Layered decomposition for the model order reduction of timescale separated biochemical reaction networks

CrossMark

Thomas P. Prescott [a,b], Antonis Papachristodoulou [a,*]

[a] *Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, United Kingdom*
[b] *Life Sciences Interface Doctoral Training Centre, University of Oxford, Parks Road, Oxford OX1 3QU, United Kingdom*

H I G H L I G H T S

- Layering reaction networks is an alternative to modular decomposition.
- Layering identifies timescale-separated variables without transformations.
- Fast and slow subsystems can be expressed in terms of original variables.
- Approximated dynamics are interpreted as summed contributions from fast and slow layers.

A R T I C L E   I N F O

A B S T R A C T

Biochemical reaction networks tend to exhibit behaviour on more than one timescale and they are inevitably modelled by stiff systems of ordinary differential equations. Singular perturbation is a well-established method for approximating stiff systems at a given timescale. Standard applications of singular perturbation partition the state variable into fast and slow modules and assume a quasi-steady state behaviour in the fast module. In biochemical reaction networks, many reactants may take part in both fast and slow reactions; it is not necessarily the case that the reactants themselves are fast or slow. Transformations of the state space are often required in order to create fast and slow modules, which thus no longer model the original species concentrations. This paper introduces a layered decomposition, which is a natural choice when reaction speeds are separated in scale. The new framework ensures that model reduction can be carried out without seeking state space transformations, and that the effect of the fast dynamics on the slow timescale can be described directly in terms of the original species.

1. Introduction

A characteristic of many complex biological systems, including biochemical reaction networks, is that dynamical processes can occur on multiple, vastly different timescales (Southern et al., 2008; Jamshidi and Palsson, 2008). Modelling such a system by taking into account all of the timescales simultaneously results in stiff models. Simulation of these models requires fine temporal resolution to capture the fastest dynamics. If we are only interested in the behaviour of the system on a slower timescale, singular perturbation (Hinch, 1991; Murray, 2002) is a well-established model approximation technique which aims to capture the important dynamics at that timescale and remove the stiffness arising from the faster dynamics, making simulation and analysis more practical.

Fig. 1 represents the method of model reduction through singular perturbation as a sequence of steps. The first step is the identification of non-dimensional fast and slow variables, by which the system can be described by interconnected dynamic subsystems. The subsystem corresponding to the timescale which is not of interest is then approximated to be a static subsystem. The final step is the expression of the approximated system in terms of the original variables. In addition to approximating a stiff model with a non-stiff model, singular perturbation is also a model reduction technique. The reduction occurs because replacing a dynamic subsystem with a static subsystem (the second step in Fig. 1) removes differential equations from the recomposed system.

Grouping reactants into fast and slow subsets is an example of *modular* decomposition (Alexander et al., 2009). In the case of biochemical networks, this classification may be difficult. The variables are concentrations of reactants, but often the parameters are such that it is instead the reaction rate constants which separate in scale. If one reactant takes part in both fast and slow

* Corresponding author. Tel.: +44 1865 273000.
E-mail address: antonis@eng.ox.ac.uk (A. Papachristodoulou).
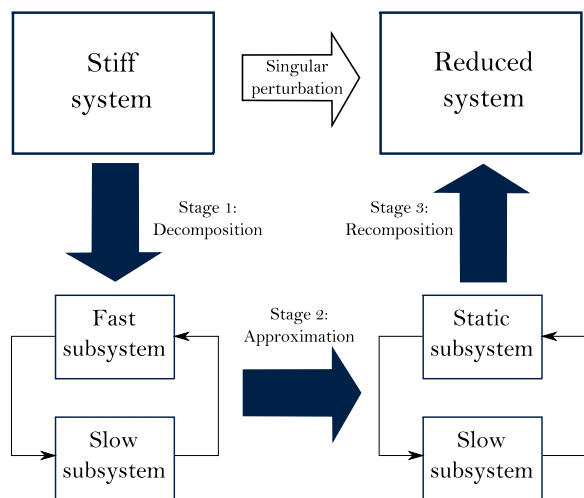
Fig. 1. The approximation of a stiff system through singular perturbation is broken down into three stages. First the system is decomposed into an interconnection of fast and slow subsystems, discussed in Section 2. In the second stage, discussed in Section 3, the fast subsystem is approximated by a static subsystem. Finally, the entire singularly perturbed system is formed from the recomposition of the original slow subsystem with the new static subsystem.

reactions, it cannot be classified as either fast or slow (Bennett et al., 2007; Lee and Othmer, 2010).

Up to now, the remedy for this has been to find a transformation of the state space to create pooled variables (Kumar et al., 1998). These are new variables, formed out of combinations of the reactant concentrations, whose dynamics allow them to be classified as fast or slow. For example, the well-known QSS approximation of the elementary enzyme reaction network is not valid when the enzyme concentration is on the same scale as that of the other reactants. The tQSS approximation (Tzafriri, 2003; Ciliberto et al., 2007) creates a new variable as a linear combination of two physical quantities that, for a given parameter range, allows a valid classification of variables as fast or slow. The improvement in approximation error (Prescott and Papachristodoulou, 2012) by taking the tQSS approach is at the expense of defining a new variable modelling a total concentration, rather than one of the original concentrations of interest.

This paper aims to perform singular perturbation of biochemical networks while avoiding finding transformations of the state variables. This will simplify the procedure of singular perturbation and also create fast and slow subsystems which have a biochemical interpretation in terms of the original variables. This is done by first reconsidering the decomposition of the system into subsystems in Section 2. It will be shown that an alternative decomposition method, termed *layering* (Prescott and Papachristodoulou, 2013), completes step 1 in Fig. 1 without requiring any transformations. We will then show in Section 3 how the approximation arising from such a decomposition results in reduced models, demonstrating our methods on two biological examples.

2. Decomposing biochemical reaction networks

This paper will focus on a standard ODE model for biochemical reaction networks (Palsson, 2006). Suppose such a network consists of N reactants X_i taking part in M reactions. The concentration of each reactant, denoted $x_i(t)$, is a component in the state variable $x(t) \in \mathbb{R}^N$. The rate at which reaction j occurs is denoted $v_j(x)$, where the argument is x to make explicit that each reaction rate may depend on the concentrations of any of the reactants at each point in time. The reaction rates form the components of the flux

vector $v(x) \in \mathbb{R}^M$. The stoichiometric coefficient of X_i in the jth reaction is denoted S_{ij}, which represents the change in concentration of X_i as reaction j proceeds. Each coefficient S_{ij} forms the stoichiometric matrix $S \in \mathbb{R}^{N \times M}$. These structures combine into a system of N coupled ODEs

$$\frac{dx}{dt} = Sv(x) \qquad (1)$$

where the initial conditions are denoted $x(0) = x_0$.

In this section we will discuss how singular perturbation techniques can be interpreted as a decomposition of stiff systems into interconnected subsystems of characteristic time scales, as in Fig. 1. With this interpretation, we will show that established methods involving transformations correspond to a modular decomposition of the system. We then introduce a complementary decomposition, termed *layering*, which results in an alternative method for singular perturbation that avoids transformations and thus retains the natural coordinates (i.e. species concentrations) for each subsystem.

2.1. Modular decomposition of timescale-separated systems

The first stage of the approximation of a stiff non-dimensionalised ODE system through singular perturbation is the attempt to classify variables (i.e. reactants) as fast or slow. Suppose that the N components of the state variable x can be partitioned into two groups, where the components of $x_f \in \mathbb{R}^{N_f}$ are N_f fast reactants and the components of $x_s \in \mathbb{R}^{N_s}$ are N_s slow reactants. Assuming a re-ordering of the reactants index, we can write $x = (x_f^T, x_s^T)^T$. The classification into fast and slow reactants can be justified through the existence of a positive parameter $\epsilon \ll 1$ such that the non-dimensional system (1) can be written

$$\begin{bmatrix} \epsilon \dot{x}_f \\ \dot{x}_s \end{bmatrix} = \begin{bmatrix} S_f \\ S_s \end{bmatrix} v(x_f, x_s),$$

where the rows of S have been re-ordered and partitioned appropriately to the components of $x = (x_f^T, x_s^T)^T$. This representation makes clear the decomposition of the system into an interconnection of fast and slow modules:

$$\epsilon \dot{x}_f = S_f v(x_f, x_s),$$
$$\dot{x}_s = S_s v(x_f, x_s).$$

The full system state x can be recovered from the modular system states through the concatenation $x = (x_f^T, x_s^T)^T$.

However, often the parameters of the dynamical system are such that there is no decomposition of the state vector $x = (x_f, x_s)$ into variables with timescale-separated dynamics. For example, consider the enzyme kinetics network

$$X + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \overset{k_2}{\to} Y + E$$
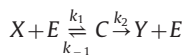
modelling enzyme E converting X to Y via an intermediate complex C. If the initial concentration $[E(0)] = e(0)$ of enzyme is on a much smaller scale than that of the substrate $[X(0)] = x(0)$ then the non-dimensionalised variables do separate in timescale through the standard quasi-steady state (QSS) approximation. However, if $e(0)$ is on a similar scale to that of $x(0)$, it is no longer valid to assume that the non-dimensionalised variables separate in timescale, and another approximation must be found.

The standard technique in this case is to find a linear transformation T of the state vector so that the new state $z = Tx$ can be decomposed into fast and slow variables $z = (z_f, z_s)$. In general the required transformations are difficult to calculate, especially in the case of large-scale systems, and are often driven by intuition. Even when found, transformations of the state vector create linear
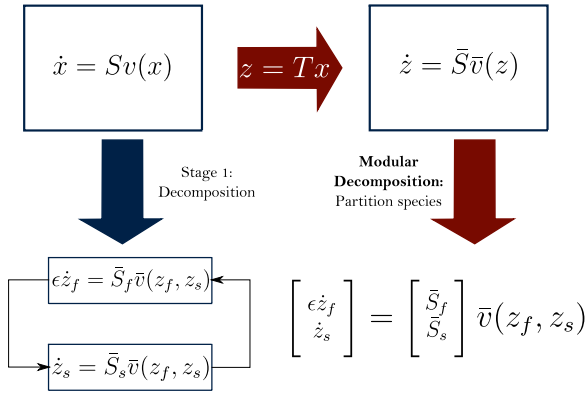
Fig. 2. Modular decomposition (modules indexed in subscript). Classify the reactant concentrations as fast or slow variables. If such a classification cannot be found then first transform the state vector into $z = Tx$ so that the components of z can be classed as fast or slow. Each group of components forms the state of a subsystem.

Fig. 3. Layered decomposition (layers indexed in superscript). Classify reactions as fast or slow. Create two variables x^f and x^s; the first evolves according to the fast reactions, the second according to the slow reactions. The evolution of each variable makes up a subsystem, and no transformations are required.

combinations of individual state components. This means that the resulting modules are expressed in coordinates which no longer represent the reactants originally modelled.

The technique described above, summarised in Fig. 2, is a *modular* decomposition, where the decomposition step in Fig. 1 is achieved by partitioning the state vector. We will now describe an alternative *layered* approach (Prescott and Papachristodoulou, 2013) to the decomposition of (1).

2.2. Layered decomposition

We now assume that, while the system of interest exhibits timescale separation, we cannot classify the non-dimensionalised variables (i.e. reactant concentrations) as fast or slow. Suppose instead that it is the reaction rate constants defining the M components of $v(x)$ which are separated in scale. In what follows we will define fast and slow subsystems through a layered decomposition, summarised in Fig. 3.

Our assumption that the reaction rates separate in scale means that we assume the existence of a positive parameter $\epsilon \ll 1$ allowing the partition of the components of v into $v = (v^f/\epsilon, v^s)^T$, corresponding to the fast and slow groups of reaction rates respectively.[1] As with x in the modular decomposition, we assume an ordering of the components of v to allow this representation. Here v^f and v^s are on the same scale, and the scale separation is entirely a consequence of ϵ. Partitioning the columns of $S = [S^f, S^s]$ conformally with the components of v means that we can write (1) as

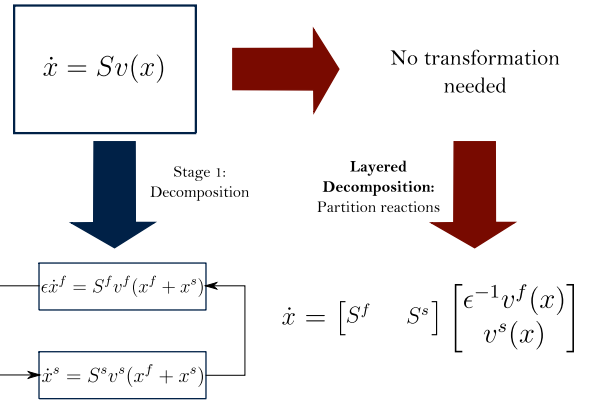$$\dot{x} = \frac{1}{\epsilon} S^f v^f(x) + S^s v^s(x). \tag{2}$$

A decomposition of the system dynamics into the sum of two vector fields is the key idea behind layered decomposition.

We now decompose (2) into two interconnected subsystems by defining two states x^f and x^s, each taking values in $\mathbb{R}^N$, with dynamics

$$\epsilon \dot{x}^f = S^f v^f(x^f + x^s), \tag{3}$$

$$\dot{x}^s = S^s v^s(x^f + x^s), \tag{4}$$

and initial conditions chosen such that $x^f(0) + x^s(0) = x_0$. The recovery of the full state x is no longer by concatenation of the subsystem states (as in the modular case), but instead through the summation

$x = x^f + x^s$ of the layered states. It is clear from summing (3) and (4) that this recovers the full system dynamics (2).

We now have fast and slow variables x^f and x^s, without requiring any transformation of the state space. However, both variables take values in $\mathbb{R}^N$, contrasting with the modular decomposition into two components $x_i \in \mathbb{R}^{N_i}$ in spaces of dimension $N_i < N$. We will show in Section 2.4 that the layered decomposition nevertheless results in subsystems of lower dimension. This will be used in Section 3 to show that a layered approach to singular perturbation does result in model order reduction in addition to removing the model's stiffness.

2.3. Example

We now illustrate the two different approaches to network decomposition on the enzyme reaction kinetics example above. Assume that the enzyme concentration e is on the same scale as the substrate x and product y concentrations, and suppose that the reversible binding reaction is on a much faster timescale than the conversion of complex to product:

$$X + E \underset{k_{-1}/\epsilon}{\overset{k_1/\epsilon}{\rightleftharpoons}} C \overset{k_2}{\rightarrow} Y + E$$
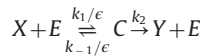
Due to the conservation relation $e + c = K$ for constant K and the decoupling of y from the system, an ODE model for this reaction network is the two-dimensional system

$$\begin{bmatrix} \dot{x} \\ \dot{c} \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} k_1 x(K-c)/\epsilon \\ k_{-1}c/\epsilon \\ k_2 c \end{bmatrix} \tag{5}$$

where $\epsilon \ll 1$ is a small parameter which makes the reversible binding/unbinding reactions much faster than the conversion of complex C into product Y. While x is clearly a fast variable, since its dynamics are on the timescale $1/\epsilon$, the timescale of c cannot be easily classified, as it takes part both in fast and slow reactions.

At this point, the two possible decomposition techniques diverge. The modular decomposition resulting from the tQSS approximation (Tzafriri, 2003) requires a transformation to the original system state $[x, c]$ to give new variables $[x, c] \mapsto [x, x + c]$. It can be seen by summing the rows in (5) that the new variable $z := x + c$ has only slow dynamics. The modular decomposition of the dynamics thus gives two interconnected one-dimensional subsystems with dynamics

$$\epsilon \dot{x} = -k_1 x(K + x - z) + k_{-1}(z - x), \tag{6}$$

$$\dot{z} = -k_2(z - x). \tag{7}$$

---

[1] We use superscript to index layers, and subscript for modules.

The remainder of the singular perturbation method proceeds by approximating the fast module (6) with a static system.

We now demonstrate layered decomposition. Consider again the flux vector in (5). The first two components are separated in scale from the final component, which means that we can write (5) in the form of (2) as

$$\begin{bmatrix} \dot{x} \\ \dot{c} \end{bmatrix} = \frac{1}{\epsilon}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} k_1 x(K-c) \\ k_{-1}c \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix}k_2 c.$$

Putting each term of this expression into the layered decomposition (3)–(4), the layers' dynamics are

$$\epsilon\begin{bmatrix} \dot{x}^f \\ \dot{c}^f \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} k_1 x(K-c) \\ k_{-1}c \end{bmatrix} \tag{8}$$

$$\begin{bmatrix} \dot{x}^s \\ \dot{c}^s \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}k_2 c \tag{9}$$

where $x = x^f + x^s$ and $c = c^f + c^s$. With this decomposition we have constructed a fast layer (8) and a slow layer (9) without resorting to a transformation.

Fig. 4 shows how the layers' trajectories overlay through summation to give the overall system dynamics, using the example of $c = c^f + c^s$. The interpretation of c^f and c^s is that each is the contribution to the dynamics of c arising from each layer. In the transient, where $t \ll 1$, reactant C is quickly produced by the fast layer through the fast equilibration of the reversible reaction $X + E \rightleftharpoons C$.

However, at the slower timescale, the contribution of the slow layer begins to decrease from its initial condition of $c^s(0) = 0$. This is because the slow reaction $C \rightarrow Y + E$ consumes reactant C. Therefore the slow layer acts to decrease the value of c by reducing its contribution c^s. Conversely, C is produced by the fast layer through the equilibration of $X + E$ with a decreasing amount of reactant C. Thus the contribution $c^f(t)$ continues to increase, partially off-setting the decrease in $c^s(t)$. Both layers eventually reach steady state, with $c^f + c^s \rightarrow 0$.

The layered system (8)–(9) illustrates the dimension problem mentioned above and also its solution, discussed in Section 2.4.

Fig. 4. The solid black line is the simulation of $c(t)$ according to (5) from initial condition $c(0) = 0$ with parameters $K = k_1 = k_2 = 1$, $k_{-1} = 2$ and $\epsilon = 0.05$. The broken lines correspond to the layered states $c^f(t)$ and $c^s(t)$ according to (8)–(9), whose sum recovers $c(t)$. Note that the fast transient in $t \ll 1$ arises in the fast layer only.

In the modular decomposition (6)–(7), the fast and slow modules are each one-dimensional subsystems. However, the layered decomposition (8)–(9) results in two subsystems whose states $[x^f, c^f]$ and $[x^s, c^s]$ each takes values in $\mathbb{R}^2$. It appears that layering expands the state space by creating two two-dimensional systems. However, the layers are in fact one-dimensional systems *embedded* in $\mathbb{R}^2$. The reduced dimension of each layer is due to additional conservation relations introduced by the decomposition. We now discuss how to ensure that the subsystems in a layered decomposition are of lower dimension.

2.4. Dimensions of layers

First consider the complete biochemical reaction network ODE model (1). Although the state x of this reaction network evolves in $\mathbb{R}^N$, the dimension of the system is not determined by N.

Let $r \le \min\{N, M\}$ denote the rank of the stoichiometric matrix S. Consider any decomposition of $S = UC$ into two matrices where $U \in \mathbb{R}^{N \times r}$ is full column rank and $C \in \mathbb{R}^{r \times M}$ is full row rank. The existence of such a decomposition is proved in the following lemma.

Lemma 1. *For any $S \in \mathbb{R}^{N \times M}$ with* rank$(S) = r$ *there exists a decomposition $S = UC$ where $U \in \mathbb{R}^{N \times r}$ and $C \in \mathbb{R}^{r \times M}$ are full rank.*

Proof. The existence of a decomposition $S = UC$ can be proved constructively using the compact singular value decomposition of $S = U\Sigma V^T$. Here the columns of $U \in \mathbb{R}^{N \times r}$ form a basis for the column space of S, the columns of $V \in \mathbb{R}^{M \times r}$ form a basis for the row space of S, and $\Sigma \in \mathbb{R}^{r \times r}$ is a diagonal matrix of the r non-zero singular values. Writing $C = \Sigma V^T$, we have the required decomposition $S = UC$. $\square$

Since $\dot{x} = Sv$ it follows that the translated trajectory $x(t) - x(0) \in$ Col(S) is in the image space of U. Since U is full column rank, at each $x(t)$ there is a unique $\xi(t) \in \mathbb{R}^r$ such that $x = x_0 + U\xi$. The dynamics in (1) can then be re-written using the internal "state" vector ξ as
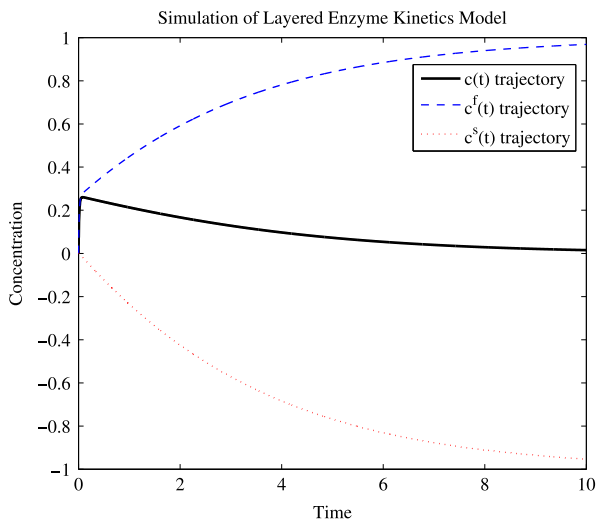
$$\dot{\xi} = Cv(x_0 + U\xi), \tag{10}$$

$$x = x_0 + U\xi. \tag{11}$$

This realisation of (1) shows that the dimension of any reaction network is determined by $r = $ rank(S).

We can now use this dimensionality argument to show how the layered decomposition (3)–(4) results in subsystems which are of smaller dimension than the original system. The dimension of the fast layer $\dot{x}^f = S^f v^f(x)$ is determined by $r^f = $ rank(S^f). Similarly rank$(S^s) = r^s$ is the dimension of the slow layer. As long as $r^f < r$ and $r^s < r$, each layer will therefore be of a lower dimension than the original system.

One interpretation of this argument follows from considering the basis vectors of the left null space of S, which correspond to $N - r$ conservation relations in the original system (1). Since the left null-vectors of $S = [S^f, S^s]$ must also be left null-vectors of S^f and S^s, the same conservation relations hold in each layer. However there are now $r - r^s$ additional conservation relations within the slow layer, and similarly $r - r^f$ additional conservations for the fast layer. Therefore, each layer's trajectory is constrained to a smaller sub-space of $\mathbb{R}^N$.

Consider again the enzyme decomposition example above. The full stoichiometric matrix in (5) has rank 2. Carrying out a layered decomposition results in two stoichiometric matrices in (8) and (9), each of which has rank 1. Therefore each layer in this example is of smaller dimension than the full system. The additional

conservation relation in (8) is that $x^f + c^f = K^f$, while that in (9) is $x^s = K^s$.

3. Layered model reduction

In Section 2 we described an alternative decomposition method for singular perturbation which does not rely on transformations or partitioning the state space. Furthermore, this decomposition results in smaller-dimensional subsystems embedded in the natural coordinate system, which therefore have an intuitive interpretation in relation to the reactant concentrations being modelled. The next stage of singular perturbation, as shown in Fig. 1, is to approximate the fast dynamic subsystem (whether it is a module or layer) into a static subsystem. In this section we will investigate the impact of the layered decomposition on the results of such an approximation.

3.1. Approximation through singular perturbation

3.1.1. Fast module
We first consider again the modular decomposition, where the transformed state $z = (z_f, z_s) = Tx$ is partitioned into fast and slow modules. The dynamics of each module are

$$\epsilon \dot{z}_f = \overline{S}_f \overline{v}(z_f, z_s),$$
$$\dot{z}_s = \overline{S}_s \overline{v}(z_f, z_s),$$

where $\overline{v} = v \circ T^{-1}$. To reduce the fast dynamic module to a static module let $\epsilon \to 0$ so that

$$0 = \overline{S}_f \overline{v}(\tilde{z}_f, \tilde{z}_s),$$
$$\dot{\tilde{z}}_s = \overline{S}_s \overline{v}(\tilde{z}_f, \tilde{z}_s),$$

where $\tilde{z}_i$ denotes the approximated value of each module's state z_i. In the limit, the value of $\tilde{z}_f$ depends statically on that of $\tilde{z}_s$ through an implicit algebraic relationship. The state of the singularly-perturbed system can then be recovered from $\tilde{z} = (\tilde{z}_f, \tilde{z}_s)$. In the original coordinate system, the approximated system therefore has the state $\tilde{x} = T^{-1}\tilde{z}$.

In the case where transformation is necessary, the approximation leads to the static module $\tilde{z}^f = \phi(\tilde{z}^s)$. The dependence is in terms of the transformed variables, so we cannot interpret the interaction of modules in terms of the original biochemical species. We will now show that the layered decomposition maintains the physical bases of the subsystems.

3.1.2. Fast layer
We now consider the layered decomposition (3)–(4). As in the modular case above, this system is approximated by letting $\epsilon \to 0$ on the left-hand side of (3) to give the singularly perturbed system

$$0 = S^f v^f(\tilde{x}^f + \tilde{x}^s), \tag{12}$$

$$\dot{\tilde{x}}^s = S^s v^s(\tilde{x}^f + \tilde{x}^s), \tag{13}$$

where again the tilde denotes approximation. Taking the sum of each layer's state

$$\tilde{x} = \tilde{x}^s + \tilde{x}^f, \tag{14}$$

gives the conversion from layered dynamics to the full system.

The algebraic relationship (12) defines an approximation of the fast layer as a static system, where the fast variable is assumed to immediately reach a steady state $\tilde{x}^f$ that depends on the slow input $\tilde{x}^s$. The following proposition proves the local existence of the function $\tilde{x}^f = \phi(\tilde{x}^s)$ that arises from the implicit relationship (12).

Proposition 1. *Consider any decomposition of the fast stoichiometry $S^f = U^f C^f$, where U^f and C^f are full column and row rank respectively, and assume that (12) holds. Suppose that*

$$U^f \partial v^f(\tilde{x}^s_c + \tilde{x}^f_c) C^f$$

is invertible at a given value of $(\tilde{x}^s_c, \tilde{x}^f_c)$, where ∂v^f denotes the Jacobian of v^f. Then there exists a function ϕ defined on a neighbourhood of $\tilde{x}^s_c$ such that $\tilde{x}^f = \phi(\tilde{x}^s)$.

Proof. By (3), it follows that $x^f(t) - x^f(0) \in \mathrm{Col}(S^f)$. Since $S^f = U^f C^f$, the columns of the full column rank matrix U^f form a basis for $\mathrm{Col}(S^f)$. Thus we can write the state of the fast layer $x^f(t) = x^f(0) + U^f \xi$ for some $\xi \in \mathbb{R}^{r^f}$, where $r^f = \mathrm{rank}(S^f)$. In the limit as $\epsilon \to 0$, the approximated fast layer's state is therefore $\tilde{x}^f(t) = x^f(0) + U^f \tilde{\xi}$.

Substituting this expression for $\tilde{x}^f$ into (12) gives

$$0 = U^f C^f v^f(\tilde{x}^s + x^f(0) + U^f \tilde{\xi})$$

where we again write $S^f = U^f C^f$. Since U^f is full column rank, this relationship holds if and only if

$$0 = C^f v^f(\tilde{x}^s + x^f(0) + U^f \tilde{\xi}) =: F(\tilde{x}^s, \tilde{\xi}). \tag{15}$$

We have constructed a function $F : \mathbb{R}^N \times \mathbb{R}^{r^f} \to \mathbb{R}^{r^f}$ that defines the manifold on which $\tilde{\xi}$ takes its values. The Implicit Function Theorem (IFT) can now be applied to F. This states that if $\partial F / \partial \tilde{\xi}$ is invertible at a given point, the equation $F = 0$ results in a function $\tilde{\xi} = \psi(\tilde{x}^s)$ locally around that point.

The required derivative is calculated via the chain rule as

$$\frac{\partial F}{\partial \tilde{\xi}} = C^f \partial v^f(\tilde{x}^s + \tilde{x}^f) U^f. \tag{16}$$

By the initial assumption that this matrix is invertible at $(\tilde{x}^f_c, \tilde{x}^s_c)$, the function $\tilde{\xi}$ must therefore exist by the IFT. Hence we can write $\tilde{x}^f = x^f(0) + U^f \tilde{\xi} = x^f(0) + U^f \psi(\tilde{x}^s) = \phi(\tilde{x}^s)$. □

There are now two ways to express the layered approximated dynamics. The first is to re-write (12)–(14) by replacing the implicit relationship (12) with the explicit relationship ϕ, to give

$$\dot{\tilde{x}}^s = S^s v^s(\tilde{x}^s + \phi(\tilde{x}^s)), \tag{17}$$

$$\tilde{x} = \tilde{x}^s + \phi(\tilde{x}^s). \tag{18}$$

This requires the construction of the function ϕ arising from (12), which may be difficult to calculate. An alternative expression of the approximate dynamics arises from combining (17)–(18) by applying the chain rule to $\dot{\tilde{x}}$. The approximate dynamics are

$$\dot{\tilde{x}} = \left(I + \frac{\partial \phi}{\partial \tilde{x}^s}\right) S^s v^s(\tilde{x}), \tag{19}$$

where we have substituted $\dot{\tilde{x}}^s$ with (17). The ODE (19) can be calculated without knowledge of ϕ using the following corollary to the IFT.

Lemma 2. *Assume again the conditions of Proposition 1. The dependence of the fast variable on the slow variable $\tilde{x}^f = \phi(\tilde{x}^s)$ given implicitly by (12) is such that*

$$\frac{\partial \tilde{x}^f}{\partial \tilde{x}^s} = -U^f [C^f \partial v^f(\tilde{x}) U^f]^{-1} C^f \partial v^f(\tilde{x}) =: M(\tilde{x}), \tag{20}$$

where ∂v^f denotes the Jacobian of v^f.

Proof. Recall the construction of F in (15), and that (12) is equivalent to $F = 0$. A standard corollary to the IFT applied in the proof of Proposition 1 is that the derivative of ψ is found in terms

of the partial derivatives of F as

$$\frac{\partial \psi}{\partial \tilde{x}^s} = -\left[\frac{\partial F}{\partial \tilde{\xi}}\right]^{-1}\frac{\partial F}{\partial \tilde{x}^s}.$$

The invertible derivative $\partial F / \partial \tilde{\xi}$ is (16), while a simple differentiation of (15) gives

$$\frac{\partial F}{\partial \tilde{x}^s} = C^f \partial v^f(\tilde{x}^s + \tilde{x}^f).$$

By writing $\tilde{x}^f = x^f(0) + U^f \tilde{\Psi}(\tilde{x}^s)$ we find

$$\frac{\partial \tilde{x}^f}{\partial \tilde{x}^s} = -U^f \left[\frac{\partial F}{\partial \tilde{\xi}}\right]^{-1}\frac{\partial F}{\partial \tilde{x}^s},$$

which matches the expression in (20). □

The preceding lemma is now used to derive the following result:

Proposition 2. *Given a partition of the flux vector v in a biochemical reaction network* (1) *into a sum of fast and slow vector fields as in* (2), *the singularly perturbed dynamics on the slow timescale are*

$$\dot{\tilde{x}} = (I + M(\tilde{x}))S^s v^s(\tilde{x}) \tag{21}$$

with perturbation matrix M given by (20).

Proof. This result follows directly from (19) and (20). □

We can thus derive two expressions for the singularly perturbed system: the layered system (17)–(18), and the combined system (21). Note that, because $M(\tilde{x})$ depends on the state, it is not immediately clear that (21) forms a reduction of (1), although the system has been approximated by removing the stiffness. In Gerdtzen et al. (2004) the authors use a result similar to (21) to reduce the model, by deriving a modular decomposition of a transformed state space.

The fact that singular perturbation of a layered decomposition is a model reduction technique is made clear through the layered expression (17)–(18) which was used to derive (21). The state $\tilde{x}^s$ in (17) is an embedding into $\mathbb{R}^N$ of an r^s-dimensional system, where $r^s = \text{rank}(S^s) < \text{rank}(S)$, and can therefore be reduced to a system of r^s equations using the arguments of Section 2.4. However, in order to use (17)–(18) we need to calculate ϕ.

We will now demonstrate the use of layered decomposition for singular perturbation on two biological examples.

3.2. Enzyme kinetics example revisited

Recall the two decompositions of the enzyme kinetics system described in Section 2.3. First consider the layered decomposition (8)–(9). For singular perturbation we let $\epsilon \to 0$ in (8). The resulting expression for (12)–(13) is

$$0 = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} k_1 \tilde{x}(K - \tilde{c}) \\ k_{-1}\tilde{c} \end{bmatrix}, \tag{22}$$

$$\begin{bmatrix} \dot{\tilde{x}}^s \\ \dot{\tilde{c}}^s \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}k_2 \tilde{c},$$

where $\tilde{c} = \tilde{c}^f + \tilde{c}^s$ and $\tilde{x} = \tilde{x}^f + \tilde{x}^s$.

Section 3.1 gives two possible representations of these dynamics: (17)–(18), or (21). By Proposition 2 the singularly perturbed system (21) is determined by the calculation of M in (20) as

$$M(\tilde{x}, \tilde{c}) = \begin{bmatrix} -\gamma & 1 - \gamma \\ \gamma & -(1-\gamma) \end{bmatrix},$$

$$\gamma = \frac{k_1(K - \tilde{c})}{k_1(K + \tilde{x} - \tilde{c}) + k_{-1}}.$$

Substituting M into (21) results in the ODE system

$$\begin{bmatrix} \dot{\tilde{x}} \\ \dot{\tilde{c}} \end{bmatrix} = \begin{bmatrix} 1 - \gamma & 1 - \gamma \\ \gamma & \gamma \end{bmatrix}\begin{bmatrix} 0 \\ -1 \end{bmatrix}k_2 \tilde{c},$$

which is written in terms of the approximate values of each concentration $\tilde{x}$ and $\tilde{c}$. The 2×2 matrix $(I + M)$ multiplying the isolated slow dynamics represents the state-dependent effect of the fast layer on the slow layer in the singularly-perturbed system.

Writing these dynamics out in full as

$$\begin{bmatrix} \dot{\tilde{x}} \\ \dot{\tilde{c}} \end{bmatrix} = \frac{-1}{k_1(K + \tilde{x} - \tilde{c}) + k_{-1}}\begin{bmatrix} k_1 \tilde{x} + k_{-1} \\ k_1(K - \tilde{c}) \end{bmatrix}k_2 \tilde{c},$$

it is not immediately clear that the approximation arising from (21) gives a model reduction, although we can see that the original system stiffness has indeed been approximated. As discussed in Section 3.1, the layered expression (17)–(18) can give a better practical model reduction scheme. However, it requires us to find $[\tilde{x}^f, \tilde{c}^f]^T = \phi([\tilde{x}^s, \tilde{c}^s]^T)$ from the algebraic relationship (22).

We can show by Proposition 1 that the function ϕ exists and by Lemma 2 that its Jacobian is M as given above. After calculating $\phi = (\tilde{x}^f, \tilde{c}^f)^T$, (which arises from the solution to a quadratic equation and is omitted for brevity) the approximated slow dynamics (17)–(18) are

$$\begin{bmatrix} \dot{\tilde{x}}^s \\ \dot{\tilde{c}}^s \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}k_2(\tilde{c}^s + \tilde{c}^f(\tilde{x}^s, \tilde{c}^s)), \quad \begin{bmatrix} \tilde{x} \\ \tilde{c} \end{bmatrix} = \begin{bmatrix} \tilde{x}^s \\ \tilde{c}^s \end{bmatrix} + \phi\left(\begin{bmatrix} \tilde{x}^s \\ \tilde{c}^s \end{bmatrix}\right).$$

This system is clearly one-dimensional, since $\tilde{x}^s$ is constant. This illustrates how the layered representation (17)–(18) ensures that the dynamics (21) imply a reduced model. The simulated trajectories of $\tilde{c}^i$ in each approximated layer can be seen in Fig. 5. The fast variable $\tilde{c}^f$ is now a static function of the slow variables (x^s, c^s), resulting in a trajectory which is clearly close to the exact trajectory shown in Fig. 4, except along the transient time period $t \ll 1$.

The interpretation of the trajectories in Fig. 4 still applies to those in Fig. 5, but now the contribution from the fast layer is approximated to be static rather than dynamic. This means that the contribution arising from the equilibration of the fast reversible reaction $X + E \rightleftharpoons C$ is assumed to occur instantaneously. Thus the approximation introduced by singular perturbation now has a clear physical interpretation when considered in terms of layered decomposition.
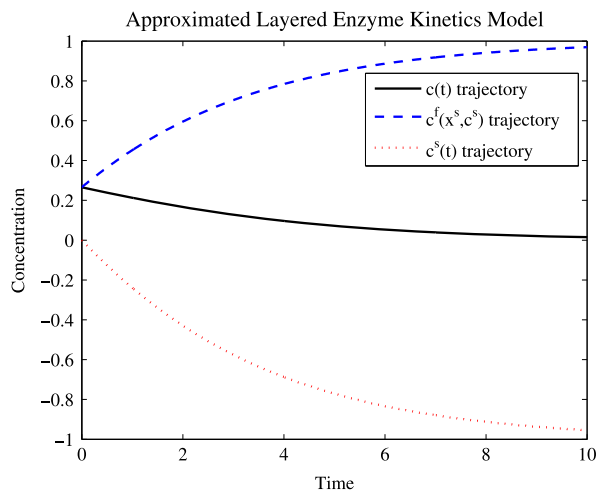
Fig. 5. These trajectories are the approximated equivalents of those in Fig. 4. The broken curves correspond to the layered states $\tilde{c}^s(t)$ and $\tilde{c}^f(t) = \tilde{c}^f([\tilde{x}^s, \tilde{c}^s])$, whose sum recovers $\tilde{c}(t)$ plotted along the black curve. Note that the fast transient has been approximated to be an instantaneous contribution from the fast layer.

3.3. The Wnt pathway

As a demonstration of layered singular perturbation on a substantial biochemical example, we will consider the reduction of the model of the Wnt pathway described in Krüger and Heinrich (2004). This model consists of 15 reactants concentrations x_i and 25 reactions $v_j(x)$. The authors make three distinctions of timescale by partitioning the flux vector v into fast, medium, and slow timescales. We will approximate the system on the medium timescale. Parameterising these scales by $\epsilon \ll 1$, the dynamics can be written in the form of (2) as

$$\dot{x} = \frac{1}{\epsilon} S^f v^f(x) + S^m v^m(x) + \epsilon S^s v^s(x)$$

for the fast, medium and slow stoichiometries and flux vectors given in Fig. 6. Each reaction parameter k_i is assumed to be on the same scale; any scale separation only arises from ϵ.

Applying a slight modification of the layered decomposition in Section 2 to this system results in the definition of three interconnected layers,

$$\epsilon \dot{x}^f = S^f v^f(x^f + x^m + x^s),\tag{23}$$

$$\dot{x}^m = S^m v^m(x^f + x^m + x^s),\tag{24}$$

$$\dot{x}^s = \epsilon S^s v^s(x^f + x^m + x^s),\tag{25}$$

where we choose to put the initial conditions into the medium timescale so that $x^m(0) = x(0)$ and $x^f(0) = x^s(0) = 0$. Note that each subsystem is embedded in the full state space $\mathbb{R}^{15}$, but the dimension of each layer is given by the rank of its stoichiometry.

Consider the modular decomposition of this model in Krüger and Heinrich (2004), in which the authors create three new medium timescale variables P_i as linear combinations, or "pools", of the reactants that take part both in reactions of more than one timescale; for example $P_3 = X_8 + X_{11} + X_{14} + X_{15}$. The variables resulting from modular decomposition are therefore not all expressed in terms of the original species of interest. We will demonstrate in this example that, in the case where reaction speeds separate in scale, layering is a natural decomposition method for singular perturbation.

We now approximate the fast and slow timescales, and consider the medium timescale only. Taking the limit as $\epsilon \to 0$ in (25), the slow layer's dynamics are approximated by $\dot{\tilde{x}}^s = 0$. The zero initial condition implies $\tilde{x}^s \equiv 0$, and there is no contribution to the medium timescale dynamics from the slow layer.

The more interesting consequence of singular perturbation comes from the approximation of the fast layer into a static subsystem. Letting $\epsilon \to 0$ in (23), the approximated state $\tilde{x}^f$ of the fast layer satisfies the algebraic relationship $S^f v^f(\tilde{x}^m + \tilde{x}^f) = 0$. As described in
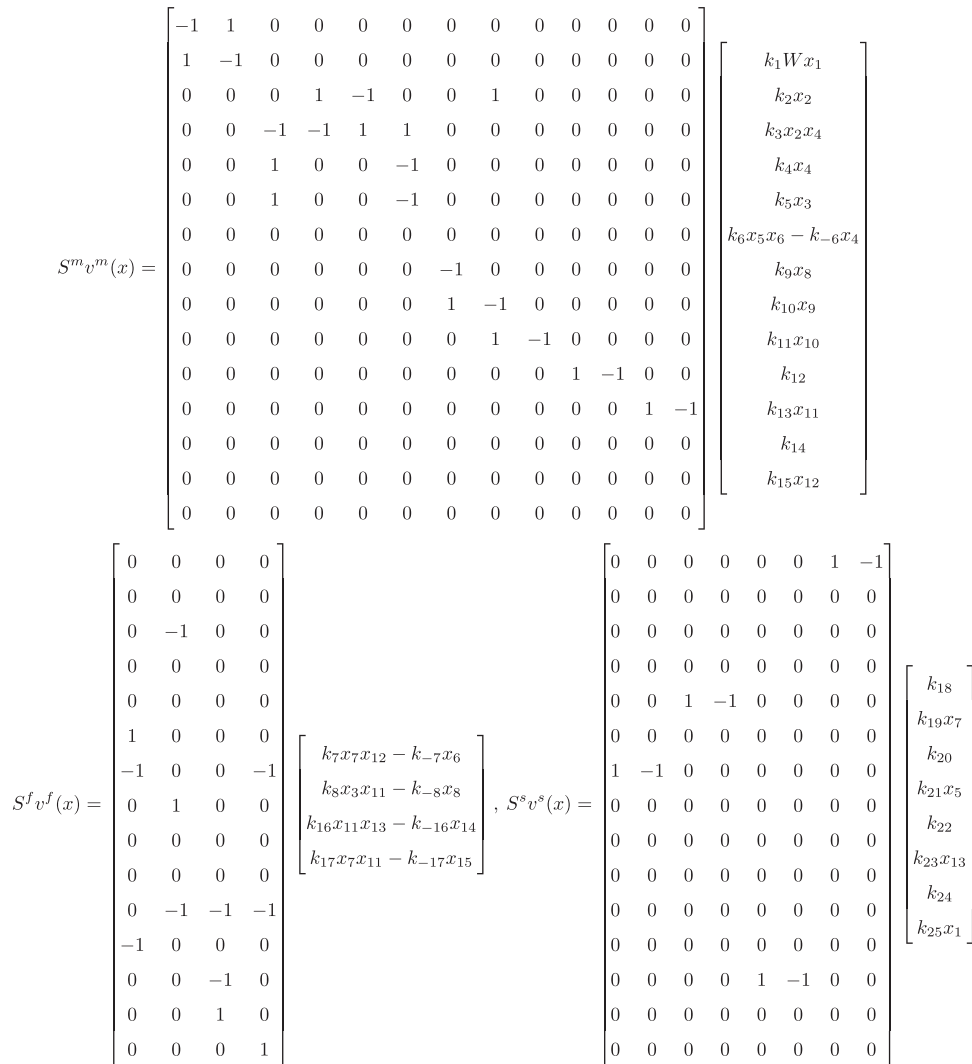

Fig. 6. The stoichiometry and flux vectors at the fast, medium and slow timescales in the model of the Wnt pathway in Krüger and Heinrich (2004).

Section 3.1, this algebraic equation implicitly defines the static system through the existence of a function ϕ such that $\tilde{x}^f = \phi(\tilde{x}^m)$.

We can show the existence of ϕ through Proposition 1. Letting $C^f = I$ and $U^f = S^f$ gives a decomposition of $S^f = U^f C^f$ that satisfies the required full row and column rank conditions respectively. The matrix $\partial v^f S^f$ is non-singular for all non-negative concentrations, and thus the condition for the existence of ϕ in Proposition 1 is satisfied. In order to construct this function recall that in the proof of Proposition 1 we show the existence of ψ such that $\phi = U^f \psi$. Here $U^f = S^f$ has rank 4, so the resulting ψ takes values in $\mathbb{R}^4$. To find ψ, and hence ϕ, the manifold given by (15) implies that $\tilde{x}^f = S^f \psi(\tilde{x}^m)$ must satisfy

$$v^f(\tilde{x}^m + S^f \psi(\tilde{x}^m)) = 0. \tag{26}$$

This condition can be used to either derive an analytical expression for ψ, or calculate ψ numerically.

Substituting $\tilde{x}^f = S^f \psi(\tilde{x}^m)$ and $\tilde{x}^s = 0$ into (24) gives the approximated reduced order system

$$\dot{\tilde{x}}^m = S^m v^m(\tilde{x}^m + S^f \psi(\tilde{x}^m)) \tag{27}$$

with initial conditions $\tilde{x}^m(0) = x(0)$. As described in Section 2.4, this approximate model is of dimension $\text{rank}(S^m) = 8$. This means it could be reduced to a system of 8 ODEs without incurring any error. Finally, the approximated system state is

$$\tilde{x} = \tilde{x}^m + S^f \psi(\tilde{x}^m). \tag{28}$$

As in the previous example, this can be interpreted as correcting the contribution to the state from the medium timescale $\tilde{x}^m$ by adding on $S^f \psi(\tilde{x}^m)$. This additional contribution arises from the fast layer instantly equilibrating the concentrations supplied by $\tilde{x}^m$ across the four fast reversible reactions.

For example, suppose we impose the initial conditions $x_i(0) = 1$ for all $i = 1, \ldots, 15$ with parameters $k_j = 1$ and $W = k_{-j} = 2$. Having let $\epsilon \to 0$, the initial value of the approximated fast layer's state $\tilde{x}^f(0) = \phi(\tilde{x}^m(0))$ is

$$\tilde{x}^f(0) = [0, 0, 0.17, 0, 0, -0.214, 0.295, -0.17,$$
$$0, 0, 0.42, 0.214, 0.17, -0.17, -0.08]^T,$$

which is the contribution of the static (i.e. approximated fast) layer at time $t = 0$. This contribution overlays onto $\tilde{x}^m(0)$ to give the approximate initial condition $\tilde{x}(0) = \tilde{x}^m(0) + \phi(\tilde{x}^m(0))$ for the full singularly perturbed system. The fact that both the fast and medium layers' states are embedded in the untransformed state space $\mathbb{R}^{15}$ means that we can immediately see the contribution of the fast dynamics to each variable in the approximated system: $\tilde{x}_{11}$ in particular receives a large correction of $+42\%$ as it "instantly" equilibrates across the four reversible reactions that comprise the fast layer.

After this initial correction, the medium timescale variable $\tilde{x}^m$ evolves according to (27). The value of $\tilde{x}^f(t) = \phi(\tilde{x}^m(t))$ at each time point provides the additive contribution of the fast layer to give the approximate state $\tilde{x}(t)$. For example, Fig. 7 plots the contributions to $\tilde{x}_{12}$ from each layer. The fast layer provides an instant correction of $\tilde{x}_{12}^f(0) = 0.214$ as the initial conditions instantly equilibrate across the fast reactions. The medium layer contributes decreasing quantities of $\tilde{x}_{12}^m$, since the net result of the reactions in this layer is to consume X_{12}. This is offset by the increase to $\tilde{x}_{12}$ contributed by the fast reactions, which equilibrate the changing concentrations supplied by the reactions in the medium layer.

This analysis illustrates that we can directly observe the effect of the approximated layer in terms of contributions to each reactants' concentration. In Krüger and Heinrich (2004) the decomposition is modular and the required transformations mean that fast variables are expressed in terms of pooled variables. Thus the interpretation of singular perturbation in terms of the original
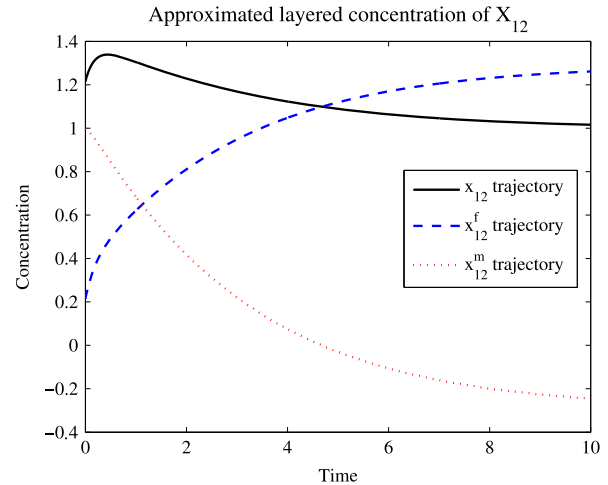

Fig. 7. These curves plot the approximate concentration of X_{12} in the singularly perturbed system (27)–(28) The broken curves correspond to the layered states $\tilde{x}_{12}^f = \tilde{x}_{12}^f(\tilde{x}^m)$ and $\tilde{x}_{12}^m$, whose sum recovers $\tilde{x}_{12}$ plotted along the black curve. Note that a fast transient is approximated to be an instantaneous contribution from the fast layer.

species is lost by a decomposition into modules. This example illustrates that this decomposition gives a direct interpretation of the fast subsystem as the contribution of the fast biochemical processes to the evolution of each species.

4. Discussion

We first discuss what the layered model reduction in Section 3.1 does *not* do. Consider again the expanded dynamics given by (2), and the approximation to the fast layer given by (12). A naïve approach to the timescale separation argument is to simply substitute (12) into (2), giving approximate dynamics equal to the dynamics of the slow layer running in isolation: $\dot{\tilde{x}} = S^s v^s(\tilde{x})$. This approach is clearly incorrect when considering the slow layer (9) of the enzyme kinetics example in isolation; this approach would imply that $\tilde{x}$ was constant. This paper has instead approximated the system as the slow layer running with an input supplied by the action of the fast layer on the slow layer. This input can be written in two ways: as a corrective term to the slow state as shown in (17)–(18); or as a state-dependent perturbation to the slow dynamics (21). As shown in Proposition 2, the second of these representations is a consequence of the first.

This work is closely related to previous work (Kumar et al., 1998; Vora and Daoutidis, 2001) which deals with stiff systems in non-standard form. In Kumar et al. (1998) the authors characterise the conditions under which transformations exist which determine fast and slow variables, while Vora and Daoutidis (2001) show that the stiffness can be directly approximated without transformation. Indeed, the derivation of (21) recovers results from Gerdtzen et al. (2004) and Vora and Daoutidis (2001). In the previous work, that result is then used to derive a model reduction by a modular timescale separation, where the "actual slow variables" are combinations of the original variables. By deriving (21) (from 17–18), this work ensures that the fast and slow variables are embedded in the original coordinates. The rank arguments of Section 2.4 are sufficient to ensure that we have a model reduction that allows the direct identification of the fast and slow contributions to each species.

The effect encoded in M as found in (20) can be interpreted as quantifying the retroactivity between two modules (Gyorgy and Del Vecchio, 2014; Del Vecchio et al., 2008). Formulating subsystems as layers, rather than modules, directly models retroactivity by summing the contributions of species taking part in

multiple layers. Since the retroactivity has been accounted for as the input of each layer, the input–output behaviour of a layer remains the same on interconnection. The timescale separation discussed in Jayanthi and Del Vecchio (2011) also recovers a version of the results in (21). By using the column and row space bases U^f and C^f of the fast layer's stoichiometry in (20), this result can be interpreted in terms of the stoichiometric structure of each subsystem's reaction network.

The result in Jayanthi and Del Vecchio (2011) also dealt with modules on more than two timescales. We can adapt the layering method to consider the approximation of the slow dynamics in the example in Section 3.3, where we originally focussed on the medium timescale. We could focus on the slow timescale by re-scaling t in (23)–(25) to give

$$\epsilon^2 \dot{x}^f = S^f v^f(x^f + x^m + x^s), \tag{29}$$

$$\epsilon \dot{x}^m = S^m v^m(x^f + x^m + x^s), \tag{30}$$

$$\dot{x}^s = S^s v^s(x^f + x^m + x^s), \tag{31}$$

where we will now put the initial conditions $x^s = x(0)$, with $x^m = x^f = 0$. Here, the timescale separation will need to be carried out in two steps, as described in Vora et al. (2006), by considering the layers in sequence from fast to slow. The timescale hierarchy (Jamshidi and Palsson, 2008) is reflected by first constructing $x^f = \phi_1(x^m + x^s)$ from (29), and then re-writing (30) as

$$0 = S^m v^m(\phi_1(x^m + x^s) + x^m + x^s)$$

to find the next implicit function $x^m = \phi_2(x^s)$. A similar method to that used to derive (21) would allow the approximated dynamics to be expressed in terms of a perturbation (depending on the column and row spaces of S^f and S^m) to the isolated slow dynamics $S^s v(x)$.

Using the M representation (21) for model reduction is difficult because, as shown in Section 3.2, the conservation relations in the resulting differential equations are obscured by the nonlinearity of M. The stiffness has been removed, which does simplify simulation, but without going through some potentially intractable further analysis there would still be as many differential equations to solve as there were before. Another benefit of the layered representation of the approximate dynamics given by (17)–(18) is therefore that we have a non-stiff approximation of the system that can be further reduced without incurring any additional error, by using the dimensionality arguments of Section 2.4 on the slow layer's dynamics (17).

Our decision to term these subsystems "layers" is influenced in part by work related to Bruggeman et al. (2008) and Hofmeyr and Westerhoff (2001). There, the authors consider a particular decomposition, which they call layering, where the reactants are partitioned into subsystems so that the reactions are also partitioned. This is a special case, only possible for block diagonal S. In the terminology of this paper, that decomposition is both layered *and* modular, since both reactions and reactants are partitioned. A layered approach is also conceptually similar to work done in static networks on multislice systems (Mucha et al., 2010). In multislice networks, nodes (i.e. reactants) are replicated across all slices (i.e. layers), and edges are either intraslice (i.e. reactions within a layer) or between the replicated nodes in different slices (i.e. the summation across all layers of each reactant's contributions).

The arguments in this paper have been in the context of biochemical networks. However, as with modular decompositions, layered decompositions for timescale separation are in theory applicable to any system with stiff dynamics. The dynamics in (2)

are a special case of the more general dynamics

$$\dot{x} = F(x) = F^s(x) + \frac{1}{\epsilon} F^f(x)$$

where F^s and F^f take values on the same scale. Defining the layered dynamics as

$$\dot{x}^s = F^s(x^s + x^f),$$
$$\epsilon \dot{x}^f = F^f(x^s + x^f),$$

with initial conditions such that $x^s(0) + x^f(0) = x(0)$, we can perform a similar approximation by letting $\epsilon \to 0$ and applying the Implicit Function Theorem to deduce the resulting approximate dependence of $\tilde{x}^f$ on $\tilde{x}^s$. In the context of biochemical systems, the specific form of F^f and F^s allows us to calculate this dependence directly, and ensure that the approximated model is of a lower dimension than the original. In full generality, this is not necessarily possible. One potential direction for future work would be to explore how to layer timescale-separated systems with different types of dynamics. Alternatively, we may be able to apply layered decomposition to other, more general, systems without necessarily partitioning the vector field based on timescale separation; for example, layered networks of multi-agent systems as in Tsubakino and Hara (2012) and Williams et al. (2004).

5. Conclusion

We have presented the dimensionality reduction of biochemical network models through singular perturbation as a process of model decomposition followed by the approximation of a fast dynamic subsystem by a static subsystem. Our contribution has been to use a new method of decomposition, introduced in Prescott and Papachristodoulou (2013), called layering.

The new decomposition technique works from a principle of partitioning reactions into subsystems, rather than reactants. The resulting layers are lower-dimensional subsystems which are embedded in the original state space, whose states sum to give the original system. When scale separation occurs in reaction speeds, layers form a natural technique for finding fast and slow variables without requiring any transformations of the state space. As no transformations have been employed, the state of each layer is interpreted as the contribution to the observed state arising from its defining subset of reactions.

We then proceed with singular perturbation by approximating the fast layer as a static subsystem. The contribution $\tilde{x}^f$ of the approximated fast layer provides an estimate of how the instant equilibrium of the fast reactions correct the contribution $\tilde{x}^s$ of the slow dynamics, resulting in an estimate $\tilde{x}$ for the observed dynamics of the singularly perturbed system.

We have demonstrated that layering is a natural system decomposition in the context of biochemical reaction networks. We expect that, beyond the question of model reduction, this framework will be able to provide additional insights to the analysis and understanding of large, complex biochemical networks.

Acknowledgements

References

Alexander, R.P., Kim, P.M., Emonet, T., Gerstein, M.B., 2009. Understanding modularity in molecular networks requires dynamics. Sci. Signal. 2 (81), pe44.

Bennett, M.R., Volfson, D., Tsimring, L., Hasty, J., 2007. Transient dynamics of genetic regulatory networks. Biophys. J. 92, 3501–3512.

Bruggeman, F.J., Snoep, J.L., Westerhoff, H.V., 2008. Control, responses and modularity of cellular regulatory networks: a control analysis perspective. IET Syst. Biol. 2 (6), 397–410.

Ciliberto, A., Capuani, F., Tyson, J.J., 2007. Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation. PLOS Comput. Biol. 3 (3), e45.

Del Vecchio, D., Ninfa, A.J., Sontag, E.D., 2008. Modular cell biology: retroactivity and insulation. Mol. Syst. Biol. 4, 161.

Gerdtzen, Z.P., Daoutidis, P., Hu, W.-S., 2004. Non-linear reduction for kinetic models of metabolic reaction networks. Metab. Eng. 6, 140–154.

Gyorgy, A., Del Vecchio, D., 2014. Modular composition of gene transcription networks. PLOS Comput. Biol 10 (3), e1003486.

Hinch, E.J., 1991. Perturbation Methods. Cambridge University Press, Cambridge, UK.

Hofmeyr, J.-H., Westerhoff, H.V., 2001. Building the cellular puzzle: control in multi-level reaction networks. J. Theor. Biol. 208, 261–285.

Jamshidi, N., Palsson, B.O., 2008. Top-down analysis of temporal hierarchy in biochemical reaction networks. PLOS Comput. Biol. 4 (9), e1000177.

Jayanthi, S., Del Vecchio, D., 2011. Retroactivity attentuation in bio-molecular systems based on timescale separation. IEEE Trans. Autom. Control 56 (4), 748–761.

Krüger, R., Heinrich, R., 2004. Model reduction and analysis of robustness for the Wnt/β-Catenin signal transduction pathway. Genome Inf. 15 (1), 138–148.

Kumar, A., Christofides, P.D., Daoutidis, P., 1998. Singular perturbation modeling of nonlinear processes with nonexplicit time-scale multiplicity. Chem. Eng. Sci. 53 (8), 1491–1504.

Lee, C.H., Othmer, H.G., 2010. A multi-time-scale analysis of chemical reaction networks: deterministic systems. J. Math. Biol. 60 (3), 387–450.

Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., Onnela, J.-P., 2010. Community structure in time-dependent, multiscale, and multiplex networks. Science 328, 876–878.

Murray, J.D., 2002. Mathematical Biology: An Introduction. Springer, New York, NY, USA.

Palsson, B.O., 2006. Systems Biology: Properties of Reconstructed Networks. Cambridge University Press, Cambridge, UK.

Prescott, T.P., Papachristodoulou, A., 2012. Guaranteed error bounds for structured complexity reduction of biochemical networks. J. Theor. Biol. 302, 172–182.

Prescott, T.P., Papachristodoulou, A., 2013. Layering in networks: the case of biochemical systems. In: Proceedings of the American Control Conference (ACC), pp. 4544–4549.

Southern, J., Pitt-Francis, J., Whiteley, J., Stokeley, D., Kobashi, H., Nobes, R., Kadooka, Y., Gavaghan, D., 2008. Multi-scale computational modelling in biology and physiology. Prog. Biophys. Mol. Biol. 96, 60–89.

Tsubakino, D., Hara, S., 2012. Eigenvector-based intergroup connection of low rank for hierarchical multi-agent dynamical systems. Syst. Control Lett. 61 (2), 354–361.

Tzafriri, A.R., 2003. Michaelis–Menten kinetics at high enzyme concentrations. Bull. Math. Biol. 65, 1111–1129.

Vora, N., Daoutidis, P., 2001. Nonlinear model reduction of chemical reaction systems. AIChE J. 47 (10), 2320–2332.

Vora, N.P., Contou-Carrere, M.-N., Daoutidis, P., 2006. Model reduction of multiple time scale processes in non-standard singularly perturbed form. In: Gorban, A.N., Kazantzis, N.K., Kevrekidis, I.G., Öttinger, H.C., Theodoropoulos, C. (Eds.), Model Reduction and Coarse-Graining Approaches for Multiscale Phenomena. Springer, Berlin Heidelberg.

Williams, A., Glavaški, S., Samad, T., 2004. Formations of formations: hierarchy and stability. in: Proceedings of the American Control Conference.