

COMPLEXITY REDUCTION OF BIOCHEMICAL NETWORKS

Ravishankar R. Vallabhajosyula
Herbert M. Sauro

Keck Graduate Institute
535 Watson Drive,
Claremont, CA 91711, U.S.A.

ABSTRACT

This paper discusses two broad approaches for reducing the complexity of large cellular network models. The first approach involves exploiting conservation and time-scale separation and allows the dimension of the model to be significantly reduced. The second approach involves identifying subnetworks that carry out well defined functions and replacing these with simpler representations. Examples include identification of functional subnetworks such as oscillators or bistable switches and replacing these with a simplified mathematical construct. This enables complex networks to be rationalized as a series of hierarchical modules and greatly simplifies our ability to understand the dynamics of complex networks.

1 INTRODUCTION

Biological networks, even at the sub-cellular level, can involve interactions between hundreds or thousands of different molecules. Such systems are complex, both in organization as well as functional properties of their subsystems. The level of organization in such systems can span spatial scales, and their dynamics invariably involves multiple time-scales. This can lead to models that are often difficult to simulate because this introduces stiffness into the model, requiring sophisticated algorithms to maintain numerical stability.

The issue of time-scale hierarchy in biological systems has been a subject of interest for a number of years, see Heinrich et al. (1977) for a review. David Park (1974) was probably one of the first to give a systematic analysis on model reduction in biological networks. He showed how species in the model could be aggregated into pools comprising of fast and slow components, thereby yielding a reduced order network which is easier to simulate. It has been argued that such a time-scale hierarchy bestows robustness to biological systems with respect to microscopic parameters, see Rojdestvenski et al. (1999). This argument

was indeed shown to have some validity in the case of the carbohydrate uptake and metabolism of *Escherichia coli* by Kremling et al. (2004) who found that the slow subsystem of the model describing the metabolic pathway was robust to some parameter changes. This issue has been addressed previously by Gardner and Ashby (1970) where it has been shown using computer studies that a system with more variables is more likely to be unstable. It is therefore inferred that time hierarchies help in stabilizing such systems by reducing it to include only those variables that are essential to the dynamics. It is of interest to note that such phenomena involving multiple time-scales are also encountered in engineering, physics, economics and many other areas.

The problem of model reduction has been a topic of interest to a number of other fields, for example chemical kinetics and combustion modeling. The results reported by investigators in these fields also apply to dynamical systems of interest to the biological community. One such review of interest is that by Okino and Mavrovouniotis (1998), who discuss various complexity reduction methods such as Lumping, which was adopted as early as 1953, Sensitivity Analysis based model reduction, both in chemical models by Rabitz et al. (1983) and Turányi (1990) and in biological models by Liao et al. (1988, 1993). In this approach the list of species is replaced by a reduced system containing only the important species. Species that have significant effect on the important species are termed necessary species while species and parameters that have only a weak influence eliminated, leading to a reduced system. However the major drawback of this method lies in the fact that, in order to obtain the sensitivity coefficients, it is necessary to evaluate the original system with the full species, reaction and parameter lists. Further, there can be numerous parameters even in a small model, and it is often a difficult task to extend this to larger models where the number of parameters can be much higher. An alternative approach to model reduction that has been used in combustion modeling makes use of an optimization method to eliminate some reactions as

described by Bhattacharjee et al. (2003). A recent study by Maurya et al. (2005) on the generation of a Reduced Order Model (ROM) for a biochemical network involved the use of a Multidimensional Sensitivity Analysis to eliminate 31 out of 48 reactions. Another reported optimization method involves the use of genetic algorithms for model reduction (Edwards et al. 1998).

2 COMPUTATIONAL COMPLEXITY

Complexity in biochemical systems is the result of a number of factors including multiple time scales, nonlinearities and often elaborate connectivity. Complexity can therefore arise even in small network models of only a few variables. Reduction of computational complexity in such models involves the application of conservation analysis along with an appropriate time scale separation method, in order to obtain a reduced model of smaller dimension.

Conservation analysis of biochemical systems involves analyzing the stoichiometry matrix for presence of dependencies. These dependencies arise due to the rank-deficiency of the stoichiometry matrix. Well known methods that have been used for this step include Gaussian Elimination, computing the null space of the stoichiometry matrix, reduction to Row Echelon form as well as the Singular Value Decomposition. A detailed review of these methods for conservation analysis has been presented by Sauro and Ingalls (2004). However, the main limitation of these methods is their lack of stability when dealing with large systems where the stoichiometry matrix is of a very high dimension. The numerical procedures implemented by these methods result in errors, leading to incorrect computation of the number of conservation relations. This adversely affects the partitioning of the original list of species into independent and dependent parts. A new method to overcome this problem was recently proposed by Vallabhajosyula et al. (2006) where a Householder-QR based algorithm making use of similarity transformations maintains high numerical accuracy for computation of the correct number of conservation laws. Validation tests on the computed conservation laws subject them to further checks, ensuring the accuracy and reliability of the final results.

It has been shown that a reduction of around 10% to 15% of the number of differential equations can be expected in the case of very large networks where there are around one thousand species and reactions. This is a significant reduction of the model size, and should be incorporated as the first step before applying other reduction methods which will be discussed later in the paper. Further, the application of conservation analysis is also essential to compute the reduced Jacobian, which must be of full rank and may be required by other analysis tools. Accurate computation of the conservation matrices is therefore important to maintain the full rank of the Jacobian.

After conservation analysis has been applied to the model, the system can be adequately described using only the independent species. However, the presence of widely separated time scales in the system brings to fore the issue of stiffness while attempting its simulation and is one of the major impediments to simulating large models. A number of methods are known that assist in the reduction of complexity in such models using time scale and other approaches. One possible approach involves the replacement of the entire model with a reduced model by eliminating some species or reactions or parameters. However, this is guided by experience, and often based on knowledge of the system being studied, and does not lend itself to automation.

Often, the presence of sufficiently varied time scales in the system results in the dynamics of the model to have contributions from two subsystems. These are the fast and slow subsystems, each of which involves reactions with fast rates and those with slow rates, along with the associated species involved in those reactions. Various methods have been developed to handle such systems by averaging the fast subsystem and obtaining algebraic relations between the slow and fast species. These methods are discussed in detail in Section 2.2.

2.1 Model Reduction by Conservation Analysis

Biochemical systems often involve conserved moieties which result from dependencies in the stoichiometry matrix. Using these conserved moieties to eliminate the dependent species can be thought of as a first order reduction and is essential for the computation of a non-singular Jacobian. In the following, we adopt the nomenclature established by Reder (1988). A biochemical network can be described by

$$\frac{d\mathbf{S}(t)}{dt} = \mathbf{N}\mathbf{v}(t) = \mathbf{f}(\mathbf{S}(t), \mathbf{p}) \quad (1)$$

where $\mathbf{S}(t) = [S_1(t)S_2(t) \dots S_m(t)]^T$ is a vector of time-dependent species concentrations, \mathbf{N} is the stoichiometry matrix and $\mathbf{v}(t) = [v_1(t)v_2(t) \dots v_n(t)]^T$ is a vector of rates of reactions. Consequently \mathbf{N} is a matrix with m rows and n columns. The right hand side in Equation (1) can be written as a function \mathbf{f} of the species vector \mathbf{S} and a vector of system parameters \mathbf{p} .

The stoichiometry matrix of the biochemical network described by Equation (1) can often be less than full rank. This is due to linear dependencies among the rows of the stoichiometry matrix, which can be interpreted as conservation relations involving the species in the network. The goal of any algorithm that computes these conservation relations should be to identify and separate m_0 independent and $m - m_0$ dependent species, where m_0 is the rank of the stoichiometry matrix. Let these be denoted as $\mathbf{S}_i(t)$ and $\mathbf{S}_d(t)$ respectively. The time evolution of the biochemical

network described in Equation (1) can then be expressed in terms of independent and dependent species as

$$\frac{d\mathbf{S}(t)}{dt} = \begin{bmatrix} d\mathbf{S}_i(t)/dt \\ d\mathbf{S}_d(t)/dt \end{bmatrix} = \begin{bmatrix} \mathbf{N}_R \\ \mathbf{N}_0 \end{bmatrix} \mathbf{v}(t). \quad (2)$$

The rows of matrix \mathbf{N} have been rearranged in Equation (2) such that the independent rows form a $m_0 \times n$ matrix \mathbf{N}_R that is of full rank, and a $(m - m_0) \times n$ matrix \mathbf{N}_0 that comprises the dependent rows of the matrix. The dependent rows can be constructed as a linear combination of the independent rows, and this can be expressed by the relation $\mathbf{N}_0 = \mathbf{L}_0\mathbf{N}_R$, where \mathbf{L}_0 is a $(m - m_0) \times m_0$ matrix. The system described by Equation (2) can then be rewritten as

$$\frac{d\mathbf{S}(t)}{dt} = \begin{bmatrix} \mathbf{N}_R \\ \mathbf{N}_0 \end{bmatrix} \mathbf{v}(t) = \mathbf{L} \mathbf{N}_R \mathbf{v}(t) \quad (3)$$

where the $m \times m_0$ matrix \mathbf{L} in Equation (3) is called the Link matrix. The dynamics of the full system described by Equation (1) can therefore be partitioned into independent and dependent parts, with the latter deriving its dynamics from the prior. These can be written as

$$\frac{d\mathbf{S}_i(t)}{dt} = \mathbf{N}_R \mathbf{v}(t) \quad \text{and} \quad \frac{d\mathbf{S}_d(t)}{dt} = \mathbf{L}_0 \mathbf{N}_R \mathbf{v}(t). \quad (4)$$

Simplification of Equation (4) yields the relation between the dependent and independent species as

$$\frac{d\mathbf{S}_d(t)}{dt} - \mathbf{L}_0 \frac{d\mathbf{S}_i(t)}{dt} = \mathbf{0}. \quad (5)$$

It can be seen by integrating Equation (5) that \mathbf{S}_d is related to \mathbf{S}_i by a constant vector $\mathbf{T} = [T_1 T_2 \dots T_{m-m_0}]^T$ such that

$$\mathbf{S}_d(t) - \mathbf{L}_0 \mathbf{S}_i(t) = [-\mathbf{L}_0 \ \mathbf{I}] \begin{bmatrix} \mathbf{S}_i(t) \\ \mathbf{S}_d(t) \end{bmatrix} = \mathbf{T}. \quad (6)$$

Introducing a matrix $\mathbf{\Gamma}$ in place of $[-\mathbf{L}_0 \ \mathbf{I}]$ in Equation (6), we can deduce that $\mathbf{\Gamma}\mathbf{S} = \mathbf{T}$. The $(m - m_0) \times n$ matrix $\mathbf{\Gamma}$ is called the conservation matrix, as it relates the species vector \mathbf{S} to \mathbf{T} , the vector of the conserved moieties. Each row of the conservation matrix $\mathbf{\Gamma}$ represents a distinct conserved cycle of the network. The values of the vector \mathbf{T} can in practice be obtained by substituting the initial conditions of the species into the relation $\mathbf{T} = \mathbf{S}_d(0) - \mathbf{L}_0\mathbf{S}_i(0)$. The key to accurate computation of the conservation matrix $\mathbf{\Gamma}$ is thus a problem of computing the matrix \mathbf{L}_0 . Sauro and Ingalls (2004) review a number of methods that are known to carry out this process. However, these methods all face the problem of numerical instability for very large networks. An alternative method was proposed recently by Vallabhajosyula et al. (2006) to overcome this problem. This method involves the use of Householder QR algorithm

for robust computation of the \mathbf{L}_0 matrix, which ensures that the correct number of dependencies in the network are identified. The reduced system, which involves the dependent species, can then be used for further analysis.

Another aspect of interest in the study of biochemical networks, especially from the point of view of complexity reduction, relates to the evaluation of the Jacobian matrix. The dynamics of the network described by Equation (1) are often nonlinear, and hence a linearized system can be obtained by perturbing the system. Ignoring second order terms, a linear equivalent of Equation (1) is given by

$$d(\delta\mathbf{S})/dt = \mathbf{J}_F \delta\mathbf{S} \quad (7)$$

where \mathbf{J}_F is the Jacobian matrix for the full system and $\delta\mathbf{S}$ represents infinitesimal changes in \mathbf{S} . The dependencies in the network between species are reflected in the Jacobian matrix \mathbf{J}_F being singular. The full Jacobian matrix can be built using the relationship between dependent and independent species, given by \mathbf{L} and \mathbf{N}_R matrices. This can be shown after modifying an expression given by Heinrich and Schuster (1996) as

$$\mathbf{J}_F = \mathbf{N} \epsilon = \mathbf{L} \mathbf{N}_R \epsilon \quad (8)$$

where ϵ is the Elasticity coefficient matrix, with $\epsilon_{ij} = \partial v_i / \partial S_j$. A non-singular reduced Jacobian matrix \mathbf{J}_R , can be constructed using conservation analysis. Indeed, it can be shown, using Equation (2) and Equation (7), that \mathbf{J}_R is given by the relation $\mathbf{J}_R = \mathbf{N}_R \epsilon \mathbf{L}$.

2.2 Model Reduction by Time-scale Separation

A number of methods have been proposed to reduce the complexity of biological systems which involve processes spread over widely different time scales. A simulator attempting to simulate such a system will have to take very small time-steps in order to account for the fastest reactions, while the slowest reactions barely change during this time-period. Hence it becomes necessary, from the point of view of simulation, to develop algorithms that take this disparity in scale into account. An early attempt at describing reduced models was by Briggs and Haldane (1925), who showed how the Michaelis-Menten ratelaw assumption is justified in enzyme kinetics. More recently, Schauer and Heinrich (1983) proposed the Quasi Steady State Assumption (QSSA) method, whose validity can be shown by carrying out a singular perturbation on the kinetics of the system, as described by Segel and Slemrod (1989). This approximation method is also known by the name of Bodenstein-Semonov kinetics or also by the name of Pseudo-Steady-State assumption (Rao and Arkin 2003). The idea behind QSSA is that the intermediate species can be removed from the system by assuming that the rate of change of these species is very

small compared to the overall dynamics. Consequently, the differential equations determining the rates of change for these species are eliminated from the dynamics. However, the choice of the species to which QSSA should be applied is ambiguous, and hence the error in the solution can vary based on which species are chosen, as has been described by Turányi (1993) in the context of computational combustion modeling. A related method is the Partial-Equilibrium Approximation (PEA) where the fast reactions are assumed to equilibrate in a short period of time relative to the system time scale. Other methods that have also been proposed for reducing the dimensionality of the system include minimal entropy production trajectories (MEPT) by Lebiecz (2004) and methods based on the thermodynamics projection operator by Gorban et al. (2000).

An important and well known method used in time-scale decomposition of dynamical systems is the Computational Singular Perturbation Method (CSP)—see Lam and Goussis (1988, 1994), Zagaris et al. (2004a, 2004b). It has been used widely in combustion modeling, see Valorani et al. (2003). This method employs an iterative algorithm to approximate the slow manifold, which is the subsystem comprising of the slow variables in the system. The true slow manifold is a geometric abstraction, and is described by Fenichel’s geometric singular perturbation theory (Fenichel 1979). The asymptotic accuracy of the approximation of the CSP manifold improves with each iteration by one order of ϵ , a small parameter, which is representative of the ratio between the slow and fast time-scales.

More recently, dynamics of systems with fast and slow time-scales have been studied from a geometrical perspective (Nguyen and Fraser 1989). The fast evolving reactions drive the dynamics onto a subsystem comprising of the slow reactions, which define what is known as the Inertial Manifold. Once the system reaches this abstract structure in the phase space of the reactions, the fast reactions will stay on this inertial manifold, while the evolution of the slow reactions themselves will be determined by the trajectory set by the initial conditions. The inertial manifold itself can be computed by using an iterative approach where the algebraic expressions for the fast reactions are solved for the slow variables. This is equivalent to the intersection of the nullclines, especially for higher order systems.

This leads to a more localized approach of Mass and Pope (1992), where the Jacobian of the system, which defines the tangential space of the dynamics, is analyzed for vanishing eigenvectors corresponding to the fast time-scales. This was further refined by Mass (1998) who proposed a more efficient algorithm for the computation of the low-dimensional manifolds. This is known as the Intrinsic Low-Dimensional Manifold (ILDm) method, and has been widely used to solve stiff problems in the study of combustion and reactive chemical processes. In comparison with the previously listed methods such as Quasi Steady

State Assumption (QSSA) the Partial Equilibrium Approximation (PEA) based methods, the ILDM method has the advantage that it can automatically partition the higher order initial system into slow and fast subsystems based on the eigenvector space of the Jacobian near the operating point. This partitioning is determined by an Eigen analysis of the linearized version of the system in Equation (1) in a small neighborhood around the operating point \mathbf{S}^* . The resulting system, ignoring higher order terms, can be written as

$$\frac{d\mathbf{S}(t)}{dt} = \mathbf{f}(\mathbf{S}^*, \mathbf{p}) + \mathbf{J}_{\mathbf{F}}(\mathbf{S}(t) - \mathbf{S}^*) \quad (9)$$

where the Jacobian $\mathbf{J}_{\mathbf{F}}$ is evaluated at the operating point. This linearized system is representative of the qualitative dynamics of the actual system, and provides a reasonably accurate picture of the system behavior close to the operating point. In particular, the direction in which the system will move depends on the eigenvalues of the Jacobian. If it can be assumed that the eigenvalues of the Jacobian are given by λ_i , $i = 1, \dots, m$, it can be inferred following a change of basis using the eigenvectors, that $\tau_i = 1/|\text{Real}(\lambda_i)|$ represents the time taken for the system to evolve along the eigenvector for the i^{th} mode. This indicates that the real parts of the eigenvalues can be used to distinguish between slow and fast modes. However this approach has its limitations when some eigenvalues are grouped closely or are degenerate, leading to an ill-conditioned system that presents numerical accuracy and stability issues for the solver.

The ILDM method of Mass and Pope (1992) overcomes this limitation by applying similarity transformations to the Jacobian matrix, which yielding a real Schur form. The advantage of using the Schur decomposition lies in the fact that the resulting matrix has a quasi upper triangular structure, with diagonal blocks representing the eigenvalues. In other words, they may be 1×1 blocks for real eigenvalues and 2×2 blocks for complex-conjugate eigenvalues. This can be written as

$$\mathbf{Q}^T \mathbf{J} \mathbf{Q} = \begin{bmatrix} \mathbf{S}_{slow} & \mathbf{S}_{coup} \\ \mathbf{0} & \mathbf{S}_{fast} \end{bmatrix}. \quad (10)$$

It should be noted that in the above relation, \mathbf{Q} is the orthogonal basis of Schur vectors and the submatrices \mathbf{S}_{slow} , \mathbf{S}_{fast} and \mathbf{S}_{coup} are quasi-upper triangular matrices, where the eigenvalues are reordered by Givens rotations (Golub and van Loan 1996). The key component of the ILDM lies in eliminating the coupling matrix \mathbf{S}_{coup} , by seeking a solution \mathbf{Z}_r of the Sylvester equation, which is given by

$$\mathbf{S}_{slow} \mathbf{Z}_r - \mathbf{Z}_r \mathbf{S}_{fast} + \mathbf{S}_{coup} = \mathbf{0}. \quad (11)$$

A change of basis can be effected once \mathbf{Z}_r is obtained by constructing the similarity transformation matrices \mathbf{Y}_r

which are given by

$$\mathbf{Y}_r = \mathbf{Q} \begin{bmatrix} \mathbf{I} & \mathbf{Z}_r \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \quad (12)$$

These similarity transformation matrices can then be used to yield the decoupled system comprising the fast and slow matrices. It can be verified that a decoupled system can indeed be obtained by substituting Equation (12) into $\mathbf{Y}_r^{-1} \mathbf{J} \mathbf{Y}_r$ and using Equation (10) to yield

$$\mathbf{Y}_r^{-1} \mathbf{J} \mathbf{Y}_r = \begin{bmatrix} \mathbf{S}_{slow} & 0 \\ \mathbf{0} & \mathbf{S}_{fast} \end{bmatrix}. \quad (13)$$

A change of basis can be carried out by applying \mathbf{Y}_r to the original species vector, thereby transforming \mathbf{S} to a new vector \mathbf{x} such that $\mathbf{x}^T = [\mathbf{x}_{slow}^T \ \mathbf{x}_{fast}^T]$. It should be noted that the dimension of \mathbf{x}_{slow} is r , which represents the size of the slow manifold, while the remaining $m - r$ dimensions represent the fast system. This transformation is also applied to the function \mathbf{f} in Equation (1) to yield a new vector \mathbf{g} such that $\mathbf{g}^T = [\mathbf{g}_{slow}^T \ \mathbf{g}_{fast}^T]$. The vectors \mathbf{x} and \mathbf{g} can therefore be written as

$$\mathbf{x} = \mathbf{Y}_r^{-1} \mathbf{S}, \quad \mathbf{g} = \mathbf{Y}_r^{-1} \mathbf{f}. \quad (14)$$

It can be seen that this similarity transformation completely decouples the fast and slow subsystems, allowing them to be solved separately. The final form of the transformed system is given by

$$\frac{d\mathbf{x}_{slow}}{dt} = \mathbf{g}_{slow}(\mathbf{x}) \quad (15)$$

and
$$\epsilon \frac{d\mathbf{x}_{fast}}{dt} = \mathbf{g}_{fast}(\mathbf{x}) \quad (16)$$

where ϵ is given by $\epsilon = \tau_{r+1}$, and is the singular perturbation parameter. An improvement on this approach involves the use of an error criterion (Deuffhard and Heroth 1996, Zobeley et al. 2005). Using an error criterion allows for adjustment of the size of the low-dimensional system based on its closeness to the equilibrium point for numerical accuracy. Shaik et al. (2005) proposed such an error-control based computational scheme for chemical kinetics, while Handrock-Meyer et al. (2001) discuss the use of a splitting parameter to obtain the minimum dimension of the slow system and illustrate their approach by means of an example.

The existence of conserved moieties in a system, while reducing the dimension of the system, can involve some moieties that include species from both the fast and slow subsystems. The analysis of such systems should exploit both the conservation matrix information as well as the partitioning into fast and slow subsystems. However, it does not appear that this topic has been studied in detail, apart from

the discussion presented by Heinrich and Schuster (1996) in their chapter on time hierarchy in metabolism, as it applies to the case of the Rapid Equilibrium Approximation. We therefore think that it would be worth investigating how the other approaches can be improved by incorporating results of conservation analysis into the time-scale separation. In this regard, one can ask if the slow and fast subsystems can be built by restricting the analysis to only the independent species obtained by conservation analysis.

It can be noted that while the conservation analysis identifies the dependent and independent species, along with the conservation laws that relate them, this analysis holds true for the total system that comprises both the fast and the slow systems. When complexity reduction methods such as ILDM are used to partition the original system into fast and slow components, they are followed by application of an approximation technique such as QSSA to average the fast system and construct an algebraic relation between the fast and slow systems. Such an approximation changes the system, where the original conservation laws no longer hold. In other words, the conserved moieties will no longer be constants, but will themselves form a part of the system, with a time-varying component. In this situation, the equations that contribute to each of the conserved moieties will have to be replaced with the averaged counterpart to yield a set of additional equations that can be appended to the averaged network. We observe at this point that only those conserved moieties that involve a species that is part of the averaged fast system will have to be included. This may result in a small increase in the number of additional equations to be solved, but would result in the proper treatment of conservation analysis in a time-scale reduction problem.

3 FUNCTIONAL COMPLEXITY

Biological systems often possess hierarchically distributed subsystems which have one or more functions such as oscillators or switches (see Tyson et al. 2003, Wolf et al. 2003 for details). It can therefore be argued that components with an identifiable function can be replaced by simpler mathematical constructs, thereby reducing the complexity. This methodology is useful in understanding the complexity of biological systems by using a modular approach, as described by Hartwell et al. (1999). Such a modular approach can indeed be shown to be feasible if it is known clearly what the inputs and outputs to each subsystem are, and care is taken to minimize external interactions with species inside the module.

In the following, we show how a network that contains a number of components can be simplified by replacing them with modules whose function is equivalent. The network shown in Figure 1 has an output response that is maximal at two frequencies and is shown in Figure 2. This network was constructed to study interactions between a set of genes and

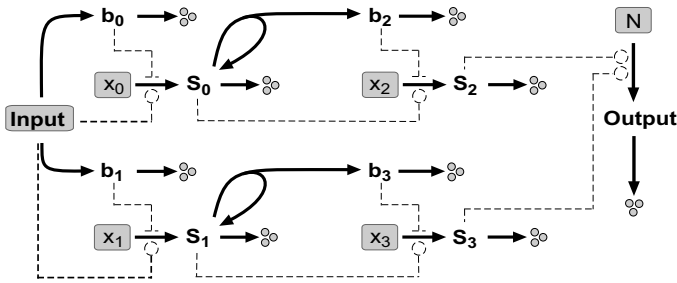


Figure 1: A Network with Maximal Response at Two Frequencies, in which Dotted Lines Indicate Regulation with Bars for Inhibition and Circles for Activation, Grey Boxes are Representative of Genes, and b_0, b_1, b_2, b_3 are Transcription Factors that Regulate the Genes.

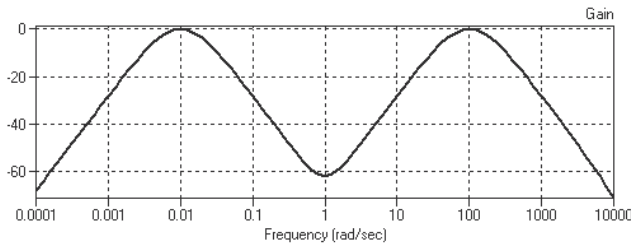


Figure 2: Frequency Response (in dB) of the Network at the Output for the Network in Figure 1.

their expressed proteins using simple relations. The low gain of this network at other frequencies far from the maximal frequencies indicates that input fluctuations at frequencies far from the maximal frequencies are not detected at the output.

The objective in this example is to simplify this network by replacing components by simpler modules. The frequency response implies that the network involves two band-pass filters, one of which operates at the lower frequency and the other operates at the higher frequency. Hence a set of band-pass filters could replace the network in Figure 1. The network then takes a conceptually simpler form shown in Figure 3.

The band-pass filters themselves could be constructed by connecting two low-pass filters, as shown in Figure 4. It can be noted that a low-pass filter is representative of the process of transcription of a gene. This process involves an effector protein whose strength has an effect on the production of expressed protein, which is the output. If the fluctuation in the strength of the effector protein is slow, the rate of protein production is high. On the other hand, if the fluctuation of the input is rapid, the transcription mechanism cannot keep up with the fluctuations, and hence the output, namely the expression of the protein, is diminished. This process is therefore representative of a low pass filter. Construction

of the band-pass filter involves the use of the feed forward loop (FFL) architecture, based on the work by Mangan and Alon (2003). This involves two low-pass filters, with the second low-pass filter being activated by the input and simultaneously, being repressed by the output of the first low pass stage. The combined effect of the activation and repression is to generate a band-pass filter behavior between the input and output.

This is a small example that illustrates how we can conceptualize the representation of a network in a form where components have been replaced by functional equivalents. This can also be achieved using a few basic modules. These basic modules as well as the principles behind the construction of more complex networks will be detailed shortly in a forthcoming paper, currently in preparation.

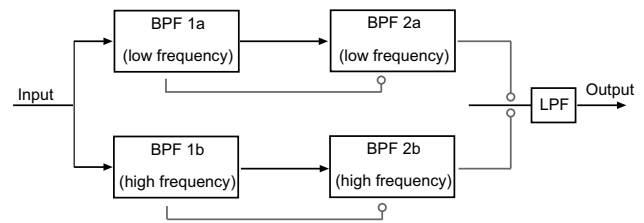


Figure 3: A Block Representation of the Network Shown in Figure 1 Using Band-Pass Filters (BPF).

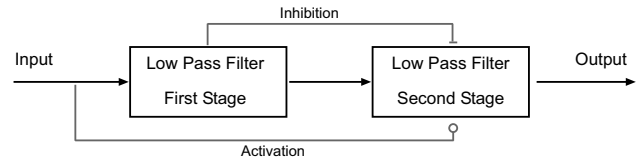


Figure 4: A Band-Pass Filter Constructed by Staging Two Low-Pass Filters.

4 SUMMARY

This paper presents a brief overview of the problem of complexity reduction of biological networks, and reviews various methods known in the literature. This is an important problem, especially from the viewpoint of modeling and simulation. Systems with a wide range of time scales tend to involve reactions with varied velocities, which in turn result in fast and slow processes. This results in a stiff system, which requires very small steps or sophisticated time stepping algorithms for the simulator to maintain numerical stability. We discuss two ways of speeding up such simulations, with the first step involving conservation analysis to extract the independent species, and a second step involving application of an appropriate time-scale reduction method to reduce the dimensionality of the model. Particular attention is paid to the times-scale reduction method

based on the Intrinsic Low Dimensional Manifold (ILDm) method. The development of a method that incorporates the salient features of both these approaches into one single method is still however an open problem.

The second issue that is addressed in this paper pertains to the representation of complex networks in a form where components with a specified function can be replaced with simpler modules with a known mathematical description. We have shown this with an example of a network with two peaks in the frequency domain, by replacing its components with band-pass filters. This is useful in simplifying a given network into components that are functionally simple, thereby reducing the complexity of the overall network.

REFERENCES

- Bhattacharjee, B., D. A. Schwer, P. I. Barton, and W. H. Green. 2003. Optimally-reduced kinetic models: Reaction elimination in large-scale kinetic mechanisms. *Combustion and Flame* 135(3):191-208.
- Briggs, G. E. and J. B. S. Haldane. 1925. A note on the kinetics of enzyme action. *Biochemical Journal* 19(2):338-339.
- Deuffhard, P. and J. Heroth. 1996. Dynamic dimension reduction in ODE models. In *Scientific Computing in Chemical Engineering*, ed. F. Keil, 24-43. Springer Verlag, Berlin.
- Dvorák, I. and J. Siska. 1989. Analysis of metabolic systems with complex slow and fast dynamics. *Bulletin of Mathematical Biology* 51(2):255-274.
- Edwards, K., T. F. Edgar, and V. I. Manousiouthakis. 1998. Kinetic model reduction using genetic algorithms, *Computers and Chemical Engineering* 22(1):239-246.
- Eggels, R., J. Louis, J. Kok, and L. DeGoey. 1997. Comparison of conventional and low-dimensional manifold methods to reduce reaction mechanisms. *Combustion, Science and Technology* 123:347-362.
- Fenichel, N. 1979. Geometrical singular perturbation theory for ordinary differential equations, *Journal of Differential Equations* 31:53-98.
- Gardner, M. R. and W. R. Ashby. 1970. Connectance of large dynamic (Cybernetic) systems: Critical values for stability. *Nature* 228:784.
- Golub, G. H. and Ch. F. van Loan. 1996. *Matrix Computation*. Johns Hopkins University Press.
- Handrock-Meyer, S., L. V. Kalachev, and K. R. Schneider. 2001. A method to determine the dimension of long-time dynamics in multi-scale systems. *Journal of Mathematical Chemistry* 30(2):133-160.
- Hartwell, L. H., J. J. Hopfield, S. Leibler, and A. W. Murray. 1999. From molecular to modular biology. *Nature* 402:C47-C52.
- Heinrich, R., S. M. Rapoport, and T. A. Rapoport. 1977. Metabolic regulation and mathematical models. *Progress in Biophysics and Molecular Biology* 32:1-82.
- Heinrich, R. and S. Schuster. 1996. *The Regulation of Cellular Systems*. Chapman & Hall.
- Kremling, A., S. Fischer, T. Sauter, K. Bettenbrock, and E. D. Gilles. 2004. Time hierarchies in the *Escherichia coli* carbohydrate uptake and metabolism. *Biosystems* 73(1):57-71.
- Lam, S. H. and D. A. Goussis. 1988. Understanding complex chemical kinetics with computational singular perturbation. In *Twenty-Second Symposium on Combustion*, The University of Washington, Seattle, 931-941. The Combustion Institute, Pittsburgh.
- Lam, S. H. and D. A. Goussis. 1994. The CSP method for simplifying kinetics. *International Journal of Chemical Kinetics* 26:461-486.
- Lebiedz, D. 2004. Computing minimal entropy production trajectories: an approach to model reduction in chemical kinetics. *Journal of Chemical Physics* 120(15):6890-6897.
- Lebiedz, D., V. Reinhardt, and J. Kammerer. 2005. Novel trajectory based concepts for model and complexity reduction in (bio)chemical kinetics. In *Workshop on Model Reduction and Coarse-graining Approaches for Multi-scale Phenomena*, Leicester, UK, in press.
- Liao, J. C., and E. N. Lightfoot. 1988. Lumping analysis of biochemical reaction systems with time scale separation. *Biotechnology and Bioengineering* 31(8):869-879.
- Liao, J. C., and J. Delgado. 1993. Advances in metabolic control analysis. *Biotechnology Progress* 9:221-233.
- Mangan, S. and U. Alon. 2003. Structure and function of the feed-forward loop network motif. In *Proceedings of the National Academy of Sciences (PNAS)*, 100(21):11980-11985.
- Mass, U. 1998. Efficient calculation of intrinsic low-dimensional manifolds for the simplification of chemical kinetics. *Computing and Visualization in Science* 1:69-91.
- Mass, U. and S. B. Pope. 1992. Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space. *Combustion and Flame* 88:239-264.
- Maurya, M. R., S. J. Bornheimer, V. Venkatasubramanian, and S. Subramanian. 2005. Reduced-order modeling of biochemical networks: application to the GTPase-cycle signalling module. *IEE Proceedings - Systems Biology* 152(4):229-242.
- Nguyen, A. H. and S. J. Fraser. 1989. Geometrical picture of reaction in enzyme kinetics. *Journal of Chemical Physics* 91:186-193.
- Okino, M. S. and M. L. Mavrovouniotis. 1998. Simplification of mathematical models of chemical reaction systems. *Chemical Reviews* 98(2):391-408.

- Park, D. J. M. 1974. The hierarchical structure of metabolic networks and the construction of efficient metabolic simulators. *Journal of Theoretical Biology* 46:31-74.
- Rabitz, H. and M. Kramer, and D. Dacol. 1983. Sensitivity analysis of chemical kinetics. *Annual Review of Physical Chemistry* 34:419-461.
- Rao, C. V. and A. P. Arkin. 2003. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. *The Journal of Chemical Physics* 118(11):4999-5010.
- Reder, C. 1988. Metabolic control theory: a structural approach. *Journal of Theoretical Biology* 135(2):175-201.
- Reich, J. G. and E. E. Selkov. 1981. *Energy metabolism of the cell*. Academic Press, London.
- Rojdestvenski, I., M. Cottam, Y. I. Park, G. and Oquist. 1999. Robustness and time-scale hierarchy in biological systems. *Biosystems* 50(1):71-82.
- Roussel, M. R. and S. J. Fraser. 2001. Invariant manifold methods for metabolic model reduction. *Chaos* 11(1):196-206.
- Sauro, H. M. and B. Ingalls. 2004. Conservation analysis in biochemical networks: computational issues for software writers. *Biophysical Chemistry* 109(1):1-15.
- Schauer, M. and R. Heinrich. 1983. Quasi-steady state approximation in the mathematical modeling of biochemical reaction networks. *Mathematical Biosciences* 65:155-170.
- Segel, L. A. and M. Selmrod. 1989. The quasi-steady-state assumption: a case study in perturbation. *SIAM Review* 31(3):446-477.
- Shaik, O. S., J. Kammerer, J. Gorecki, and D. Lebiedz. 2005. Derivation of a quantitative minimal model from a detailed elementary-step mechanism supported by mathematical coupling analysis. *The Journal of Chemical Physics* 123:234103.
- Turányi, T. 1990. Analysis of Complex Kinetic Systems. Tools and Applications. *Journal of Mathematical Chemistry*, 5, 203-248.
- Turányi, T., A. Tomlin, and M. Pilling. 1993. On the error of the quasi steady state approximation. *Journal of Physical Chemistry* 97:163-172.
- Tyson, J. J., K. C. Chen, and B. Novak. 2003. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory signaling pathways in the cell. *Current Opinion in Cell Biology* 15:221-231.
- Vallabhajosyula, R. R., V. Chickarmane, and H. M. Sauro. 2006. Conservation analysis of large biochemical networks. *Bioinformatics* 22(3):346-353.
- Valorani, M., H. Najm, and D. Goussis. 2003. CSP analysis of a transient flame-vortex interaction: time scales and manifolds, *Combustion and Flame* 134:35-53.
- Wolf, D. M. and A. P. Arkin. 2003. Motifs, modules and games in bacteria. *Current Opinion in Microbiology* 6:125-134.
- Zagaris, A., H. G. Kaper, and T. J. Kaper. 2004a. Analysis of the computational singular perturbation reduction method for chemical kinetics. *Journal of Nonlinear Science* 14:59-91.
- Zagaris, A., H. G. Kaper, and T. J. Kaper. 2004b. Fast and slow dynamics for the computational singular perturbation method. *Multiscale Modeling and Simulation* 2:613-638.
- Zobeley, J., D. Lebiedz, J. Kammerer, A. Ishmurzin, and U. Kummer. 2005. A new time-dependent complexity reduction method for biochemical systems. In *Transactions on Computational Systems Biology*, ed. C. Priami et al. Springer Verlag, LNBI 3380:90-110.

AUTHOR BIOGRAPHIES

RAVISHANKAR RAO VALLABHAJOSYULA is an engineer by training, and is currently a postdoctoral fellow in Computational & Systems Biology at the Keck Graduate Institute. He has 15 years of experience related to modeling and simulation of dynamical systems and was previously involved with development of hydraulic simulators for the oil and gas pipeline transportation industry, where he developed software for Real-Time and Predictive Modeling tools. He then took a research position at the Keck Graduate Institute, where his current interests in Systems Biology include dynamical aspects of biochemical systems, stochastic modeling and complexity reduction problems. His PhD in Mechanical Engineering from Rice University addressed the topic of efficient wavelet based numerical methods for stochastic mechanics. He obtained his B.Tech degree in Mechanical Engineering from Indian Institute of Technology Madras, Chennai, India. His e-mail address is <rrao@kgi.edu>.

HERBERT SAURO was originally educated as a biochemist/microbiologist but became interested in simulation and theory to understand cellular networks after accidentally coming across a paper by David Garkfinkel on the simulation of glycolysis. He wrote one of the first bio-chemical simulators for the PC (SCAMP) in the 1980s to assist work on extending metabolic control analysis (a theory closely related to biochemical systems theory). However, with the lack of community interest in systems biology during the late 80s and early 90s, he left science to start a software company and offer consultancy work to finance firms in the UK. With the surge in interest in systems biology in the US in the late 90s, he secured a position at Caltech to assist in the development of the Systems Biology Markup Language. Since then he moved to a faculty position at the Keck Graduate Institute where he continues to do research on network motifs, theory and software. His email address is <hsauro@kgi.edu>.