Linking Discrete and Stochastic Models: The Chemical Master Equation as a Bridge between Process Hitting and Proper Generalized Decomposition

Courtney Chancellor[1,2], Amine Ammar[4], Francisco Chinesta[2], Morgan Magnin[1,3], and Olivier Roux[1]

[1] L'UNAM Université, École Centrale de Nantes, IRCCyN UMR CNRS 6597, France
[2] L'UNAM Université, École Centrale de Nantes, GeM UMR CNRS 6183, France
[3] National Institute of Informatics, Tokyo 101-8430, Japan
[4] Arts et Metiers ParisTech. Angers, France

Abstract. Modeling frameworks bring structure and analysis tools to large and non-intuitive systems but come with certain inherent assumptions and limitations, sometimes to an inhibitive extent. By building bridges in existing models, we can exploit the advantages of each, widening the range of analysis possible for larger, more detailed models of gene regulatory networks. In this paper, we create just such a link between Process Hitting [6,7,8], a recently introduced discrete framework, and the Chemical Master Equation in such a way that allows the application of powerful numerical techniques, namely Proper Generalized Decomposition [1,2,3], to overcome the curse of dimensionality. With these tools in hand, one can exploit the formal analysis of discrete models without sacrificing the ability to obtain a full space state solution, widening the scope of analysis and interpretation possible. As a demonstration of the utility of this methodology, we have applied it here to the p53-mdm2 network [4,5], a widely studied biological regulatory network.

1 Introduction

Our ability to gather data in the context of gene regulatory networks has skyrocketed in the past decades: technology has given scientists an unprecedented ability to take in large amounts of raw data on the genome and genomic expression. The scale of this newly available data is massive and uninterpretable without applying formal analysis. Computational tools are invaluable in this respect: to put an otherwise incomprehensible data set into the context of a modeling framework allows scientists to understand the behaviors of a system and make predictions thereof. Once one imposes a model, however, one is confined to the inherent assumptions and limitations of that method. The resulting compromise is why there exist many varieties of structures, each exploiting certain advantages while accepting certain limitations. There is a great interest in systems biology in how to best navigate this choice or, better yet, how to avoid it

altogether: by building bridges in existing models, we can utilize the best aspects of each, widening the range of analysis possible for larger, more realistic models of gene regulatory networks.

In this paper, we begin what will become a body of work dedicated to this goal: approaching biological systems in inventive or innovative ways to maximize the accuracy and utility of modeling structures. We start with a recently introduced modeling structure called Process Hitting [6,7,8], a discrete model in which one tracks qualitative shifts of the system. Like other discrete frameworks, it possesses certain advantages in model fitting and model checking. In addition, Process Hitting offers straightforward ways of incorporating temporal and stochastic properties and, moreover, can be thought to contain the Generalized Logical Network in that it has been shown that the interaction graph and discrete Thomas parameters can be derived from the Process Hitting model. Simulation is required to obtain a full description of the local behaviors of the system, a computationally expensive and sometimes inhibitive aspect. Here, we propose a method to solve the system, obtaining a solution for all time, by translating Process Hitting to an equivalent differential equation form, namely the Chemical Master Equation (or CME), with the intention of applying a novel numerical technique,the Proper Generalized Decomposition (or PGD)[1,2,3], to overcome problems of dimensionality. Since this paper makes use of state-of-the art methods which readers may not yet be familiar with, we will briefly introduce Process Hitting, Proper Generalized Decomposition and the Chemical Master Equation in Section 2, outlining advantages, disadvantages and important features of each. In this section, we will also introduce the p53-mdm2 network [4,5,11], a biologically relevant model for which an extensive body of work exists and to which we will apply our methodology. Once these structures have been defined, the translation of Process Hitting to a special, discretized form of the Chemical Master Equation is outlined in Section 3. Application to the p53-mdm2 model and the analysis of results can be found in Section 4, along with a clearly defined path for future work. When fully realized, our proposed methodology has the potential to allow scientists to construct models without explicit knowledge of reaction kinetics, use a variety of analysis tools which exploit the discrete, formal structure of Process Hitting and, finally, efficiently solve this complete system for in-depth analysis.

2 Frameworks and Methods Used

The foundation of this paper is in patching together novel techniques to form a better, more complete modeling structure. While most readers may be familiar with discrete and stochastic modeling, for example, it may not be the case for Process Hitting or the Chemical Master Equation in particular. Here, we have included a brief description of all of the frameworks and methods used: although no one section gives neither a full nor formalized description, we hope to give an intuitive notion of each, as well as an understanding of their significance. We begin by introducing the p53-mdm2 system, followed by the relevant modeling frameworks, and finally by the numerical method of Proper Generalized Decomposition.

2.1 Introductory Example: The p53-mdm2 Network

We have chosen to apply our methodology to a proven and biologically relevant network, the p53-mdm2 regulatory system [4,5,11]. The protein p53 is a transcription factor for a variety of systems, particularly those relating to arrested cell growth, DNA repair and cell death. When a cell incurs DNA damage, p53 concentration levels rise, inducing cell repair or, if the damage persists, apoptosis, thus preventing the spread of genetically unstable cells. p53 is strictly regulated by ubiquitin ligase mdm2 via a negative circuit. Since this system is of such interest in cancer research, there exists an extensive body of experimental and modeling work dedicated to its study.

In the presence of DNA damage, high concentrations of p53 can promote cellular repair but can be lethal with long-term exposure. As a counter balance, these same high concentration levels of p53 up-regulate the transcription of gene mdm2, increasing the concentration of its protein in the cytoplasm of the cell. When this concentration is high enough, mdm2 moves into the nucleus, where it blocks further transcription of p53 and facilitates the degradation of existing proteins, preventing the cell from going into apoptosis. Since the ability to express p53 at sufficient levels is required for cellular repair, this inhibition cannot always prevail: byproducts of cellular damage help the degradation of mdm2 in the nucleus, and translocation itself is inhibited by p53 at a level much lower than the activation of mdm2. These interactions are summarized as a directed graph in Figure 1.
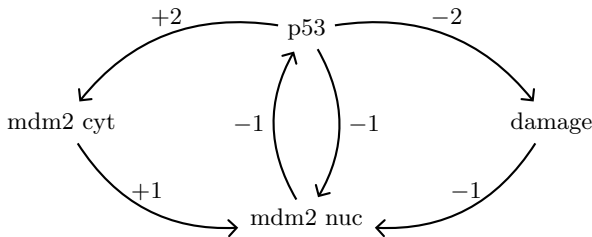

Fig. 1. Representation of p53-mdm2 network as a directed graph: interactions are summarized as activation or inhibition (+ or −), including discretized concentration levels ([0, 1, 2...]) defined by threshold values. Since mdm2 behaves differently in regards to the system depending on its location in the cell, cytoplasmic and nucleic mdm2 are represented separately.

2.2 A Brief Introduction to Process Hitting

Given the biological description of the system, we wish to construct a model for further analysis and study. Process Hitting is, more generally, a framework for modeling concurrent processes but is particularly apt for biological regulatory networks in that it conceals the kinetic mechanisms by which the system moves, describing instead the qualitative changes that may occur. In this fashion, it is possible to model a system with observational data and only partial knowledge of

its inner workings. Although its structure seems simplistic, Process Hitting can capture complex dynamics and easily lends itself to model-checking, by which one can determine whether or not certain desired features are preserved by the model. In addition, temporal and stochastic properties can be naturally integrated into the Process Hitting structure. A full description of this framework including its implementation, can be found in [6,7,8].

In Process Hitting, here on referred to as PH, all interacting species —enzymes, genes, proteins, etc. —are abstracted as *sorts*. The sorts of the p53-mdm2 system are cellular damage, nuclear mdm2, cytoplasmic mdm2 and p53, which are given the labels Dam, Mn, Mc, and p53, respectively. These sorts are then subdivided into *processes*, which could represent concentration levels, spatial configuration, or any other form which has a distinct qualitative impact on the system. Dam, for example, has two processes, Dam 0 and Dam 1, the absence and the presence of cellular damage. Conversely, p53 contains three processes which represent the relevant concentration ranges of p53 in the cell. Processes interact with one another via *actions*, in which processes *hit* one another to create a *bounce* to some new level of the same sort, wherein we find the namesake of "Process Hitting". These actions move the state space one level at a time and are, therefore, asynchronous. For gene regulatory networks, processes are often abstractions of relevant concentration ranges, discretized domains of real numbers, and actions represent varying action and inhibition reactions. For instance, we know that, when at a very high level, p53 up-regulates the level of cytoplasmic mdm2. In PH action terms, this is demonstrated as

$$p53\_2 \rightarrow Mn\_0 \nearrow 1$$
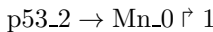
which reads "p53_2 hits Mn_0 to bounce to Mn_1", as demonstrated in Figure 2.

In this structure, the absence of an activator is equivalent to inhibition and vice versa. Therefore, whenever p53 is below its activating threshold, it is effectively an inhibitor:

$$p53\_1 \rightarrow Mn\_1 \nearrow 0$$

$$p53\_0 \rightarrow Mn\_1 \nearrow 0$$
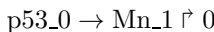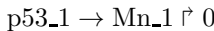
Whether or not this is a biologically valid assumption is subject to the modeler, who may remove any unwanted hits as suits the system in question. But what if a process can be influenced by more than one factor, as is the case with nucleic mdm2 which is influenced negatively by p53 and cellular damage but positively by cytoplasmic mdm2? PH does not enforce separability in its framework, that is, that effects be additive: the impact of two activators, for example, is not necessarily equivalent to the addition of their individual impacts. It is instead up to the modeler to define the results of combined influences. In order to remain consistent with the formalism of PH, we must introduce *cooperative sorts*, a representation of the collective influence of species. In our network, this leads to the creation of collective sort p53McDam which interacts with the normal sort Mn. In addition to defining the actions each combination of p53McDam will perform on Mn, we must also add actions between the normal sorts p53,
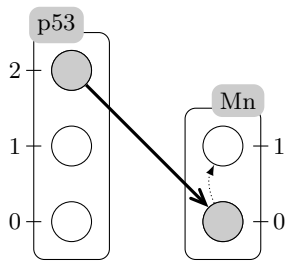
Fig. 2. Example of a Process Hitting action. Here, we show sorts p53 and Mn, boxes which contain processes. If the system is such that process p53_2 and Mn_0 are actice at the same time; p53_2 will have the chance to "hit" Mn_0, indicated by the solid arrow, which will then "bounce" to Mn_1, indicated by the light, dashed arrow.

Mc and Dam, and p53McDam which update the cooperative sort such that its state instantaneously reflects the current state of the system. Note that we do not need to indicate how an action is performed (the kinetics) nor dictate any particular behaviors as is the case in other discrete frameworks. Although we may abstractly incorporate kinetic concepts such as threshold values, as in the example of p53 via the discretization of concentration levels ([0, 1, 2]), we do not need to quantitatively define these thresholds. A list of the actions and the diagram of the Process Hitting of the p53 system can be found in the appendix.

2.3 The Chemical Master Equation

The Chemical Master Equation, or CME [1], is considered a canonical stochastic model in biological regulatory networks in which the modeler assumes full kinetic knowledge of the system. The vector $z = [z_i] \in \mathbb{Z}_{\geq 0}^{N_{sp}}$, $i = 1 \cdots N_{sp}$ contains the discretely valued counts of molecules for a given species i. Rather than tracking the state of the system as it varies in time according to an underlying deterministic process, we think about the system in terms of the probability $P(z, t|z_0, t_0)$ of existing at a certain state z at any given time t given some initial condition. From any state, reactions occur which move the system to a new configuration according to the reaction's known stoichiometry. This is a stark contrast to the Process Hitting framework, in which observational data and qualitative knowledge was sufficient. The Chemical Master Equation describes the evolution of the probability of the system existing at any given state z by considering the propensities, a, of all reactions r_j which leave z and those which enter z,

$$\frac{\partial P(z, t|z_0, t_0)}{\partial t} \equiv \sum_j [a_j(z - v_j)P(z - v_j, t|z_0, t_0) - a_j(z)P(z, t|z_0, t_0)]$$

To simplify this representation, we may aggregate these terms to express the CME in matrix form, $\frac{\partial P}{\partial t} = \mathcal{A}P$ where $\mathcal{A}$ is known as the *connectivity matrix* of

the regulatory network. $\mathcal{A}$ is sparse, with nonzero elements $A_{i,j}$ where a reaction links states i and j:

$$A_{ij} = \begin{cases} -\sum a_k(z), & i = j \text{ (reactions leaving state i)} \\ \sum a_k(z), & i \neq j \text{ (reactions moving from state i to state j)} \end{cases}$$

While the CME is a natural and rich description of the physical system in a biological regulatory network, it demonstrates what has come to be known as the curse of dimensionality, growing exponentially with the number of species, N_{sp}: that is, if, for each species, we limit the range of possible values to N, the total state space of the system will be $(N_{sp})^N$. Since biological regulation can have many subtly interacting factors, this is clearly an impasse in the application of the CME to gene regulatory networks, which can become impossible even for simplified models. Although simulation techniques are common, the structure of the CME permits the application of state-of-the-art numerical methods. To this end, PGD has already been shown to effectively and flexibly solve the Chemical Master Equation.

2.4 Proper Generalized Decomposition

Proper Generalized Decomposition [2,3] is an emerging numerical tool in the field of mechanical engineering, though it has been applied to a wide range of problems, including the CME [1]. The foundation of this method is to assume that the target, in this case, the probability, can be written as a sum of a product of separable functions.

$$P(z,t) \cong \sum_{j=1}^{M} F_1^j(z_1) \cdot F_2^j(z_2) \cdot ... \cdot F_N^j(z_N) \cdot F_t(t)$$

This is not an entirely novel idea, but its recent applications have proven promising in dimension reduction problems. Although the accuracy increases with every addition, only a limited number, M, of functions are needed to capture the behavior of the system. Note that, with each function of size N (that is, the state space of each variable is limited to size N), with N_{sp} functions, the resulting dimensionality is the M sum of N_{sp} functions of size N, or $M(N \times N_{sp})$ in contrast to the original N_{sp}^N. The inclusion of a time as a separated function means that the solution is not incremental but complete for all time.

PGD is performed iteratively, searching for each product of separable functions which will minimize the residual of the running sum. These functions are colloquially called "modes", although there is no underlying notion that they represent the greatest source of variance, as is the case with Principal Component Analysis [10] (PCA). At each step, one is searching for a single one-dimensional function, in this case a $N \times 1$ vector, with the remainder of the state space known. These sets of one dimensional functions are found until their sum, the resulting approximation, meets some stopping criterion. Since all operations can

be performed by canonical techniques and are highly parallelizable, iterations are generally fast and computationally inexpensive. In addition, the form of PGD offers a natural way of handling unknown parameters by incorporating them as additional state space dimensions without changing the original algorithm. This, in particular, is a very desirable characteristic for application to biological regulatory networks since many parameters are often either unknown or come with some degree of uncertainty.

3 Translation of PH to CME

It is clear that Process Hitting contains certain desirable properties as a modeling framework for gene regulatory networks, such as the ability to construct a model from partial knowledge of the system, and that PGD is a fast, efficient way to numerically solve differential equations by breaking high dimensional systems into a search for one-dimensional modes. If we are able to move Process Hitting from its current format to that more like the CME, we have the potential to apply PGD and solve the system, not only for one desired configuration, but for the probability of existing at any configuration at any time given the initial state. An intuitive indication that such a link exists can be found in the simulation method used in Process Hitting: Gillespie's Next Reaction Algorithm.[13] This simulation technique allows for concurrent and competing processes and was developed by Gillespie as an *exact* simulation for the CME. That is, the Chemical Master Equation with any set of reactions can be simulated precisely by the Next Reaction Algorithm: if Process Hitting uses the Next Reaction Algorithm, there must exist some corresponding Chemical Master Equation that defines exactly the same behavior derived from the PH qualitative description of the system which does not require the addition of kinetic knowledge.

The key to translating the Process Hitting to its Chemical Master Equation structure is to re-imagine the kinetic reactions found in CME to be much more abstract, that is, not true physical reactions, but some collection of events which move the system from one state to another. These abstracted reactions cannot be interpreted in any greater detail than a PH action, though it is clear that, in reality, the system is being driven by a series of physical and chemical reactions. The hiding of kinetic processes was considered an advantage of PH, one which is conserved in translation to the CME. It is important that we concretize how one can derive these abstracted reactions, which we refer to as "faux-reactions", from any Process Hitting action. To illustrate this process, we will take a particular example

$$\text{p53McDam } 0 \rightarrow \text{Mn } 0 \nrightarrow 1 @ \left(4h^{-1}\right)$$
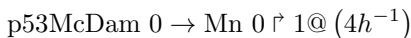
That is, that when p53 AND cytoplasmic mdm2 AND cellular damage are at their lowest levels, the amount of nuclear mdm2 increases at the propensity of 4 per hour. The combination of those factors activate the uptake of mdm2 to the nucleus, bringing the system from state set $\{0, 0, 0, 0\}$ to state set $\{0, 0, 0, 1\}$ where it will then be susceptible to other actions. It is easy to write this in a more readable form

$$0p53 + 0Dam + 0Mc + 0Mn \xrightarrow{4h^{-1}} 0p53 + 0Dam + 0Mc + 1Mn$$
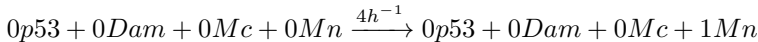
Granted, we have intentionally mimicked the structure of a stoichiometry equation, though it does not follow stoichiometric rules. Rather, this syntax merely states that there is some process (indicated by an arrow) which occurs at propensity $4h^{-1}$ that brings the system from state set $\{0, 0, 0, 0\}$ to state set $\{0, 0, 0, 1\}$. Any Process Hitting action can be written in these terms: for actions that do not dictate the full state set as our example has, undetermined variables give rise to multiple unique reactions with identical rates. So p53 0 → Dam 0 ↱ 1, for example, will result in four individual reactions wherein Mn and Mc take on values 0 and 1.

At this point, we must stretch our understanding of the Chemical Master Equation syntax to accommodate these new faux-reactions, though not by much. Recall that in the CME the system is described in terms of molecule counts, or whole numbers. In the same way that one discretizes the real number line in Generalized Logical Networks [12], so can we discretize the whole number line: $1p53$ does not signify one molecule of p53, but a quantity of p53 within some range. Thus, the faux-reaction equation effectively abstracts the mechanism by which the system moves as well as its actual quantitative contribution. The resulting set of equations can be but in the functional form of a discretized Chemical Master Equation which retains the syntax of its original counterpart, a sum of the effects of each reaction r_j which brings species to or away from a certain state, z at a given propensity $a(z)$.

4 Application and Results

We begin by constructing the Process Hitting for the p53 system as described in Section 1.1. This includes the creation of a cooperative sort p53McDam and all of the actions needed to update such that the state of the cooperative sort reflects instantaneously the state of the system. A full list of actions, as well as the PH diagram, can be found in the appendix. Reaction propensities were taken from [4,5]. At this early stage, we can do model-checking to find out if the fundamental structure of the graph supports certain desirable dynamics, exploiting the formal structure of the discrete model. Process Hitting has been implemented in a freely available software called PINT [9], which, among other things, can search for steady states, perform reachability analysis, and run simulation. In addition, PINT has the ability to import and export data from a number of other systems biology syntaxes, making it a flexible platform for newcomers to Process Hitting. From here, we can take the approach outlined in Section 3 to translate all actions into their corresponding faux-reactions.

In this paper, we begin by considering the most basic, intuitive representation of the Chemical Master Equation, that is, the case where the state space is represented by a one-dimensional vector of enumerate states, where the probability is a function of the state of the system and time, $P(Z, T)$. Here, Z is a vector of all possible states that can be occupied by the system; so, for example,

$\{Mc, Dam, p53, Mn\} = \{0,0,0,0\}$ is state 1, $\{1,0,0,0\}$ is state 2, and so on. This enumeration of states is acceptable for this particular problem but unsupportable as the number of species increases, making it impractical for most gene regulatory networks. However, by using this construction, we retain the potential to capture important emergent properties, to be addressed later on in the following subsection. In addition, the method by which we construct the discretized Chemical Master Equation is very straightforward in the case of a one dimensional state space. We can begin by considering the a matrix form of the CME as described in Section 2.3, $\frac{\partial P(z,t|z^0, t^0)}{\partial t} = \mathcal{A}P(z,t|z^0, t^0)$. As an example, the reaction

$$0p53 + 0Dam + 0Mc + 0Mn \xrightarrow{4h^{-1}} 0p53 + 0Dam + 0Mc + 1Mn$$
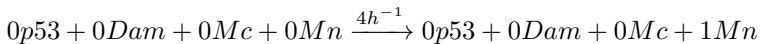
results in a nonzero element at $A_{\{0,0,0,0\},\{0,0,0,0\}}$ and $A_{\{0,0,0,1\},\{0,0,0,0\}}$, or, using arbitrary numbering for states, $A_{1,1}$ and $A_{7,1}$. The first element represents the system leaving state $1 \equiv \{0,0,0,0\}$ and the second the system moving to state $7 \equiv \{0,0,0,1\}$ from state $1 \equiv \{0,0,0,0\}$. Once the connectivity matrix is constructed, we can apply PGD to find the decomposition of $P(Z, t)$ and solution of the Chemical Master Equation for all time.

4.1 Results of One-Dimensional Problem

In Figure 3, we have taken a snapshot in time to compare results of Process Hitting and our discretized version of the Chemical Master Equation. We begin the system in state $3 \equiv \{0, 1, 0, 0\}$, which represents the presence of damage ($Dam = 1$) without any other element ($p53 = Mn = Mc = 0$). Since the system is small, we can solve the resulting CME using implicit finite differences as a gold standard to which we can compare the PGD results. PH simulations were executed using PINT, averaging over 1000 runs. Although this particular graph only relates to a single instance in time, the PGD solution obtained is for all time, as can be seen in Figure 4. We stopped the iterative algorithm after reaching a precision of 10^{-3}. Although there is clearly good agreement in this solution, even better approximations can be made by continuing the iterative scheme, though only a limited number of modes are needed to obtain the basic information of the system.

4.2 Increased Depth of Analysis

By connecting two very different modeling frameworks, we are able to exploit the advantages of each and, potentially, fill in gaps of analysis. From Process Hitting, we can quickly and efficiently analyze the global behaviors of the system, using tools such as steady state or reachability analysis. The application of formal methods allows a modeler to ask fundamental questions of the system. However, there exist blind spots in this modeling type like any other. For example, the p53-mdm2 system as given in the appendix has no focal steady states and, although we can use model-checking to obtain further information, a complete
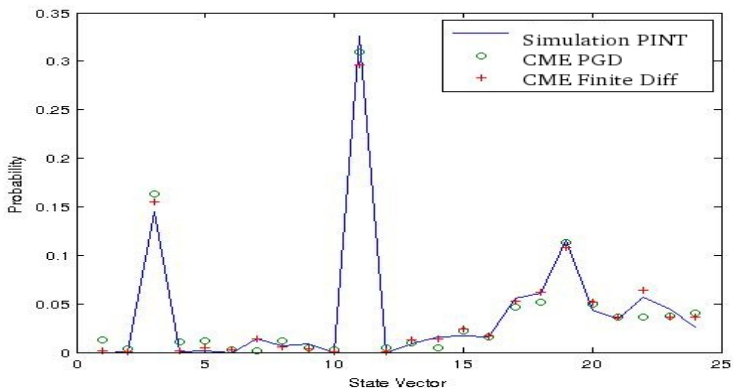
Fig. 3. Comparison of PH simulation with 1000 trials and translated CME, solved via Finite Differences and PGD

response requires simulation techniques. While manual investigation is possible in this small network, this ceases to be a viable strategy as the number of species increases. Moving to the probabilistic syntax of the Chemical Master Equation opens the potential to solve, rather than simulate, the complete system. In the p53-mdm2 network, we observe states whose probabilities do not change after a certain measure of time. This represents a basin of attraction, or a limit cycle, not yet capturable in PH analysis and predicted in [4,5]. In linear algebra terms, this is the null space of $\mathcal{A}$, all state vectors Z which satisfy $\mathcal{A}x = 0$. The PGD results for this particular problem can be seen on the left of Figure 4. Thus, even though the experimenter knows that the system does not settle into a particular steady state, he can know what states will be most or least prominent and the interactions (the faux-reactions) which connect them. Since the PGD can be implemented quickly, it facilitates experimentation with the model: one can easily test the results of adding or removing actions, working up to a model which demonstrates the correct dynamic behavior. Knowing that our system contains a limit cycle, we can induce a steady state by removing a single action in the Process Hitting. In this particular example, we have removed the action $Mn\_0 \to p53\_1 \restriction 2$. That is, the absence of mdm2 in the nucleus is insufficient to bring p53 to its maximum level. As we can see on the right hand side of Figure 4, we do arrive at the desired focal steady state, which can be confirmed using PH analysis.

4.3 Separability, Emergence and Multidimensional Problems

We can effectively solve a PH system in the context of its one-dimensional state vector form, $P(Z, T)$, but the true goal of this work is to consider the multi-dimensional problem in which the probability is a function of all species of the system, $P(Mc, Dam, p53, Mn, T)$. The size of neither the solution nor the prob-
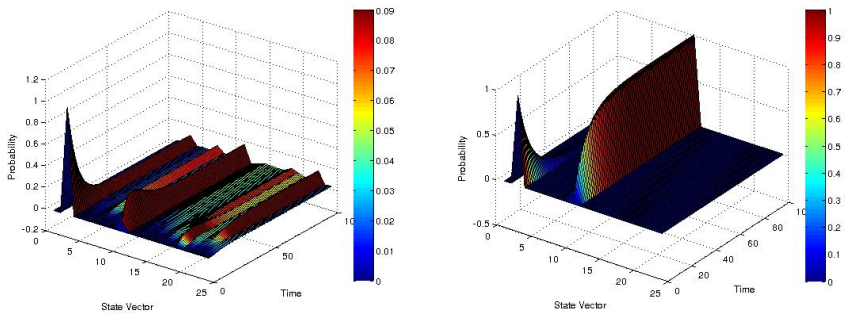
Fig. 4. The PGD solution of two systems. The original p53-mdm2 system as described in Section 2.1 (as seen on the left) contains a basin of attraction, or a limit cycle. The probabilities of the states in this cycle become constant as all trajectories from the initial condition enter the basin. By removing a single link in this cycle, we can induce a focus, or a steady state (as seen on the right). In this particular example, we have removed the action Mn_0 → p53_1 ↾ 2. That is, the absence of mdm2 in the nucleus is insufficient to bring p53 to its maximum level.

lem has changed, but the structure of the algorithm avoids the same problems of dimensionality. Where the connectivity matrix $\mathcal{A}$ was of size 24×24, it will now be $2 \times 2 \times 3 \times 2 \times 2 \times 2 \times 3 \times 2$, which contains the same number of elements. In this form, however, we do not directly enumerate the states and, at each step in the PGD algorithm, we search for a function of a very limited size corresponding to the number of processes in the sort, typically two or three. Without moving to a multi-dimensional structure, we cannot address larger and more interesting systems, however, moving into this new representation poses its own difficulties. A growing awareness of emergent properties in biological systems has developed from the fields of synthetic and systems biology. Though difficult to explain at a detailed level, the principle of emergence is, in effect, that the whole is somehow greater than the sum of its parts. The effect of activators and inhibitors are not additive but, rather, activators and inhibitors work synergistically to create amplified signals. For us, this means that we cannot assume that gene regulatory networks are separable, a feature we depend on for the PGD representation. However, that is not to say that it cannot be put into a *more or less* separated form: most likely, not all species demonstrate important emergent properties. The behavior of a system could very well be captured with the inclusion of a limited number of combined variables, perhaps $P(Dam, Mn, p53Mc, T)$, for example, in which $p53Mc$ captures the combined effects of p53 and cytoplasmic mdm2 while maintaining the individual influences of $p53$ and Mc. Developing a formal method for discovering these significant non-separable elements will be the topic of future work and brings the promise of opening up our methodology to the multidimensional formulation which, in turn, frees us from dimensionality.

5 Conclusion

The linking between very different modeling frameworks to arrive at a more complex, more powerful level of analysis promises to be an interesting vein of research in synthetic biology. In this paper, we have laid the foundation for just such a bridge between the discrete framework of Process Hitting and the Chemical Master Equation, in the hopes of overcoming the obstacle of dimensionality via the application of Proper Generalized Decomposition. This particular work considers the one dimensional "state vector" case, a form which conserves emergent– or non-separable– properties of the system. In order to broaden our scope to larger systems, we must be able to move to a multidimensional form. To do so, we must be able to re-introduce a limited number of non-separable, or combined, elements, which will be the topic of future works. The methodology introduced here has the potential to enrich analysis of gene regulatory networks and permit the study of larger, more complicated, and more realistic models, a necessity if we hope to make modeling frameworks fruitful in an applied setting.
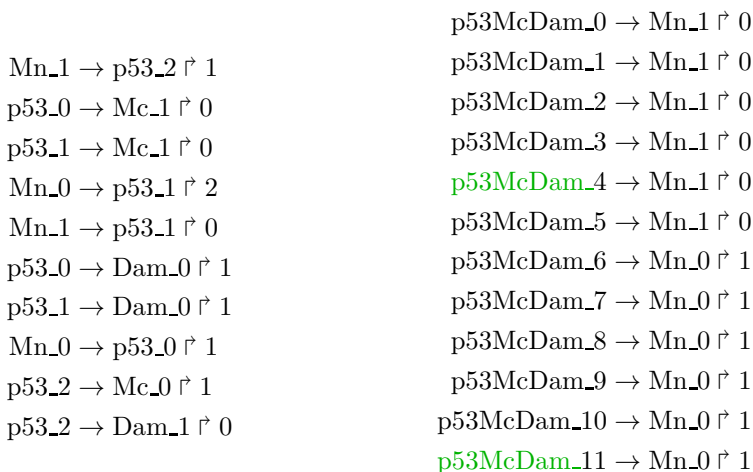
References

1. Ammar, A., Cueto, E., Chinesta, F.: Reduction of the chemical master equation for gene regulatory networks using proper generalized decompositions. International Journal for Numerical Methods in Biomedical Engineering 28(9), 960–973 (2012)
2. Chinesta, F., Ammar, A., Leygue, A., Keunings, R.: An overview of the proper generalized decomposition with applications in computational rheology. Journal of Non-Newtonian Fluid Mechanics 166(11), 578–592 (2011)
3. Chinesta, F., et al.: PGD-Based computational vademecum for efficient design, optimization and control. Archives of Computational Methods in Engineering 20(1), 31–59 (2013)
4. Abou-Jaoudé, W., Ouattara, D., Kaufman, M.: Frequency tuning in the p53-mdm2 network. I. Logical approach. Journal of Theoretical Biology 258, 561–577 (2009)
5. Abou-Jaoudé, W., Ouattara, D., Kaufman, M.: Frequency tuning in the p53-mdm2 network. II. Differential and stochastic approaches. Journal of Theoretical Biology 264(4), 1177–1189 (2010)
6. Paulevé, L., Magnin, M., Roux, O.: Refining dynamics of gene regulatory networks in a stochastic π-calculus framework. In: Priami, C., Back, R.-J., Petre, I., de Vink, E. (eds.) Transactions on Computational Systems Biology XIII. LNCS (LNBI), vol. 6575, pp. 171–191. Springer, Heidelberg (2011)
7. Folschette, M., Paulevé, L., Inoue, K., Magnin, M., Roux, O.: Concretizing the process hitting into biological regulatory networks. In: Gilbert, D., Heiner, M. (eds.) CMSB 2012. LNCS, vol. 7605, pp. 166–186. Springer, Heidelberg (2012)
8. Paulevé, L., Magnin, M., Roux, O.: Tuning Temporal Features within the Stochastic-Calculus. IEEE Transactions on Software Engineering 37(6), 858–871 (2011)
9. Paulevé, L., Magnin, M., Roux, O.: Pint-Process Hitting Related Tools (October 10, 2010), http://processhitting.wordpress.com (April 16, 2013)
10. Jolliffe, I.: Principal component analysis, vol. 487. Springer, New York (1986)
11. Leenders, G., Tuszynski, J.: Stochastic and deterministic models of cellular p53 regulation. Frontiers in Molecular and Cellular Oncology 3(64) (2013)

12. Bernot, G., et al.: Application of formal methods to biological regulatory networks: extending Thomas asynchronous logical approach with temporal logic. Journal of Theoretical Biology 229(3), 339–347 (2004)
13. Paulevé, L., Youssef, S., Lakin, M., Phillips, A.: A Generic Abstract Machine for Stochastic Process Calculi. In: Proceedings of the 8th International Conference on Computational Methods in Systems Biology, pp. 43–54 (2010)

A   The Process Hitting for p53-mdm2

In Section 2.2 we introduced the concept of sorts, processes and actions, as well as giving a brief introduction as to how actions are constructed based on a qualitative description of an activation or inhibition reaction. The actions constructed from the information outlined in Section 2.1 on the p53-mdm2 system are as follows:

$$Mn\_1 \to p53\_2 \nearrow 1$$
$$p53\_0 \to Mc\_1 \nearrow 0$$
$$p53\_1 \to Mc\_1 \nearrow 0$$
$$Mn\_0 \to p53\_1 \nearrow 2$$
$$Mn\_1 \to p53\_1 \nearrow 0$$
$$p53\_0 \to Dam\_0 \nearrow 1$$
$$p53\_1 \to Dam\_0 \nearrow 1$$
$$Mn\_0 \to p53\_0 \nearrow 1$$
$$p53\_2 \to Mc\_0 \nearrow 1$$
$$p53\_2 \to Dam\_1 \nearrow 0$$

$$p53McDam\_0 \to Mn\_1 \nearrow 0$$
$$p53McDam\_1 \to Mn\_1 \nearrow 0$$
$$p53McDam\_2 \to Mn\_1 \nearrow 0$$
$$p53McDam\_3 \to Mn\_1 \nearrow 0$$
$$p53McDam\_4 \to Mn\_1 \nearrow 0$$
$$p53McDam\_5 \to Mn\_1 \nearrow 0$$
$$p53McDam\_6 \to Mn\_0 \nearrow 1$$
$$p53McDam\_7 \to Mn\_0 \nearrow 1$$
$$p53McDam\_8 \to Mn\_0 \nearrow 1$$
$$p53McDam\_9 \to Mn\_0 \nearrow 1$$
$$p53McDam\_10 \to Mn\_0 \nearrow 1$$
$$p53McDam\_11 \to Mn\_0 \nearrow 1$$

Note that, for clarity and brevity, we have not listed those actions which update the cooperative sort p53McDam to reflect the current status of p53, Mc and Dam, nor have we included them in the graphical representation of the system shown in Figure 5. There are 48 such actions compared to the 22 listed here. As example, we have included two actions of Mc on the cooperative sort,

$$Mc\_1 \to p53McDam\_0 \nearrow 2$$

$$Mc\_1 \to p53McDam\_1 \nearrow 3$$

In both cases, p53McDam reflects the state space such that the sort Mc is at process Mc_0, which is no longer true if Mc_1 and must be updated. These examples are represented by blue, tightly dashed arrows in the figure, whereas the full updating action list is generally represented by the red, solid arrows linking the individual sorts p53, Mc and Dam, to their respective cooperative sort.
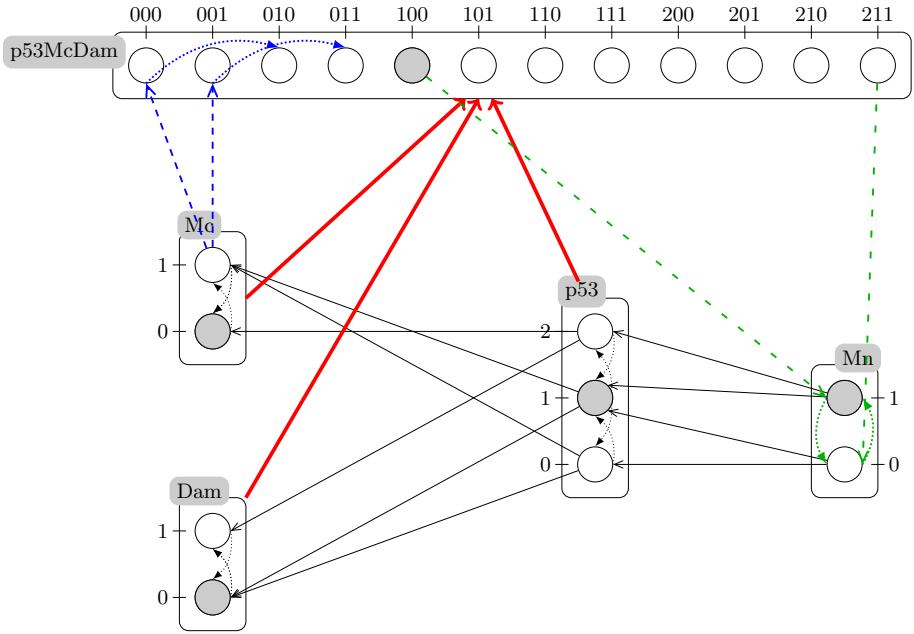
Fig. 5. A graphic representation of the Process Hitting. Each box represents a sort, which contains processes indicated by circles. Actions linking individual sorts are demonstrated by black arrows and can be found in the left hand column of the appendix chapter on PH. The p53-mdm2 system contains one cooperative sort, p53McDam. For the sake of clarity, not all actions linking the individual sorts and the cooperative sort have been shown; rather, red, solid arrows are drawn to indicate these updating actions, two examples of which are shown in tightly dashed blue lines. Likewise, not all actions linking p53McDam and Mn have been drawn, with two examples given in loosely dashed, green lines. As we have depicted it, our system is in state $\{0, 0, 1, 1\}$. There exists two actions which can be played, p53_1 → Dam_0 ↱ 1 and p53McDam_4 → Mn_1 ↱ 0. Process Hitting is asynchronous, thus, only one action can be played at any given time. Which action actually occurs depends on the rates, stochastic and temporal features assigned to each.