots, m$. Now we express z via y using (6.7) assuming that z_i are fast and y_j are slow. We obtain $z_i = Z_i(y)$, where Z_i are rational functions. If $\tilde{R}_{ij} = 0$, we have simple formulas for Z_i:

$$Z_i(y) = \frac{\sum_{j=1}^{m} \bar{R}_{ij}y_j}{\tilde{k}_i + \sum_{j=1}^{m} \tilde{K}_{ij}y_j}. \tag{6.9}$$

Note that if we seek equilibria of (6.7), (6.8), then (6.9) is an exact relation. Substituting (6.9) into (6.8) one has

$$\frac{dy_j}{dt} = -y_i(\bar{k}_i + \sum_{j=1}^{m} \tilde{M}_{ij} y_j + \sum_{l=1}^{N-m} S_{ij} Z_j(y) + r_i + \sum_{l\neq j}^{N-m} T_{ij} Z_j(y) + \sum_{l\neq j}^{m} P_{il} y_l. \tag{6.10}$$

We call these equations the vertex cover system. It is a reduction, which can be found in a Poly(N) running time by symbolic methods.

Step 2

For example, we would like to check that the reduced system (6.6) exhibits an Andronov-Hopf bifurcation.

a) We seek an equilibrium y_{eq} for the vertex cover system. Further, we use the Taylor series for the right hand sides at $y_{\rm eq}$ setting

$$w = y - y_{\rm eq}.$$

Then the vertex cover system takes the form

$$\frac{dw_j}{dt} = \xi_i(S,T) + \sum_{j=1}^{m} \eta_{ij}(S,T) w_j + \sum_{j=1}^{l=1} \sum \phi_{ijl}(S,T) w_j w_l, \tag{6.11}$$

where ϕ_{ijl}, η_{ij} and ξ_i are linear functions of $S_{i'j'}, T_{i'l'}$.

b) By symbolic algebra [4], we can find a condition that system (6.10) admits an Andronov-Hopf bifurcation in an explicit symbol form via ξ, ϕ, η, say

$$\eta_{11} < \phi_2 + 1.2. \tag{6.12}$$

c) Then we substitute $\xi_i(S,T)$ and other expressions into (6.12) and obtain a condition for the existence of Andronov-Hopf bifurcations.

7. Why networks are large?

Numerical realizations show some interesting results. Let us consider realizations of quadratic vector fields. Then we can use identity (5.9) and the problem reduces to an approximation of quadratic polynomials $\mu(q) = c_1 q + c_2 q^2$ by sums $g(q)_X = \sum_{i=1}^{N_0} X_i (a_i + q)^{-1}$ on some interval $[0, \beta]$, where $\beta > 0$. It is clear that without loss of generality one can set $\beta = 1$. Since the approximation problem is linear, we can consider two cases: $\mu_1 = q$ and $\mu_2 = q^2$.

We have applied the method of least squares (the method 2). We set $a_i = i\bar{a}$, where $i = 1, \ldots, N_0$ and $N_0 = 2, 3, 5$. By a program in MATLAB2009 we have estimated the following quantities:

$$d_0 = \sup_{q\in[0,1]} |\mu_i - g_X(q)|, \quad d_1 = \sup_{q\in[0,1]} |\frac{d\mu_i}{dq} - g'_X(q)|,$$

and $|X| = \max |X_i|$. In these relations, $X = X(\bar{a}, N_0)$ are optimal vectors that can be found by the method of least squares by relations (5.14), (5.15) and (5.16). The quantities d_i give the approximation precisions and $\bar{X}$ gives an estimate of the maximum of the kinetic rates that involved in (y, z) interaction (it is a force of the satellite action on the hubs). They depend on $\bar{a}$ and N_0. The number N_0 can be interpreted as a complexity of the network. The number $\bar{a}$ estimates the magnitude of some kinetic rates. These rates determine the action of a center on satellites.

We have found that one can obtain a good approximation for μ_1 by $N_0 = 2, 3$ and for μ_2 by $N_0 = 3$. For $\mu_1 = q$ and $\bar{a} = 5, N_0 = 3$ it is found

$$d_0 \approx 0.002, \quad d_1 \approx 0.024, \quad \bar{X} \approx 1220, \tag{7.1}$$

and for $\bar{a} = 20, N_0 = 3$ it is found that

$$d_0 \approx 10^{-5}, \quad d_1 \approx 0.0015, \quad \bar{X} \approx 17000. \tag{7.2}$$

Similar results can be obtained for $\mu_2 = q^2$. We observe therefore that the approximation precision increases when $\bar{a}$ grows but, on the other hand, then the kinetic rates increase. To see this dependence and obtain analytical results let us consider the case $\bar{a} \gg 1$. Some rough estimates based on method 2 then show (we omit technical details) that

$$d_i = O(\bar{a})N_0^{-1/2}, \quad \bar{X} = O(\bar{a}^3)N_0^{-1}. \tag{7.3}$$

Thus, given a network of complexity N_0, in order to obtain a prescribed dynamics of a network, it is necessary to use either a strong action of the centers to the satellites, or from the satellites to the centers (or both interactions should be strong). Note that in real applications the kinetic rates are bounded by some constants and cannot become arbitrarily big. Therefore, the networks, which are flexible and show different kinds of dynamics, should be complex ($N_0 \gg 1$). Increasing N_0 (which a polynomial function of N, actually it is the connectivity of hubs, see [11]), we can conserve the same precision of approximation of a required dynamics having restricted kinetic rates.

8. Shock testing for random systems

Let us consider system (5.19) assuming that $N \gg 1$ and coefficients K_{ij}, and R_{ij} are random. More precisely, let each K_{ij}, R_{ij} and r_i be distributed according to the normal laws $\mathcal{N}(b_0, \sigma_0^2)$ $\mathcal{N}(b_1, \sigma_1^2)$ and $\mathcal{N}(b_3, \sigma_3^2)$, respectively, where $b_i, \sigma_i > 0$. We assume that all K_{ij} and K_{kl} are independent if $(i, j) \neq (k, l)$, and the same holds for R_{ij} and r_i. Moreover, all K_{ij} , R_{kl} and r_i are independent. Note that results of this section are valid for all other distribution laws, if the corresponding densities $\rho(x)$ are C^2-smooth and positive for $x \geq 0$.

Together with a large random system (5.19) let us consider a smaller system of the same structure, namely

$$\frac{du_i}{dt} = -u_i(\bar{k}_i + \sum_{j=1}^{n} \bar{K}_{ij}u_j) + \bar{r}_i + \sum_{j=1}^{n} \bar{R}_{ij}u_j \tag{8.1}$$

where $i = 1, \ldots, n$. We assume that the dynamics generated by this system is dissipative, i.e., there is a compact absorbing set D with open interior, which attracts all trajectories. Let us consider the realization of (8.1) by (5.19). This realization can be found as follows. We can call it "realization by shock decay", since we sharply increase some decay constants k_i.

We choose n indices $S = \{i_1, \ldots i_n\} \subset I_N = \{1, \ldots, N\}$. Let us introduce $u_k = x_{i_k}$, where $k = 1, \ldots, n$. So, we have a decomposition $I_N = S \cup F$, where F is a set of indices corresponding to fast variables. Furthermore, let us take $k_i = a$ for all $i \neq i_1, i_2, \ldots i_n$, where a is a large enough and $k_i = \bar{k}_i$ otherwise. Then, by slow manifolds, we see that fast variables are defined by

$$x_i = a^{-1}(r_i + \sum_{j \in S} R_{ij}x_j) + a^{-2}\eta_i(x_{i_1}, \ldots., x_{i_n}, a), \quad i \in F, \tag{8.2}$$

where η_i are smooth functions of slow variables bounded in $C^1(D)$-norm for each compact domain D as $a \to \infty$. We substitute these relations into eqs. (5.19) for slow components x_i with $i \in S$. As a result, we obtain system (8.1) with small perturbations

$$\frac{du_i}{dt} = -u_i(\bar{k}_i + \sum_{j=1}^{n} \bar{K}_{ij}u_j) + \bar{r}_i + \sum_{j=1}^{n} \bar{R}_{ij}u_j + \phi_i(u, a), \tag{8.3}$$

where

$$\bar{K}_{kl} = K_{i_k i_l}, \quad \bar{R}_{kl} = R_{i_k i_l}, \quad \bar{k}_l = k_{i_l}, \quad \bar{r}_l = r_{i_l}, \tag{8.4}$$

and $\phi_i(u,a)$ are smooth functions such that

$$|\phi_i(u,a)|_{C^1(D)} < c_D(N,n)a^{-1}, \tag{8.5}$$

for any compact domain D and some constant $c_D(N,n) > 0$. If $N \gg n$ then almost all systems (8.1) can be represented in such a way. Let us formulate a lemma.

Lemma 8.1. *Under the above assumptions, the probability $P(N,n,\varepsilon)$ that system (5.19) realizes a given short system (8.1) with accuracy ε can be estimated by*

$$P(N,n,\varepsilon) > 1-(1-p_n(\varepsilon))^m, \qquad p_n(\varepsilon) > C(n)\varepsilon^{2n^2+n}. \tag{8.6}$$

where m is the integer closest to N/n.

Proof. We consider the following decompositions generated by the disjoint sets: $S_1 = \{1,\dots,n\}$, $S_2 = \{n+1,\dots,n\}$, $\dots$, $S_m = \{mn+1,\dots,mn+n\}$. Let us denote the sets $M_{\bar{K},\varepsilon}$, $M_{\bar{R},\varepsilon}$ and $M_{\bar{r},\varepsilon}$ of $n\times n$ matrices X and the vectors y by

$$M_{\bar{K},\varepsilon} = \{X:\ |X_{kl}-\bar{K}_{kl}| < \varepsilon,\ k,l = 1,\dots,n\},$$

$$M_{\bar{R},\varepsilon} = \{X:\ |X_{kl}-\bar{R}_{kl}| < \varepsilon\ k,l = 1,\dots,n\},$$

and

$$M_{\bar{r},\varepsilon} = \{y:\ |y_k-\bar{r}_k| < \varepsilon\ k = 1,\dots,n\}.$$

Let us now consider the sets T_s, $s = 1,\dots,m$ of triples consisting of $n\times n$ submatrices X of the matrix K, submatrices Z of the matrix R and subvectors y of the vector r defined by $X_{kl} = K_{i_k i_l}$, $Z_{kl} = K_{i_k i_l}$ and $y_k = r_{i_k}$, where $k \to i_k$, $k = 1,\dots,n$ is a map of indices associated with the set S_s, where $s = 1,\dots,m$. For any s the probability $P_n(\varepsilon,s)$ that the corresponding set T_s intersects with $M_{\bar{K},\varepsilon}$ does not depend on N and it admits the estimate $P_n(\varepsilon,s) = 1-p_n(\varepsilon)$, where $p_n > C(n)\varepsilon^{2n^2+n}$. Since all submatrices and vectors are independent, we obtain (8.6).

□

Combining this lemma and Theorem 5.4, we obtain the following main result.

Theorem 8.2. *Consider the family $S^t(P)$ of semiflows defined by system (5.19) with random independent parameters. Consider a vector field $Q(y)$ defined on a ball $B^n \subset \mathbb{R}^n_>$ and satisfying assumptions of Theorem 5.4. For any $\varepsilon > 0$ let us denote by $P(N,n,\varepsilon,Q)$ the probability that—as a result of some "realization by shock decay" for an appropriate variable choice—the corresponding semiflow $S^t(P)$ satisfies assertions (***i***) and (***ii***) of Theorem 5.4. Then this probability can be estimated by*

$$P(N,n,\varepsilon,Q) > 1-(1-q_n(\varepsilon,Q))^m, \qquad q_n(\varepsilon,Q) > 0, \tag{8.7}$$

where m is the integer closest to N/n.

This means that the dynamics defined by a large random system of class $\mathcal{C}$ can generate a number of possible structurally stable dynamics. Note, however, that it is difficult to give an explicit estimate for $q_n(\varepsilon,Q)$.

9. Application to real biochemical systems

We consider a small and two larger two biochemical systems important in applications.

9.1. Calcium oscillation model from [6]

9.1.1. Input System

This model explains the enzymatic transfer of calcium across cell membranes via CICR ("calcium induced by calcium release"). The model consists of 4 chemical species and 6 reactions (eqs. (9.1)–(9.6) below), where S_i represent the chemical species and x_i represent the corresponding concentrations, and k_{ij} are the parameters (experimental rate constants):

$$0 \xrightarrow{k_{21}} S_1 \tag{9.1}$$
$$S_1 \xrightarrow{k_{12}} 0 \tag{9.2}$$
$$S_1 + S_2 \xrightarrow{k_{43}} 2S_1 \tag{9.3}$$
$$S_1 + S_3 \xrightarrow{k_{65}} S_4 \tag{9.4}$$
$$S_4 \xrightarrow{k_{56}} S_1 + S_3 \tag{9.5}$$
$$S_4 \xrightarrow{k_{76}} S_2 + S_3 \tag{9.6}$$

The resulting system of differential equations with quadratic nonlinearities is given by the following equations:

$$dx_1/dt = -k_{12}x_1 + k_{21} + k_{43}x_1x_2 + k_{56}x_4 - k_{65}x_1x_3 \tag{9.7}$$
$$dx_2/dt = -k_{43}x_1x_2 + k_{76}x_4 \tag{9.8}$$
$$dx_3/dt = k_{56}x_4 - k_{65}x_1x_3 + k_{76}x_4 \tag{9.9}$$
$$dx_4/dt = -k_{56}x_4 + k_{65}x_1x_3 - k_{76}x_4 \tag{9.10}$$

There is a conservation law in this model, namely

$$x_3 + x_4 = c_1, \tag{9.11}$$

where c_1 is a constant. We now handle the conservation law as described in Sect. 3.1(b). We eliminate one variable (for example, x_3). Then system (9.7)–(9.10) reduces to

$$dx_1/dt = k_{43}x_1x_2 - k_{65}x_1(-x_4 + c_1) - k_{12}x_1 + k_{56}x_4 + k_{21}, \tag{9.12}$$
$$dx_2/dt = -k_{43}x_1x_2 + k_{76}x_4, \tag{9.13}$$
$$dx_4/dt = k_{65}x_1(-x_4 + c_1) - k_{56}x_4 - k_{76}x_4. \tag{9.14}$$

Now we implement the steps mentioned in Sect.6.1 for eqs. (9.12)–(9.14).

9.1.2. Step 1

A monomial in $x_1, x_2, \ldots, x_n$ is a product of form $x_1^{a_1}x_2^{a_2}\ldots x_n^{a_n}$ where all the exponents $a_1, a_2, \ldots, a_n$ are non negative integers. The total degree of this monomial is the sum $a_1 + a_2 + \ldots + a_n$. As the input system is quadratic the monomials have degrees ≤ 2. The monomials of total degree 2 were used for the vertex cover computation. To separate the monomials the right hand side of the system was factorised into a symbolic matrix and a vector of monomials as shown below:

$$\begin{bmatrix} k_{21} & -c_1k_{65} - k_{12} & k_{56} & k_{43} & k_{65} \\ 0 & 0 & k_{76} & -k_{43} & 0 \\ 0 & c_1k_{65} & -k_{56} - k_{76} & 0 & -k_{65} \end{bmatrix} \tag{9.15}$$

$$\begin{bmatrix} 1 \\ x_1 \\ x_4 \\ x_1\,x_2 \end{bmatrix} \tag{9.16}$$

The vertex cover of the undirected graph is computed by an greedy algorithm. We use the decomposition

$$y = \{x_1\}, \quad z = \{x_2, x_4\}.$$

Let us consider the subsystem for z variables, i.e., equations for $dx_2/dt, dx_4/dt$:

$$\begin{aligned} dx_2/dt &= -k_{43} x_1 x_2 + k_{76} x_4, \\ dx_4/dt &= k_{65} x_1 (-x_4 + c_1) - k_{56} x_4 - k_{76} x_4. \end{aligned}$$

We set $dx_2/dt, dx_4/dt = 0$, resolve these equations with respect to x_2, x_4 and find an equilibrum $x_2(x_1), x_4(x_1)$:

$$\left\{ x_2 = \frac{c_1\ k_{65}\ k_{76}}{k_{43}\ (k_{65}\ x_1 + k_{56} + k_{76})}, x_4 = \frac{k_{65}\ x_1\ c_1}{k_{65}\ x_1 + k_{56} + k_{76}} \right\}.$$

The vertex cover system is constructed as above for eq. (6.10), i.e., by substituting eq. (9.1.2) in eqs. (9.12)–(9.14):

$$\begin{aligned} dx_1/dt = {} & \frac{k_{76}\ k_{65}\ x_1\ c_1}{k_{65}\ x_1 + k_{56} + k_{76}} - k_{65}\ x_1 \left(-\frac{k_{65}\ x_1\ c_1}{k_{65}\ x_1 + k_{56} + k_{76}} + c_1 \right) - k_{12}\ x_1 \\ & + \frac{k_{56}\ x_1\ c_1\ k_{65}}{k_{65}\ x_1 + k_{56} + k_{76}} + k_{21}, \\ dx_2/dt = {} & 0, \\ dx_4/dt = {} & k_{65}\ x_1 \left(-\frac{k_{65}\ x_1\ c_1}{k_{65}\ x_1 + k_{56} + k_{76}} + c_1 \right) - \frac{k_{56}\ x_1\ c_1\ k_{65}}{k_{65}\ x_1 + k_{56} + k_{76}} \\ & - \frac{k_{76}\ k_{65}\ x_1\ c_1}{k_{65}\ x_1 + k_{56} + k_{76}}. \end{aligned}$$

9.1.3. Step 2

Solving the vertex cover system for equilibrium with respect to y variables, finally we obtain an equilibrium

$$\left\{ x_1 = \frac{k_{21}}{k_{12}} \right\}. \tag{9.17}$$

9.2. Model of ERK activation

9.2.1. Input System

This model investigates the extracellular signal-regulated kinase (ERK) signaling pathway by angiotensin II type 1A receptor. This signaling pathway is involved in cell surface receptors and hence is regarded as a drug target. Therefore, understanding such a pathway mechanism is a key problem. The model presented in [9] consists of 18 chemical species and 26 chemical reactions. The model represented in Systems Biology Markup Language (SBML) can be obtained from `http://www.nature.com/msb/journal/v8/n1/suppinfo/msb201222_S1.html#msb201222-s1`. The corresponding system of differential equations

from the SBML was computed using the PoCab system environment [17]. This model has the form

$$
\begin{aligned}
dx_1/dt &= -k_{10}x_1 - k_{18}x_1 - k_{19}x_{11}x_1 + k_{15}x_{14} + k_{16}x_{15} + k_{17}x_{16} + k_{23}x_{17}\\
dx_2/dt &= -k_1x_2x_{13} - k_2x_2x_1 - k_0x_2 + k_6x_{12}\\
dx_3/dt &= -k_5x_3x_7 - k_{21}x_3x_{17} - k_{22}x_3x_{18} + k_9x_8 + k_{25}x_9\\
dx_4/dt &= -k_3x_{12}x_4 + k_7x_5\\
dx_5/dt &= k_3x_{12}x_4 - k_7x_5\\
dx_6/dt &= -k_4x_5x_6 + k_8x_7\\
dx_7/dt &= k_4x_5x_6 - k_8x_7\\
dx_8/dt &= k_5x_3x_7 - k_9x_8\\
dx_9/dt &= k_{21}x_3x_{17} + k_{22}x_3x_{18} - k_{25}x_9\\
dx_{10}/dt &= -k_{11}x_{10}x_{13} + k_{13}x_{14} + k_{15}x_{14}\\
dx_{11}/dt &= -k_{12}x_{11}x_{13} - k_{19}x_{11}x_1 - k_{20}x_{11}x_{16}\\
&\quad +k_{14}x_{15} + k_{16}x_{15} + k_{23}x_{17} + k_{24}x_{18}\\
dx_{12}/dt &= k_1x_2x_{13} + k_2x_2x_1 + k_0x_2 - k_6x_{12}\\
dx_{13}/dt &= k_{10}x_1 - k_{11}x_{10}x_{13} - k_{12}x_{11}x_{13} + k_{13}x_{14} + k_{14}x_{15}\\
dx_{14}/dt &= k_{11}x_{10}x_{13} - k_{13}x_{14} - k_{15}x_{14}\\
dx_{15}/dt &= k_{12}x_{11}x_{13} - k_{14}x_{15} - k_{16}x_{15}\\
dx_{16}/dt &= k_{18}x_1 - k_{20}x_{11}x_{16} - k_{17}x_{16} + k_{24}x_{18}\\
dx_{17}/dt &= k_{19}x_{11}x_1 - k_{23}x_{17}\\
dx_{18}/dt &= k_{20}x_{11}x_{16} - k_{24}x_{18}
\end{aligned}
$$

The model has 7 conservation laws:

$$
\begin{aligned}
x_2 + x_{12} = C_1, \quad x_1 + x_{13} + x_{14} + x_{15} + x_{17} + x_{16} + x_{18} = C_2\\
x_4 + x_5 = C_3, \quad x_6 + x_7 = C_4, \quad x_{10} + x_{14} = C_5\\
x_{11} + x_{15} + x_{17} + x_{18} = C_6, \quad x_3 + x_9 + x_8 = C_7
\end{aligned}
$$

The conservations laws were taken into account and the pre-processing step was done analogously to the previous example. The variables that were eliminated are $x_1, x_2, x_3, x_4, x_6, x_{10}, x_{11}$. Therefore, the input system becomes a 11 dimensional model.

9.2.2. Step 1 and Step 2

The vertex cover from the 11 dimensional model is computed as in the previous example and variables are decomposed into y and z variables as shown below:

$$
\begin{aligned}
y &= \{x_7, x_{12}, x_{13}, x_{15}, x_{17}, x_{18}\}\\
z &= \{x_5, x_8, x_9, x_{14}, x_{16}\}
\end{aligned}
$$

The z sub-system consists of equations for $x_5, x_8, x_9, x_{14}, x_{16}$ and these equations are solved with respect to z variables. The resulting rational functions as well as the vertex cover system is constructed similarly to the previous example. We fail to solve the vertex cover system with respect to y variables in symbolic form using the computer algebra system Maple, since its solver didn't find solutions in 2000 seconds of computation time. To overcome this difficulty, we have used the numeric values of parameters and conservation law constants from [9]. Finally, the following single equilibrium is obtained:

$$
\begin{aligned}
\{x_7 = 0.04184, x_{12} = 0.01184, x_{13} = 0.00022, x_{15} = 0.00270,\\
x_{17} = 0.0000091, x_{18} = 0.00189\}\,.
\end{aligned}
$$

9.3. Model of MAPK cascade

9.3.1. Input System

We consider the polynomial equation system of single layer Mitogen-activated protein kinase cascade system (MAPK) from [22]. The model is a well studied biological network that displays the property of multistationarity i.e. more than one steady state solution for fixed parameters and conservation constants. Such properties are also well documented in cell cycle and signal transduction networks in biology. Below, we show that the multistationarity of given nine dimensional MAPK cascade system is preserved in the vertex cover system with a dimension of just two.

The input system is given below

$$\begin{aligned}
dx_1/dt &= -k_1x_1x_2 + k_{12}x_9 + k_2x_3\\
dx_2/dt &= -k_1x_1x_2 - k_4x_2x_4 + k_2x_3 + k_3x_3 + k_5x_5 + k_6x_5\\
dx_3/dt &= k_1x_1x_2 - k_2x_3 - k_3x_3\\
dx_4/dt &= -k_{10}x_4x_7 - k_4x_2x_4 + k_{11}x_9 + k_3x_3 + k_5x_5 + k_9x_8\\
dx_5/dt &= k_4x_2x_4 - k_5x_5 - k_6x_5\\
dx_6/dt &= -k_7x_6x_7 + k_6x_5 + k_8x_8\\
dx_7/dt &= -k_{10}x_4x_7 - k_7x_6x_7 + k_{11}x_9 + k_{12}x_9 + k_8x_8 + k_9x_8\\
dx_8/dt &= k_7x_6x_7 - k_8x_8 - k_9x_8\\
dx_9/dt &= k_{10}x_4x_7 - k_{11}x_9 - k_{12}x_9
\end{aligned}$$

The model has three conservation laws:

$$\begin{aligned}
x_2 + x_3 + x_5 &= C_1\\
x_7 + x_8 + x_9 &= C_2\\
x_1 + x_3 + x_4 + x_5 + x_6 + x_8 + x_9 &= C_3
\end{aligned}$$

9.3.2. Step 1 and Step 2

The conservation laws were handled as per Sect. 3.1(a). The vertex cover from the 9 dimensional model is computed as in the previous example and variables are decomposed into y and z variables as shown below

$$\begin{aligned}
y &= \{x_2, x_7\}\\
z &= \{x_1, x_3, x_4, x_5, x_6, x_8, x_9\}
\end{aligned}$$

The z sub-system consists of equations for $x_1, x_3, x_4, x_5, x_6, x_8, x_9$ and these equations are solved with respect to z variables along with the conservation law for z sub-system (i.e., $x_1 + x_3 + x_4 + x_5 + x_6 + x_8 + x_9 - C_3$). The resulting rational functions as well as the vertex cover system is constructed similar to the previous example. We could solve for the symbolic solution using Maple resulting in five solutions (the expressions are not presented due to space constraints). The solutions converted to float using the parameters values and conservation constants obtained from [22] are as follows:

$$\begin{aligned}
x_2 &= 0.4118, \quad x_7 = 0.0097\\
x_2 &= 0.4260, \quad x_7 = 0.0136\\
x_2 &= 1.1584, \quad x_7 = 0.7460\\
x_2 &= -3.1805, \quad x_7 = -9.6338\\
x_2 &= 32.0989, \quad x_7 = -16.0470
\end{aligned}$$

The vertex cover system has 3 positive biologically interesting solutions consistent with experimental findings of multistationarity in the original MAPK cascade system.

10. Conclusion and discussion

In this paper, we have considered a class of systems of differential equations with quadratic nonlinearities modelling biochemical networks. We have presented algorithms, which allow to describe low dimensional reductions of dynamics of these systems. These algorithms need no numerical integration of stiff systems and include the following ingredients: **i**) algebraic methods to find so-called vertex covers; **ii**) the theory of slow invariant manifolds; and **iii**) some special approximations. Moreover, we have shown, for some important class of systems of biochemical kinetics, that the large time behavior of a large network may be very complicated. In a sense, any given dynamics can be realized, within arbitrarily small precision, by a sufficiently large network. The second theorem asserts that even a *fixed* large quadratic kinetic system with $N \gg 1$ variables and random kinetic rates can simulate different smaller systems with $n \ll N$ variables. In a sense, a random big system from our class can simulate a number of low dimensional ones belonging to the same class. This set of low dimensional dynamics can correspond to adaptive reactions of the organism. An algorithm producing such simulations is simple and it can be easily performed by gene control. Namely, this gene control can modify some kinetic rate coefficients describing reagent decays. This increase of the decay rates turns off the corresponding reagents. The rest reagents perform a needed low dimensional dynamics.

As it was noticed in a famous textbook on biochemistry [14], the metabolisms of mammals and mushrooms are wonderfully similar. Even the number of reagents involved in the human and the yeast metabolisms do not differ very dramatically and have the same order. The results of this paper possibly shed a light on this wonderful phenomenon. In fact, one can assume that evolution could use any random metabolic network. According to our results, this network can perform any dynamical adaptive answer under appropriate gene control.

Acknowledgements. The third author was supported by the grant of Russian Ministry of Education, 2012-1.2.1-12-000-1013-016 and his work was done in the in the frame of the project We 1945/7-1 funded by *Deutsche Forschungsgemeinschaft*. The third author was also financially supported by Government of Russian Federation, Grant 074-U01. The first author is grateful to Max-Planck Institut für Mathematik, Bonn for its hospitality.

References

[1] D. Bartel. MicroRNAs: target recognition and regulatory functions. *Cell*, 136:215–233, 2009.

[2] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, 2nd edition, 2001.

[3] E. N. Dancer and P. Poláčik. Realization of vector fields and dynamics of spatially homogeneous parabolic equations. *Memoirs of Amer. Math. Society*, 140(668), 1999.

[4] H. Errami, W. M. Seiler, M. Eiswirth, and A. Weber. Computing Hopf bifurcations in chemical reaction networks using reaction coordinates. In V. P. Gerdt, W. Koepf, E. W. Mayr, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing*, volume 7442 of *Lecture Notes in Computer Science*, pages 84–97. Springer Berlin Heidelberg, 2012.

[5] M. Gary and D. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. H. Freeman, New York, 1979.

[6] K. Gatermann, M. Eiswirth, and A. Sensse. Toric ideals and graph theory to analyze Hopf bifurcations in mass action systems. *Journal of Symbolic Computation*, 40(6):1361–1382, 2005.

[7] H. Haken. *Synergetics—An Introduction*. Springer, Berlin, 3rd edition, 1983.

[8] A. Halmschlager, L. Szenthe, and J. Tóth. Neural networks and physical systems with emergent collective computational abilities. *Electronic Journal of Qualitative Theory of Differential Equations*, 14:1–14, 2004.

[9] D. Heitzler, G. Durand, N. Gallay, A. Rizk, S. Ahn, J. Kim, J. D. Violin, L. Dupuy, C. Gauthier, V. Piketty, P. Crépieux, A. Poupon, F. Clément, F. Fages, R. J. Lefkowitz, and E. Reiter. Competing G protein-coupled receptor kinases balance G protein and β-arrestin signaling. *Molecular systems biology*, 8(590):590, 2012.

[10] D. Henry and D. B. Henry. *Geometric theory of semilinear parabolic equations*. Lecture Notes in Mathematics. Springer, Berlin, 1981.

[11] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:641–654, 2000.

[12] M. D. Korzuchin. Kolebaltelnie processi v biol. i chimich. sistemach. Thesis. 231 pp. Moscow, 1967. In Russian.

[13] N. V. Kuznetsov and G. A. Leonov. Lyapunov quantities and limit cycles of two-dimensional dynamical systems. In *Dynamics and Control of Hybrid Mechanical Systems*, pp. 7–28. World Scientific, 2010.

[14] A. L. Lehninger, D. L. Nelson, and M. M. Cox. *Principles of Biochemistry.* Worth, New York, 2nd edition, 1993.

[15] A. Lesne. Complex networks: from graph theory to biology. *Letters in Mathematical Physics*, 78:235–262, 2006.

[16] D. Ruelle and F. Takens. *Elements of differentiable dynamics and bifurcation theory.* Academic Press, Boston, 1989.

[17] S. S. Samal, H. Errami, and A. Weber. PoCaB: a software infrastructure to explore algebraic methods for bio-chemical reaction networks. In V. P. Gerdt, W. Koepf, E. W. Mayr, and E. V. Vorozhtsov, editors, *Computer Algebra in Scientific Computing*, volume 7442 of *Lecture Notes in Computer Science*, pages 294–307. Springer Berlin Heidelberg, 2012.

[18] Y. Shimoni, G. Friedlander, and G. Hetzroni. Regulation of gene expression by small non-coding RNAs: a quantitative view. *Molecular Systems* , vol. 3, N1, 2007.

[19] A. M. Zhabotinsky. *Konzentrzionnie avtokolebania.* Nauka, Moscow, 1974. in Russian.

[20] J. Zhao, H. Yu, J. Luo, Z. W. Cao, and Y. X. Li. Hierarchical modularity of nested bow-ties in metabolic networks. *BMC bioinformatics*, 7, 2006.

[21] V. Zhdanov. Kinetic models of gene expression including non-coding RNAs. *Physics Reports*, 500:1–42, 2011.

[22] Conradi, C., Flockerzi, D., & Raisch, J. (2008). *Multistationarity in the activation of a MAPK: parametrizing the relevant region in parameter space.* Mathematical biosciences, 211(1), 105-31, doi:10.1016/j.mbs.2007.10.004.