On the numerical solution of the chemical master equation with sums of rank one tensors

Markus Hegland[1] Jochen Garcke[2]

August 4, 2011

Abstract

We show that sums of rank one tensors (or separable functions) representing the so-called Candecomp/Parafac or CP-*decomposition* is used effectively to solve the chemical master equations as in many cases the effective tensor rank (CP rank) of the probability distribution only grows slowly with time. Both theoretical bounds and computational experiments are presented which support this claim.

The proposed numerical algorithm is thought to provide an effective tool for the computational study of stochastic biochemical systems involving large numbers of different chemical species.

Contents

1 Introduction

A well-stirred chemical system is characterised by a vector $X(t)$ of copy numbers of its relevant chemical species. According to Kurtz and collabora-

tors [2], such a system is modelled as a stochastic process $X(t)$ satisfying

$$X(t) = X(0) + \sum_{i=1}^{k} Y_i \left(\int_0^t \lambda_i(X(s))\, ds \right) z_i. \tag{1}$$

Here $X(0)$ is the initial condition, $Y_i(\nu)$ are independent standard Poisson processes with parameter ν, $\lambda_i(x) \in \mathbb{R}$ is the propensity and $z_i \in \mathbb{Z}^d$ the stoichiometric vector of reaction i for $i = 1, \ldots, k$. One can use this equation to generate simulation algorithms like the Gillespie method [3] and even the kinetic rate equations as by the application of the law of large numbers one gets $Y_i(\nu) \to \nu$ for $\nu \to \infty$. In Section 2 we use a related consequence of the law of large numbers.

Here we consider the conditional distribution $p(t;x) = P(X(t) = x \mid X(0))$ of the stochastic process $X(t)$ defined by equation (1). If $x \in \mathbb{Z}^d$ is interpreted as an index then $p(t;x)$ is the xth entry of a tensor $p(t)$. Such a conditional distribution is used for example to compute the likelihood of data $X(t_1), \ldots, X(t_N)$ in a maximum likelihood method for the determination of parameters of the propensities λ_i, see for example the paper by Tian et al. [8].

VanKampen [9] has shown that $p(t)$ satisfies the *chemical master equations* (CME)

$$\frac{dp(t)}{dt} = A\, p(t) \tag{2}$$

where

$$A\, p = \sum_{i=1}^{k} (S_{z_i} - I)\, \lambda_i \circledast p \tag{3}$$

and where $S_z u(x) = u(x+z)$ denotes the shift along z and $\lambda_i \circledast p$ is the Hadamard or point-wise product.

Typically, $p(t)$ is a d-dimensional tensor with M^d entries where M is the size of state space in one dimension. This exponential dependency of the complexity on the dimension is the curse of dimension. There are tensors, however, which can be represented in a sparse format and therefore do not suffer under this curse [6]. In particular, if

$$p(t;x) = \prod_{i=1}^{d} p_i(t;x_i)$$

for some p_i where x_i is the i-th component of x one only requires the storage of dM numbers, namely M for each p_i. This is a rank-one tensor. More generally, a tensor of rank r is of the form

$$p(t;x) = \sum_{j=1}^{r} s_j \prod_{i=1}^{d} p_{j,i}(t;x_i). \tag{4}$$

The storage requirement here is rdM and there is no curse for low r. In the following we provide some evidence that such low rank r approximations of the solution of the master equation are effective.

A related approach to the one discussed here has been suggested by Jahnke and Huisinga [5]. Their approach is based on the Dirac–Frenkel–McLachlan variational principle while the algorithm suggested here is based on more general approximation with low rank tensors. We also provide some explanation why the chemical master equation admits low-rank approximations. Alternatives to tensor-product approximations have been suggested by Munsky and Khammash [7] using the finite-state projection method and by Hegland et al. [4] using (dimension-) adaptive sparse grids. While especially the adaptive sparse grids have been shown to work well even for very high dimensions, the theory of these earlier approaches for high dimensions is less clear. As the chemical master equations have a natural low rank structure we think that the tensor-product approaches are more promising for some very high dimensional problems, and in particular for modelling of reaction cascades.

In Section 2 we discuss when and why the solution of the chemical master equation can be approximated by low rank tensors. In particular, the law of large numbers is used to demonstrate the accuracy of (local) rank one approximations. In Section 3 we introduce the new algorithm which is based on a combination of a standard initial value problem solver and low rank compression. To show the potential of the approach we provide in Section 4 some computational results and show that the new method works well at least for two dimensional problems where it appears to have a spectral-like complexity.

2 Low rank approximations of $p(t)$

When solving the CME for the conditional probability $p(t;x) = p(X(t) = x \mid X(0) = x_0)$ one often observes that there exist ϵ-approximations for which the rank grows initially (note that $p(x,0)$ has rank one), then has some phases where the rank is relatively stable and others where the rank decreases. We show under which conditions these situations arise. For this we consider ϵ approximations which are obtained by approximating the propensities λ_i of the CME. The effect of such a modification on the solution is bounded by the following lemma.

Lemma 1 *Let $p(t)$ and $\tilde{p}(t)$ solve the chemical master equations*

$$\frac{dp}{dt} = \sum_{i=1}^{k} (S_{z_i} - I)\, \lambda_i \circledast p, \tag{5}$$

$$\frac{d\tilde{p}}{dt} = \sum_{i=1}^{k} (S_{z_i} - I)\, \tilde{\lambda}_i \circledast \tilde{p}, \tag{6}$$

for $p(0) = \tilde{p}(0)$. If $X(t)$ is the stochastic process with distribution $p(t)$ then

$$\|p(t) - \tilde{p}(t)\|_1 \leq 2 \sum_{i=1}^{k} \int_0^t E(|\lambda_i(X(s)) - \tilde{\lambda}_i(X(s))|)\, ds.$$

Proof: Let A and $\tilde{A}$ be the operators of the equations (5) and (6), respectively such that

$$A\, u = \sum_{i=1}^{k} (S_{z_i} - I)\, \lambda_i \circledast u,$$

$$\tilde{A}\, u = \sum_{i=1}^{k} (S_{z_i} - I)\, \tilde{\lambda}_i \circledast u,$$

for $u \in \ell_1(\mathbb{Z}^d)$. Let in the following $\|u\|_1 = \sum_{x \in \mathbb{Z}^d} |u(x)|$ be the $\ell_1(\mathbb{Z}^d)$ norm. A standard variation of the constant argument gives

$$p(t) - \tilde{p}(t) = \int_0^t e^{\tilde{A}(t-s)} (A - \tilde{A})\, p(s)\, ds.$$

Using the triangle inequality and the ℓ_1-contractivity of $e^{\tilde{A}t}$ one gets the bound

$$\|p(t) - \tilde{p}(t)\|_1 \leq \int_0^t \|(A - \tilde{A})p(s)\|_1\, ds.$$

Another application of the triangle inequality combined with the contractivity of the shifts S_{z_i} is then used to show

$$\|p(t) - \tilde{p}(t)\|_1 \leq 2 \sum_{i=1}^{k} \int_0^t \|\lambda_i \circledast p(s) - \tilde{\lambda}_i \circledast p(s)\|_1\, ds.$$

The bound follows from the definitions of the ℓ_1 norm and expectation. ♠

If A has operator rank one, that is $A = A_1 \otimes \cdots \otimes A_d$, then the rank of the solution $p(t)$ of the corresponding CME does not change over time. The

previous lemma then provides a condition under which a constant rank approximation of $p(t)$ exists — when A applied to $p(t)$ is well approximated by a rank one operator.

If the propensities λ_i have tensor rank one, then the rank of $\lambda_i \circledast p$ is the same as the rank of p. As in this case the rank of $S_{z_i}\lambda_i$ is also one, the rank of Ap (which has $2k$ terms) is $2k$ times the rank of p. For sufficiently small time steps Δt one approximates $p(t+\Delta t)$ using the Euler method as

$$p(t+\Delta t) = p(t) + \Delta t\, A\, p(t).$$

The rank of $p(t+\Delta t)$ is then at most $2k+1$ times the rank of $p(t)$. Consequently we expect to find that ϵ approximations of $p(t)$ have a tensor rank which grows over time. While the discussion suggests exponential growth in time the growth is slower in practice.

In order to understand what may happen in practice, we consider a simple example with two reactions $\mathcal{S}_1 \to \mathcal{S}_1'$ and $\mathcal{S}_2 \to \mathcal{S}_2'$. The state is defined by the counts of $\mathcal{S}_1$ and $\mathcal{S}_2$. Assume that the propensities are $\lambda_1(x) = \kappa_1 x_1$ and $\lambda_2(x) = \kappa_2 x_1 x_2/n_1$ where n_1 is the initial count of $\mathcal{S}_1$. Choose $\tilde{\lambda}_1 = \lambda_1$ and $\tilde{\lambda}_2(x) = \lambda_2(E(X_1(t)), x_2) = \kappa_2 e^{-\kappa_1 t} x_2$. The identity $E(X_1(t)) = e^{-\kappa_1 t} n_1$ is obtained from the explicit determination of the marginal distribution of $X_1(t)$. This distribution satisfies a master equation with propensity $\lambda_1(x_1) = \kappa_1 x_1$ and is thus binomial. The approximating system has two statistically independent reactions where the second propensity depends on time. The approximating tensor $\tilde{p}(t)$ thus has rank one.

Proposition 2 *Let $p(t)$ and $\tilde{p}(t)$ be the probability distributions for the two systems defined in the previous paragraph. Then the error is bounded by*

$$\|p(t) - \tilde{p}(t)\|_1 \leq \sqrt{\frac{2}{n_1\kappa_1}}\kappa_2(1-e^{-\kappa_1 t})\sqrt{\int_0^t E(X_2(s)^2)\, ds}$$

Proof: An application of Lemma 1 gives

$$\|p(t) - \tilde{p}(t)\|_1 \leq 2\int_0^t E(|\lambda_2(X(s)) - \tilde{\lambda}_2(X(s))|)\, ds.$$

Inserting the values of the propensities one gets by Cauchy–Schwarz

$$\begin{aligned} E(|\lambda_2(X(s)) - \tilde{\lambda}_2(X(s))|) &\leq \frac{\kappa_2}{n_1}\sqrt{\mathrm{var}(X_1(s))}\,\sqrt{E(X_2(s)^2)} \\ &= \kappa_2\, n_1^{-1/2}\sqrt{e^{-\kappa_1 s}(1-e^{-\kappa_1 s})}\,\sqrt{E(X_2(s)^2)}. \end{aligned}$$

Integration over t and another application of Cauchy–Schwarz gives

$$\int_0^t \sqrt{e^{-\kappa_1 s}(1-e^{-\kappa_1 s})}\,\sqrt{E(X_2(s)^2)}\,ds \\ \leq \sqrt{\int_0^t e^{-\kappa_1 s}(1-e^{-\kappa_1 s})\,ds}\,\sqrt{\int_0^t E(X_2(s)^2)\,ds},$$

from which the result follows. ♠

One can see that the error of the approximation $\tilde{\lambda}_2$ first increases and then decreases with time. Interestingly, the error also decreases with the initial count n_1 of $\mathcal{S}_1$, this is a consequence of the law of large numbers.

When the error grows to a certain threshold, one partitions the domain into two parts and does the rank one approximation for each part separately. (An alternative to using domain partitioning would be to use a partition of unity.) This is how one gets higher rank approximations and one can use the proposition to get error bounds for all the components and thus for $p(t)$.

To further illustrate the consequence of the law of large numbers consider the original system from above. This has been solved for different parameters n_1 using the CMEPY software mentioned in Section 4 to get the exact solution. An approximation was obtained using the singular value decomposition (SVD) and truncating all components with singular values $\sigma_i < 10^{-5}$. One can see that the ranks of the resulting approximation decrease with increasing n_1. Thus the probability of larger systems are approximated by tensors with smaller ranks, see Figure 1.

3 The algorithm

The numerical solution of the initial value problem (2) uses a discretisation of time $t_0 = 0 < t_1 < \cdots < t_K$. For example, one may choose $t_k = k\tau$ but more generally, the grid widths $\tau_k = t_k - t_{k-1}$ are not all equal to some value τ. Denote a numerical approximation to $p(t_k)$ by $p_k \in \mathbb{R}^\Omega$ where Ω contains the support of $p(t_k)$. As the differential equation associated with the initial value problem (2) is stiff, one needs to use stable solvers. To illustrate the suggested approach we use very simple solvers.

Furthermore, one needs a tensor class which has two operations: addition of tensors and (point-wise, Hadamard) multiplication. The sum of a rank r_1 and a rank r_2 tensor is (at most) of rank $r = r_1 + r_2$, whereas the Hadamard product is a tensor of rank $r_1 r_2$.

We consider the *implicit Euler or Rothe method* where the approxima-

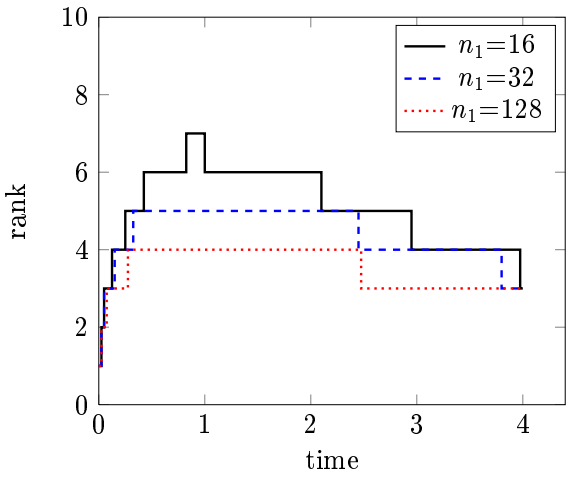


Figure 1: Ranks of truncated SVD approximation

tion p_k is defined as the solution of the linear system of equations

$$(I - \tau_k A)p_k = p_{k-1}, \quad k = 1, \ldots, K.$$

The discretisation t_k is chosen such that the error of the Rothe method is at most some ϵ.

The determination of p_k is done with a *gradient descent method*. This class of iterative methods defines a sequence $p_k^{(n)}$ by

$$p_k^{(n+1)} = p_k^{(n)} - \omega_n((I - \tau_k A)\, p_k^{(n)} - p_{k-1}).$$

Iterate from $n = 0, \ldots, N$ and choose N such that the error of the descent method is again bounded by ϵ. In the analysis one utilises the fact that the p_{k-1} is actually also only known approximately. The parameter ω_n is for example chosen according to the steepest descent method.

The main computational ingredient is the matrix vector product

$$Ap = \sum_{i=1}^{k} (S_{z_i} - I)\Lambda_i p.$$

As Λ_i are diagonal we first consider the values of (the tensor) $(\lambda_i(x))_x$. We note that in the simplest case which uses counts of species for the states and the law of mass action all these tensors are of rank one. Consequently the $\Lambda_i p$ all have the same rank as p, thus no rank reduction is required. In other cases a low rank representation of the tensor containing the values

Algorithm 1: $A^{\mathrm{CP}}(p)$: Operation of A on p

propensities λ_i are given in CP-format (exact or approximated initially)
shift operators S_{z_i} are given in CP-format
$Ap = 0$
for $i \in \{1, \dots, k\}$ **do**
  $A_i p = (S_{z_i} \lambda_i) \circledast (S_{z_i} p) - \lambda_i \circledast p$
  possible rank reduction of $A_i p$ ▷ $cp^{\#}(A_i p) = 2 \cdot cp^{\#}(\lambda_i) \cdot cp^{\#}(p)$
  $Ap = Ap + A_i p$
end
possible rank reduction of Ap ▷ $cp^{\#}(Ap) = \sum_{i=1}^{k} cp^{\#}(A_i p)$
return Ap

of the propensities is explicitly computed; here an approximation might be necessary to keep the rank low while controlling the error made. In two dimensions one uses the singular value decomposition (SVD), and for more dimensions one can use alternating least squares or other procedures [6]. The employed rank is chosen adaptively to get a desired accuracy. This is possible as the rank of the exact solution is finite and small. The value $S_{z_i} p$ is obtained by shifting all the components representing p by the value of component z_i. This shift operation is rank-invariant.

The algorithm $A^{\mathrm{CP}}(p)$ to compute Ap in the CP-format is shown in Algorithm 1. The rank of Ap after the application of the (approximated) tensors $(\lambda_i(x))_x$ can be as large as

$$2 \sum_{i=1}^{k} cp^{\#}(\lambda_i) \cdot cp^{\#}(p),$$

where $cp^{\#}$ equals r in the representation (4) of a tensor. Therefore a rank reduction algorithm often needs to be applied.

The full procedure is shown in Algorithm 2. The residual used in the gradient descent algorithm also needs to be stored as a low rank tensor. Due the multiplication with A the ranks of the residual might grow as well. Therefore a rank reduction algorithm needs to be applied which again needs to keep the error of the approximation under control.

Algorithm 2: Implicit Euler in CP-format

initial distribution p_0 is given in CP-format
time interval $[0, t_K]$ is divided into time steps $t_0 = 0 < t_1 < \cdots < t_K$
for $k \in \{1, \ldots, K\}$ **do** ▷ `implicit Euler`
  $\tau_k = t_k - t_{k-1}$
  $p^{it} = p_{k-1}$
  $res = \delta_t \cdot A^{\mathrm{CP}}(p^{it})$
  while *not good enough* **do** ▷ `grad. desc.:` $(I - \tau_k A)p_k = p_{k-1}$
    $res^{op} = res - \tau_k \cdot A^{\mathrm{CP}}(res)$
    $\alpha = (res^{op}, res)/(res^{op}, res^{op})$
    $p^{it} = p^{it} + \alpha \cdot res$
    possible rank reduction of p^{it} ▷ $cp^{\#}(p^{it}) = cp^{\#}(p^{it}) + cp^{\#}(res)$
    $res = res - \alpha \cdot res^{op}$
    possible rank reduct. of res ▷ $cp^{\#}(res) = cp^{\#}(res) + cp^{\#}(res^{op})$
  end
  $p_{k-1} = p^{it}$
  possible rank reduction of p_{k-1}
end

4 Computational experiments

We report on some experiments with a 2-dimensional problem with propensities

$$\lambda_i = 60e^{-t}\left(\frac{0.5}{x_i + x_{1-i}} + \frac{0.5}{x_i + 1000}\right) x_i, \quad i = 1, 2. \tag{7}$$

This model was suggested to us by Kevin Burrage originally. It models competing T-cell clonotypes. The basis for this study is the CMEPY software [1]. For our purposes we have added low rank approximations to CMEPY. The system converges towards a bistable distribution where one clonotype dominates in each of the branches, see Figure 2. It has some diagonal structure which is often unfortunate for tensor based approaches.

While overall, smaller time steps seem to result in smaller ranks the rank increases over the observation interval as seen in Figure 3(a). This tendency can also be seen in Figure 3(b), where in addition one can see that the error is proportional to the threshold. The cost for a reduction of the error by a factor of 10 is a rank increase by 2. This means that by a linear increase in complexity one can reduce the error by a constant factor. This exponential error dependency is what is encountered for spectral methods and for radial basis function approximation and needs to be further investigated. (For grid based methods the error is usually proportional to a power of the cost.) In

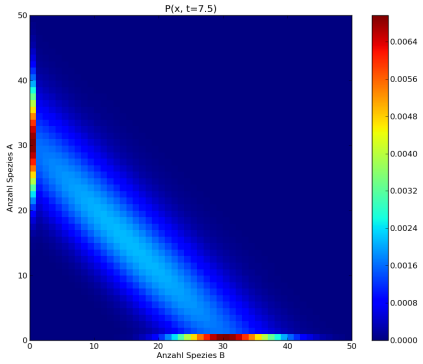


Figure 2: Solution p for the T-cell clonotypes model at time $t = 7.5$.

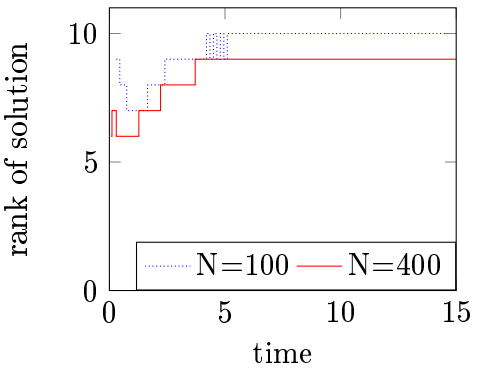
Figure 3(c) one sees that the largest observed rank of the residual *res* in Algorithm 2 after thresholding is proportional to the step size, so smaller time steps allow a more compact representation. The final plot, Figure 3(d), provides an insight into the complexity of the solver as it displays the number of gradient descent steps required at every integration step. With smaller step sizes one requires fewer iterations but it is interesting to note that the actual number of iterations required is mostly very small. These last two observations together explain that the observed total run time decreases with smaller time steps, in contrast with what one would expect. But this observation has to be taken with a grain of salt since our current implementation is far from optimal.

Acknowledgements The work on CMEPY was supported by the ARC Centre in Bioinformatics. The work by the second author was supported through the MATHEON Project A18. We would like to thank R. Fletcher-Costin for his work on CMEPY and S. Kamprath and M. Asher for their work on the tensor product approximation codes used in this project.

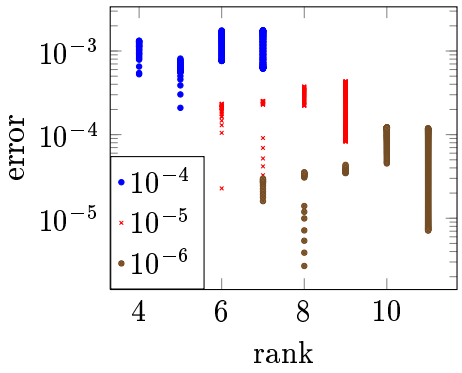
References

[1] CmePy Documentation. `http://fcostin.github.com/cmepy/`, 2010.

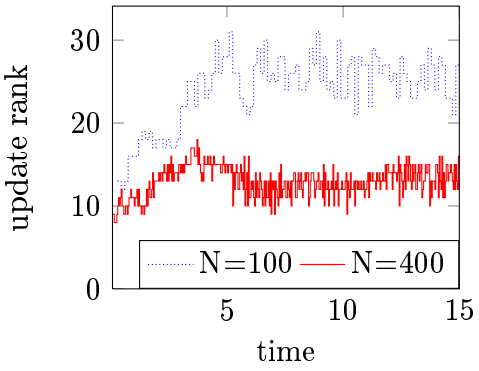
[2] K. Ball, T. G Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Probab.*, 16(4):19251961, 2006.
`http://dx.doi.org/10.1214/105051606000000420`



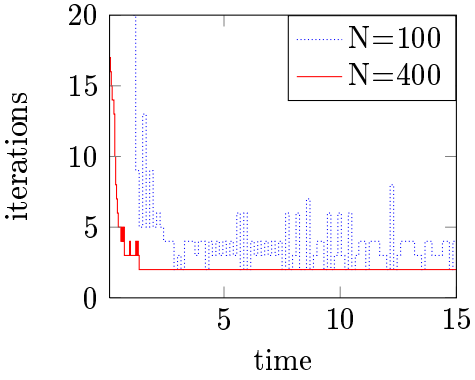
(a) solution rank over time



(b) ranks vs error for different thresholds, each mark represents one time step



(c) max. rank of update over time



(d) iterations for gradient descent

Figure 3: Some results for the T-cell clonotypes model

[3] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22(4):403–434, 1976.
http://dx.doi.org/10.1016/0021-9991(76)90041-3

[4] 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. http://dx.doi.org/10.1016/j.cam.2006.02.053

[5] Tobias Jahnke and Wilhelm Huisinga. A dynamical low-rank approach to the chemical master equation. *Bull. Math. Biol.*, 70(8):2283–2302, 2008. http://dx.doi.org/10.1007/s11538-008-9346-x

[6] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, 2009.
http://dx.doi.org/10.1137/07070111X

[7] Brian Munsky and Mustafa Khammash. A multiple time interval finite state projection algorithm for the solution to the chemical master equation. *J. Comput. Phys.*, 226(1):818–835, 2007.
http://dx.doi.org/10.1016/j.jcp.2007.05.016

[8] T. Tian, S. Xu, J. Gao, and K. Burrage. Simulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics*, 23(1):84–91, 2006.
http://dx.doi.org/10.1093/bioinformatics/btl552

[9] N. G. van Kampen. *Stochastic processes in physics and chemistry*, volume 888 of *Lecture Notes in Mathematics*. North-Holland Publishing Co., Amsterdam, 1981.

Author addresses

1. **Markus Hegland**, Mathematical Sciences Institute, The Australian National University, Australian Capital Territory 0200, Australia.
mailto:markus.hegland@anu.edu.au

2. **Jochen Garcke**, Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany
mailto:garcke@math.tu-berlin.de