Approximating the solution of the chemical master equation by aggregation

M. Hegland¹

(Received 13 August 2008; revised 9 November 2008)

Abstract

The chemical master equation is a continuous time discrete space Markov model of chemical reactions. The chemical master equation is derived mathematically and it is shown that the corresponding initial value problem has a unique solution. Conditions are given under which this solution is a probability distribution. We present finite state and aggregation-disaggregation approximations and provide error bounds for the case of piecewise constant disaggregation. The aggregation-disaggregation approximation allows the solution of the chemical master equation for larger state spaces and is also an important tool for the solution of multidimensional problems.

Contents

1 The chemical master equation

C372

http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/1426 gives this article, © Austral. Mathematical Soc. 2008. Published November 20, 2008. ISSN 1446-8735. (Print two pages per sheet of paper.)

2 Approximation of probabilities over \mathbb{N}

C377

References C382

1 The chemical master equation

We consider a simple stochastic model for a closed system of s chemical species (molecules, chemical complexes, parts of molecules, etc) which may react according to a given set of reactions. The state of the system is described by the copy numbers of each species and is an integer vector $\mathbf{x} \in \mathbb{N}^s$. A chemical reaction is then modeled as a state transition

$$x \rightarrow x + z$$

where $z \in \mathbb{Z}^s$ is the *stoichiometric vector* of the reaction. We assume that the system is well stirred and in constant (random) motion and that as a consequence reactions occur randomly and the state of a system is a stochastic process, that is, a time dependent random variable. If one assumes that the propensity of any one reaction to take place only depends on the current state of the system one gets a Markovian model. Admittedly, this is a very simple model of chemical processes as they occur inside biological cells. Nonetheless, such models have been successfully used to gain insight into many cellular processes [1, 10, e.g.]. Moreover, a further justification for this model is that the widely used kinetic rate equations are obtained as a limiting case when the copy numbers tend to infinity. However, it has been shown that the stochastic model cannot be replaced by the kinetic rate equations—even if one is only interested in expectations of X(t)—and consequently the 'intrinsic noise' of chemical reactions is an important driver in molecular biology.

Currently, the main computational tool to study the behaviour of stochastic chemical reactions is the *stochastic simulation algorithm* (SSA) introduced

1

by Gillespie in 1977 [5]. This algorithm uses a random number generator to generate sample paths of the reaction. For complex systems with large numbers of reactions and systems with fast reactions the SSA was found to be too slow as the SSA updates the sample path for every single reaction taking place. An improvement is obtained with the tau-leap method which updates the sample path over larger time intervals [5] using an approximation which models the chemical reactions by independent Poisson processes. The time intervals are also determined as part of the procedure.

In many applications one is interested in statistical information about the stochastic process like the expectations, moments, correlations, quantiles and statistics of arrival times. Such information can be obtained from sample paths but one typically needs to average over a large number of them to get sufficient accuracy. One uses the sample paths $x^{(i)}(t)$ to get an empirical distribution function for the random variables X(t) and determines the statistics from the empirical distribution function which takes the form

$$p^{\text{emp}}(x,t) = \frac{1}{N} \sum_{k=1}^{N} \delta(x - x^{(k)}(t))$$

where $\delta(0) = 1$ and $\delta(x) = 0$ for $x \neq 0$.

The large computational effort required to generate large numbers of sample paths required motivates the search for alternative approximations of the probability distribution p(x,t) of X(t). Such approximations may be based on the chemical master equation which is an (infinite) system of ordinary differential equations describing the dynamics of p(x,t). In terms of probabilities, the chemical reaction i with a stoichiometric vector z_i increases the probability of state $x + z_i$ by some amount and decreases the probability of state x by the same amount. Chemical reactions thus define fluxes of probabilities j_i and one gets the following differential equations for a system with r reactions [4]:

$$\frac{\partial p(x,t)}{\partial t} = \sum_{i=1}^{r} (S_{z_i} - I) j_i(x,t). \tag{1}$$

Here I is the identity, S_{z_i} is the shift operator defined by

$$S_{z_i}p(x,t)=p(x+z_i,t),$$

and z_i is the stoichiometric vector of reaction i. The choice of the flux should be such that p(x,t) is a probability distribution with

$$\sum_{\mathbf{x} \in \mathbb{N}^s} p(\mathbf{x}, \mathbf{t}) = 1.$$

In order for this conservation of total probability to apply, no probability should flow out of the domain,

$$j_{i}(x,t) = 0 \quad \text{if } x + z_{i} \notin \mathbb{N}^{s}. \tag{2}$$

By definition one also has $j_i(x,t) \geq 0$ for all x and t. The chemical interpretation of this condition is that reactions will not occur unless sufficient amounts of the substrates are available. The second defining property of probabilities is that they cannot be less than zero. For this property to be maintained there can be no fluxes away from points x with p(x,t)=0 as such fluxes would lead to negative probabilities. It follows then that there must exist finite functions $\lambda_i(x,t)$ (which in principle could depend on p) such that

$$j_i(x,t) = \lambda_i(x,t)p(x,t) \tag{3}$$

and by the previous condition one has

$$\lambda_i(x,t) = 0$$
 if $x + z_i \notin \mathbb{N}^s$.

In summary, one has the following simple result

Lemma 1 Let p(x,t) be such that p(x,0) is a probability distribution and furthermore satisfies

$$\frac{\partial p(x,t)}{\partial t} = \sum_{i=1}^{r} (S_{z_i} - I) \lambda_i(x,t) p(x,t)$$

where $\lambda_i \geq 0$ satisfy

$$\lambda_{i}(x,t) = 0$$
 if $x + z_{i} \notin \mathbb{N}^{s}$,

then p(x, t) is a probability distribution.

The system of differential equations satisfied by p(x,t) are called the *chemical master equations* [11]. One introduces the multiplication operators Λ_i by

$$\Lambda_i p(x,t) = \lambda_i(x,t) p(x,t),$$

the linear operator A by

$$A = \sum_{i=1}^{r} (S_{z_i} - I) \Lambda_i \tag{4}$$

and writes the chemical master equations (CME) as

$$\frac{\partial p}{\partial t} = Ap. \tag{5}$$

In this general form the Hille–Yosida theorem guarantees existence and uniqueness of the solution of the CME:

Theorem 2 Let $A = \sum_{i=1}^{r} (S_i - I) \Lambda_i$ where S_i are shift operators and where $\Lambda_i \geq 0$ are densely defined diagonal operators on $\ell_1(\mathbb{N}^d)$. Then the initial value problem

$$\frac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{t}} = \mathrm{A}\mathfrak{p}$$

with $p(0) = p_0 \in \ell_1(\mathbb{N}^d)$ has a unique continuous solution $p(t) \in \ell_1(\mathbb{N}^d)$.

Proof: For any $\lambda > 0$ and $f \in \ell_1(\mathbb{N}^d)$ in the domain of A one has

$$\begin{split} \|(\lambda I - A)f\| &= \|\lambda f + \sum_{i=1}^r \Lambda_i f - \sum_{i=1}^t S_i \Lambda_i f\| \\ &\geq \|\lambda f + \sum_{i=1}^r \Lambda_i f\| - \sum_{i=1}^r \|\Lambda_i f\| \\ &= \sum_{x \in \mathbb{N}^d} \left[\lambda + \sum_{i=1}^r \lambda_i (x)\right] |f(x)| - \sum_{x \in \mathbb{N}^d} \sum_{i=1}^r \lambda_i (x) |f(x)| \\ &= \lambda \|f\| \end{split}$$

by the definition of the ℓ_1 norm and as $\|S_i\| \leq 1$. As A is densely defined in $\ell_1(\mathbb{N}^d)$ it follows that the residual $(\lambda I - A)^{-1}$ is a bounded linear operator on $\ell_1(\mathbb{N}^d)$ and that

$$\|(\lambda I - A)^{-1}\| \le \frac{1}{\lambda}.$$

By the Hille–Yosida theorem, A is the infinitesimal generator of a strongly continuous semigroup T_t and consequently, $p(t) = T_t p_0$ is the unique continuous solution of the CME.

Some simple relevant biochemical examples with one reaction include the Poisson process with $\lambda(x) = \alpha x$, the decay with $\lambda(x) = \alpha(\alpha - x)_+$, the (hetero and homo) dimerization with $\lambda(x) = \alpha(\alpha - x)_+(b - x)_+$ and the reduced Michaelis–Menten process with $\lambda(x) = \alpha(\alpha - x)_+/(b + \alpha - x)_+$ where α , α , α are all larger than zero and where $(\alpha - x)_+ = \max(\alpha - x, 0)$ etc. Except for the first case, all the cases result in an effectively finite state space. However, this is usually not the case of reversible reactions and more than one species. Systems with multiple reactions are obtained as combinations of these simple reactions. Interactions between species are obtained as the rates λ_i related to the production or decay of one species may depend on the levels of other species and, in addition, multiple species may be produced or consumed in the same reaction.

Currently there seem to be three major computational approaches to solve the CME which were all developed at about the same time. The discrete Galerkin [3, 2] method uses a the Rothe approach by first discretizing time and then solving the resulting system of equations approximately. This approximation is based on a space partitioning and uses a local spectral basis in each partition. The finite state projection (FSP) method [8] is based on a truncation of the state space combined with an interpolation using piecewise polynomials on the space partitions. The aggregation approach [6] uses a finite volume method and local polynomial approximation for state space approximation and the expokit method [9] for the solution of the resulting ODE. This method has been extended by an adaptive sparse grid approximation in order to solve large systems. It was shown to be able to solve the CME with up to 100 species [7].

While all the methods are based on local polynomial approximations it was so far unclear why such polynomial approximations should perform well for approximating the probability distribution which does not have the usual smoothness properties required for such approximations as it is an integer function. In the next section we answer this question by providing approximation error bounds for a class of functions which contains the probability distributions of interest.

2 Approximation of probabilities over \mathbb{N}

The main tool for the analysis will be the characteristic function or Fourier transform of the probability distributions defined as

$$\hat{p}(\omega,t) = \sum_{x \in \mathbb{N}} p(x,t)e^{ix\omega}.$$

As $p \in \ell_1$ (as a function of x) the characteristic function is a continuous 2π -periodic function of ω . In case of the Poisson process one has rates

 $\lambda(x) = \text{constant and}$

$$p(x,t) = \frac{(\lambda t)^x}{x!} \exp(-\lambda t), \quad x \in \mathbb{N},$$

and

$$\label{eq:problem} \boldsymbol{\hat{p}}(\boldsymbol{\omega},t) = \exp\left[\lambda t (e^{i\boldsymbol{\omega}} - 1)\right].$$

The special form of $\mathfrak p$ for this and similar cases motivated Engblom [3] and Deuflhard [2] to consider weighted ℓ_2 norms with weights of the form

$$w(x) = \frac{x!}{\alpha^x}, \quad x \in \mathbb{N},$$

for some $\alpha>0$. Similarly, we suggest to consider a weighted supremum norm with slightly weaker weights but include a weighted norm on the Fourier space. Specifically, let $\alpha>0$, $\beta>0$ and $\kappa\geq0$ be constants and define

$$\|\alpha\|_{\alpha,\beta,\kappa} = \sup_{x\in\mathbb{N}} e^{\alpha|x-\kappa|} |\alpha(x)| + \sup_{\omega\in[-\pi,\pi]} e^{\beta|\omega|} |\widehat{\alpha}(\omega)|.$$

This defines a norm on the set of finite length sequences $\mathfrak a$ (sequences where only a finite number of terms are nonzero) and the topological closure is then a Banach space defined as

$$M_{\alpha,\beta,\kappa} = \{\alpha \mid \|\alpha\|_{\alpha,\beta,\kappa} < \infty\}.$$

It follows that for every $a \in M_{\alpha,\beta,\kappa}$ one has

$$|a(x)| \le e^{-\alpha|x-j\kappa|} ||a||_{\alpha,\beta,\kappa}, \quad k \in \mathbb{Z},$$

and

$$|\hat{\mathbf{a}}(\omega)| \le e^{-\beta|\omega|} \|\mathbf{a}\|_{\alpha,\beta,\kappa}, \quad \omega \in [-\pi,\pi].$$

This means that for large enough β and α the $p \in M_{\alpha,\beta}$ admit sparse approximations both in the original and in the Fourier domain. For example, one has for the number of terms larger than some $\epsilon \in (0, \|p\|_{\alpha,\beta,\kappa}]$:

$$\#\{p(x)>\varepsilon\}\leq \frac{2}{\alpha}\log\frac{\|p\|_{\alpha,\beta}}{\varepsilon}.$$

(For simplicity we treat time t as an implicit parameter and write p(x) instead of p(x,t).) In passing we note that a consequence of the bound on $\hat{p}(\omega)$ and because we have $\hat{p}(0) = 1$ (as p is a probability) one has

$$\|\mathbf{p}\|_{\alpha,\beta,\kappa} \geq 1.$$

We consider a combination of two approximations. First we truncate the state space and set the approximation $\mathfrak{p}^{(\mathfrak{m},\mathfrak{n})}$ to be defined by

$$p^{(m,n)}(x) := \begin{cases} p(x), & x = m, \dots, m+n-1 \\ 0 & \text{otherwise.} \end{cases}$$

This approximation has n consecutive (potentially) nonzero terms. The best possible approximation with n consecutive nonzero terms $T_n(p) = p^{(m_{\rm best},n)}$ then satisfies

$$\|T_n(\mathfrak{p})-\mathfrak{p}\|_1\leq \|\mathfrak{p}^{(\mathfrak{m},\mathfrak{n})}-\mathfrak{p}\|_1,\quad \mathfrak{m}\in\mathbb{N}\,.$$

This approximation is a nonlinear approximation. One has the error bound

Proposition 3 For $\mathfrak{p} \in M_{\alpha,\beta,\kappa}$ the error of the 'best consecutive \mathfrak{n} -term approximation' $T_{\mathfrak{n}}$ is bounded by

$$\|\mathsf{T}_{\mathsf{n}}(\mathfrak{p}) - \mathfrak{p}\|_1 \leq \mathsf{C}e^{-\alpha \mathsf{n}/2}$$

for
$$C = 2\|p\|_{\alpha,\beta,\kappa}/(1-e^{-\alpha})$$
.

Note that approximation procedure does not require the knowledge of any of the parameters α , β nor κ .

Proof: We show that there exists an approximation $p^{(m,n)}$ with this bound. First observe that for the n-term approximation starting at m one has

$$\|p^{(m,n)} - p\|_1 = \sum_{x=0}^{m-1} p(x) + \sum_{x=m+n}^{\infty} p(x),$$

 $\mathrm{and}\ \mathrm{so}, \mathrm{as}\ p(x) \leq \|p\|_{\alpha,\beta,\kappa} e^{-\alpha|k-\kappa|} \ \mathrm{one}\ \mathrm{has}, \mathrm{for}\ m\ \mathrm{satisfying}\ \kappa-n+1 \leq m \leq \kappa\,,$

$$\begin{split} \|p^{(m,n)} - p\|_1 &\leq \left(\sum_{\kappa=0}^{m-1} e^{-\alpha(\kappa-k)} + \sum_{\kappa=m+n}^{\infty} e^{-\alpha(k-\kappa)}\right) \|p\|_{\alpha,\beta,\kappa} \\ &\leq \frac{e^{-\alpha(\kappa-m+1)} + e^{-\alpha(n+m-\kappa)}}{1 - e^{-\alpha}}. \end{split}$$

For even n choose $m=\kappa-n/2$ and for odd n choose $m=\kappa-(n-1)/2$ to get

$$\|\mathfrak{p}^{(\mathfrak{m},\mathfrak{n})} - \mathfrak{p}\|_1 \leq \frac{2e^{-\alpha\mathfrak{n}/2}}{1 - e^{-\alpha}} \|\mathfrak{p}\|_{\alpha,\beta,\kappa}.$$

Thus there exists an \mathfrak{m} such that the bound holds and by the definition of T_n the claim follows.

This bound is now combined with a theorem by Khammash and Munsky [8] which provides a relation between $T_n(\mathfrak{p})$ and the numerical solution of the CME using the Finite State Projection method to obtain an error bound for the FSP method.

Consider now the aggregation based approximation. This consists of two steps, an aggregation step and a disaggregation step. An alternative interpretation uses three steps, an initial smoothing step, a sampling step and a final smoothing step. While the interpretations are quite different, the results of the two procedures are identical. Furthermore, the second interpretation provides access to the tools used in mathematical signal processing. The first smoothing step is given by the mapping $\mathfrak{p} \to \mathfrak{a} * \mathfrak{p}$ where

$$(\tilde{\mathfrak{a}}*\mathfrak{p})(\mathfrak{y}) = \sum_{\mathfrak{x}\in\mathbb{N}} \mathfrak{a}_{\mathfrak{x}-\mathfrak{y}}*\mathfrak{p}_{\mathfrak{x}}, \quad \mathfrak{y}\in\mathbb{Z}.$$

The 'filter' a(x) is defined for $x \in \mathbb{Z}$. For notational simplicity and to be consistent with the reconstruction phase we use the reverse filter \tilde{a} defined by

$$\tilde{a}(x) = a_{-x}$$

and using the convolution theorem one gets the Fourier transform of the smoothed density $(\tilde{\mathfrak{a}}\star \mathfrak{p})(\omega)=\hat{\mathfrak{a}}(\omega)^*\hat{\mathfrak{p}}(\omega)$ where $\hat{\mathfrak{a}}(\omega)^*$ denotes the conjugate complex of $\hat{\mathfrak{a}}$. In the special case of aggregation based smoothing one simply takes the average of neighboring values, more specifically,

$$(\tilde{a} \star p)(x) = \frac{1}{h} \sum_{y=x}^{x+h-1} p(y)$$

and

$$\alpha(x) = \begin{cases} 1/h, & x = 0, \dots, h-1, \\ 0, & \text{otherwise.} \end{cases}$$

The Fourier transform is then

$$\widehat{\mathfrak{a}}(\omega) = e^{2\pi i (h+1)/2} \frac{\sin(\omega h/2)}{h \sin(\omega/2)}.$$

In a second step the convoluted signal $\tilde{\alpha}\star p$ is sampled with an operator S defined by

$$S\left(\tilde{\alpha}\star p\right)(x) = \begin{cases} (\tilde{\alpha}\star p)(x) & \text{for } x=jh \text{ with } j\in\mathbb{Z} \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

The sampled sequence provides the computational savings of course as it requires h times less space to store than the original distribution. An approximation of the original distribution is then recovered with a second convolution with some sequence b so that one gets the approximation

$$p^{\mathrm{appr}} = b \star S(\tilde{a} \star p).$$

The sampling S introduces aliasing in the Fourier domain and one has for the Fourier transform (or characteristic function)

$$\widehat{\mathfrak{p}^{\mathrm{appr}}}(\omega) = \widehat{\mathfrak{b}}(\omega) \sum_{k=0}^{h-1} \widehat{\mathfrak{a}}(\omega + 2\pi k/h)^* \widehat{\mathfrak{p}}(\omega + 2\pi k/h).$$

References C382

In the simplest case of piecewise constant approximation one has $b=\alpha$ and one gets the following bound for the error.

Proposition 4 Let $\mathfrak{p} \in \mathsf{M}_{\alpha,\beta,\kappa}$ and \mathfrak{h}/β sufficiently small. Then there exists a constant C such that

$$\|\mathfrak{p}^{\mathit{appr}} - \mathfrak{p}\|_{\ell_2} \le C(h/\beta)^{3/2} \|\mathfrak{p}\|_{\alpha,\beta,\kappa}.$$

Proof: Using a triangle inequality and $|\hat{p}(\omega)| \leq e^{-\beta|\omega|} ||p||_{\alpha,\beta,\kappa}$ one gets

$$\begin{split} \|p^{\mathrm{appr}} - p\|_{\ell_2} &= \|\widehat{p^{\mathrm{appr}}}(\omega) - \widehat{p}(\omega)\| \\ &\leq \|p\|_{\alpha,\beta,\kappa} \Bigg(\sqrt{\int_{-\pi}^{\pi} (|\widehat{\alpha}(\omega)|^2 - 1)^2 e^{-2\beta|\omega|} \, d\omega} \\ &+ \sum_{k=1}^{h-1} \sqrt{\int_{-\pi}^{\pi} |\widehat{\alpha}(\omega) \widehat{\alpha}(\omega + 2k\pi/h)|^2 \, e^{-2\beta|\omega|} \, d\omega} \Bigg) \end{split}$$

The claimed result can then be obtained using standard bounds for $\hat{\mathfrak{a}}(\omega)$.

Note that h/β is the length scale of the approximation relative to the length scale β defined by the Fourier transform. A similar proof is used to obtain error bounds for approximations with other piecewise polynomials b(x). The application of these approximation error bounds proceeds in the same way as in the finite state space method using a combination with some stability estimate. We plan to present this explicitly in a forthcoming paper.

References

[1] Adam Arkin, John Ross, and Harley H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage

References C383

lambda-infected Escherichia coli cells. *Genetics*, 149(4):1633–1648, 1998.

http://www.genetics.org/cgi/content/abstract/149/4/1633 C372

- [2] P. Deuflhard, W. Huisinga, T. Jahnke, and M. Wulkow. Adaptive discrete Galerkin methods applied to the chemical master equation. Technical Report ZIB-Report 07-04, Konrad–Zuse Zentrum fuer Informationstechnik Berlin, 2007.
 - http://www.zib.de/Publications/Reports/ZR-07-04.pdf C377, C378
- [3] Stefan Engblom. Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comput.*, 180(2):498–515, 2006. doi:10.1016/j.amc.2005.12.032 C377, C378
- [4] Lars Ferm and Per Lötstedt. Numerical method for coupling the macro and meso scales in stochastic chemical kinetics. *BIT Numerical Mathematics*, 47:735–762, 2007. doi:10.1007/s10543-007-0150-z C373
- [5] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, December 1977. C373
- [6] Markus Hegland, Conrad Burden, Lucia Santoso, Shev MacNamara, and Hilary Booth. A solver for the stochastic master equation applied to gene regulatory networks. J. Comput. Appl. Math., 205(2):708–724, 2007. doi:10.1016/j.cam.2006.02.053 C377
- [7] Leo Huber and Markus Hegland. Dimension adaptive sparse grids for the chemical master equations. submitted, 2008. C377
- [8] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, 124:044104–1–044104–1, 2006. doi:10.1063/1.2145882 C377, C380

References C384

[9] Roger B. Sidje. EXPOKIT: Software package for computing matrix exponentials. *ACM Transactions on Mathematical Software*, 24(1):130–156, March 1998. doi:10.1145/285861.285868 C377

- [10] T. Tian and K. Burrage. Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage-λ. Journal of Theoretical Biology, 227:229–237, 2004. doi:10.1016/j.jtbi.2003.11.003 C372
- [11] N. G. van Kampen. Stochastic Processes in Physics and Chemistry. North Holland, Amsterdam, the Netherlands, 1981. C375

Author address

1. M. Hegland, Mathematical Sciences Institute, Australian National University, Canberra, Australia.

mailto:markus.hegland@anu.edu.au