18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Solving the Chemical Master Equation with the Aggregation-Disaggregation Method

M. Hegland[1]

[1] *Center for Mathematics and Its Applications (CMA), Mathematical Sciences Institute, The Australian National University, Australian Capital Territory*
Email: Markus.Hegland@anu.edu.au

Abstract. Originally, aggregation and disaggregation were considered as acceleration techniques similar to multigrid methods for the solution of linear systems of equations. Recently we have demonstrated that these methods can also be used for the numerical solution of the chemical master equation. Here three scenarios are discussed where aggregation and disaggregation accelerate convergence, reduce complexity and lead to approximate solutions.

Keywords: *Chemical master equation, aggregation / disaggregation, numerical algorithms*

1. INTRODUCTION

The chemical master equation (CME) describes how the probability distribution of the chemical reaction counts or chemical species counts changes over time. The numerical solution of the CME is challenged by problems relating to the number of reactions involved, large copy numbers, reactions with different time scales and singular non-symmetric matrices. The reduction of the system size is an essential approach which addresses many of these problems. Such a reduction which maintains total probability is aggregation. In order to make the solution of a reduced system useful to the solution of the "big problem" one requires an interpolation type approach called disaggregation which associates a probability distribution on the full state space with the reduced distribution. Different disaggregation methods lead to algorithms with different properties. In the rest of this paper it will be shown how three different disaggregations lead to multigrid-style methods, to faster solvers for the reduced problems and to piecewise linear approximations.

2. ACCELERATING CONVERGENCE

Here we introduce some of the basic concepts for aggregation-disaggregation methods and illustrate how a chemical master equation solver could be constructed using these methods. We will assume that the state space is finite but large with elements $\{0, \ldots, M-1\}$. A random variable on this state is denoted by X and a stochastic process by $X(t)$. A probability distribution on this state space is then a vector with M non-negative elements. In particular we define p and $p(t)$ to have the components $p_i = P(X = i)$ and $p_i(t) = P(X(t) = i)$ respectively, for $i = 0, \ldots, M-1$ and note that the indices of the components of p start with zero. We assume that $X(t)$ is a Markov process and thus p satisfies a master equation of the form

$$\frac{dp}{dt} = Ap$$

and some initial condition $p(0) = p_0$. An application of the backward Euler method to this problem gives for $t_0 = 0$ and $t_n = t_{n-1} + \Delta t$ the linear system of equations

$$(I - \Delta t A)p(t_n) = p(t_{n-1}) \tag{1}$$

which provides an approximation of the probability distribution (which for simplicity and as we will not discuss time discretisation we also denote by $p(t_n)$). As the number of unknowns M is large, iterative methods to solve these equations would be used. The simplest methods including Richardson, Jacobi and Gauss-Seidel methods, may, however converge slowly. This is a motivation for multigrid methods in general, and the aggregation-disaggregation approach in particular.

The approach has been discussed by [2] and uses two matrices, an aggregation matrix E_h (which provides the restriction in multigrid terminology) and the disaggregation matrix F_h (which provides the prolongation). The *aggregation matrix* $E_h \in \mathbb{R}^{M/h,M}$ is characterised by the width of the aggregation cell h. (We assume here for simplicity that M and h are powers of two.) Specifically, aggregation defines a new probability distribution $\overline{p}$ defined by

$$\overline{p}_k = (E_h p)_k = p_{kh} + p_{kh+1} + \cdots p_{(k+1)h-1}, \quad \text{for } k = 0, \ldots, M-1. \tag{2}$$

One can see that $\overline{p}$ is the probability distribution of the random variable $Y = \lfloor X/h \rfloor$ where $\lfloor \cdot \rfloor$ denotes the truncation to the next integer. The following very simple but fundamental proposition shows how the original distribution is recovered from the aggregated one using disaggregation.

Proposition 1. *Let $p \in \mathbb{R}^M$ be a probability distribution with $p_i = P(X = i)$ and $Y = \lfloor X/h \rfloor$ and let $D \in \mathbb{R}^{M,M}$ be a diagonal matrix with components $D_{i,i} = P(X = i \mid Y = \lfloor i/h \rfloor)$ Then*

$$p = D E_h^T \overline{p}$$

where $\overline{p}$ is the probability distribution of Y and E_h^T is the transpose of the aggregation matrix defined in equation (2).

Proof. As $\overline{p}_k = P(Y = k)$ one sees that $(E_h^T \overline{p})_i = P(Y = \lfloor i/h \rfloor)$ and thus

$$\begin{aligned} [D E_h^T \overline{p}]_i &= D_{i,i}\, P(Y = \lfloor i/h \rfloor) \\ &= P(X = i, Y = \lfloor i/h \rfloor) \\ &= P(X = i) = p_i \end{aligned}$$

as $Y = \lfloor X/h \rfloor$. □

It follows that for the *disaggregation matrix* $F_h = D E_h^T$ one has $p = F_h \overline{p}$ and as $\overline{p} = E_h p$ one has

$$p = F_h E_h p.$$

This was made possible because F_h depends on p. Note that in particular $F_h E_h$ has a rank less or equal to M/h and it thus cannot be equal to I which has a rank of M. The aggregation-disaggregation methods however, approximate this "ideal" F_h by using earlier iterates in an iterative method for the solution of equation (1). The convergence of the technique is based on a contractivity of the underlying procedure.

3. REDUCING COMPLEXITY

A major application of aggregation is the reduction of the size of the state space and consequently, the computational complexity of the CME. In this section we consider aggregation which is based on properties of the propensity functions. The state space is here a set of integer tuples e.g. $x \in \{0, \ldots, M-1\}^d$ where d is the number of reactions. The CME has the form

$$\frac{dp_x}{dt} = \sum_{j=1}^{d} \lambda_j(x - e_j) p_{x-e_j} - \lambda_j(x) p_x. \tag{3}$$

The propensities $\lambda_j(x)$ are assumed to be defined on all of $\mathbb{Z}^d$ but are zero if any of the components x_i of the argument x are either less than zero or larger or equal to M. It is furthermore assumed here that there exist functions $\nu_j : \mathbb{Z}^{\overline{d}} \to \mathbb{R}_+$ and an integer matrix $Z \in \mathbb{Z}^{\overline{d},d}$ such that

$$\lambda_j(x) = \nu_j(Zx), \quad j = 1, \ldots, d. \tag{4}$$

If X is a random variable with probability distribution p then $Y = ZX$ is a random variable with probability distribution $\overline{p}$ defined by an aggregation E_Z where

$$\overline{p}_y = (E_Z p)_y = \sum_{x \in Z^{-1}y} p_x.$$

One then has

Proposition 2. *Let $X(t)$ by a stochastic process with a probability distribution $p(t)$ satisfying equation (3). If the propensities λ_j satisfy the equations (4) then the probability distribution $\overline{p}$ of the stochastic process $Y(t) = ZX(t)$ satisfies the master equation*

$$\frac{d\overline{p}_y}{dt} = \sum_{j=1}^{d} \nu_j(y - z_j)\overline{p}_{y-z_j} - \nu_j(y)\overline{p}_y,$$

where $z_j = Ze_j$ are the columns of the matrix Z.

Proof. The claim follows directly from the relation (4) and equation (3):

$$\begin{aligned}
\frac{d\overline{p}_y}{dt} = \left(E_Z \frac{dp}{dt}\right)_y &= \sum_{x \in Z^{-1}y} \sum_{j=1}^{d} \lambda_j(x - e_j)p_{x-e_j} - \lambda_j(x)p_x \\
&= \sum_{j=1}^{d} \sum_{x \in Z^{-1}y} \nu_j(Zx - z_j)p_{x-e_j} - \nu_j(Zx)p_x \\
&= \sum_{j=1}^{d} \nu_j(y - z_j) \sum_{x \in Z^{-1}y} p_{x-e_j} - \nu_j(y) \sum_{x \in Z^{-1}y} p_x \\
&= \sum_{j=1}^{d} \nu_j(y - z_j)\overline{p}_{y-z_j} - \nu_j(y)\overline{p}_y
\end{aligned}$$

as $\sum_{x \in Z^{-1}y} p_{x-e_j} = \sum_{x' \in Z^{-1}(y-z_j)} p_{x'} = \overline{p}_{y-z_j}$ where $z_j = Ze_j$. □

The most important example is the transformation from the interpretation of chemical reactions as counting processes to actual reactions between chemical species defined the by counts of the various species involved. In this case the vector y corresponds to the vector of counts of species and the matrix Z is the *stoichiometric matrix*. If the number of species is much lower than the number of reactions then this approach may substantially reduce the complexity of the master equation. The trade-off is that one has to deal with shifts z_j and thus more dense and less regular matrices of the CME. One also looses the triangular structure. However, as one is only interested in $\overline{p}$ methods to solve this reduced CME are of interest.

A second example is based on the fact that many reactions are reversible and thus the propensities (which depend on the species counts) do depend on the differences of the counts $x_{j_f} - x_{j_b}$ of the forward and backward reactions only. Thus again one can introduce this difference as a new variable y and can get a reduction of the state space of up to a factor two. This comes at the cost of a matrix A which will have a symmetric structure (but not symmetric values) which is roughly a reduced version of $A + A^T$ if A is the original matrix.

While in general the existence of a (best Z) is an algebraic problem usually such a Z is given by the application and is often related to species counts and to conservation of sums of such counts.

One may now introduce a disaggregation matrix F_Z which leads to algorithms for the solution of the reduced system based on solvers for the larger system for a reduced state space. To get this one assumes that Z and has been suitable transformed (using results from number theory), such that it is of the form $Z = [I, Z_1]$ where I is the identity such that $y = Zx = x_0 + Z_1 x_1$ where x_0 and x_1 are the vectors of the first $\overline{d}$ components of x and the remaining $d - \overline{d}$ components, respectively. If one now chooses the disaggregation to be $F = \begin{bmatrix} I \\ 0 \end{bmatrix}$, i.e., $(F\overline{p})_{(y,0)} = \overline{p}_y$ and $(F\overline{p})_{(x_0,x_1)} = 0$ else, one has $EF = I$. One can show that the CME for $\overline{p}$ given in proposition (2) then takes the form

$$\frac{d\overline{p}}{dt} = E_Z A F_Z \, \overline{p}$$

and the initial condition is $\overline{p}(0) = E_Z p(0)$. Based on this equation one can now solve the original CME for a very short time so that the state space where the probability distribution is only a small factor larger than the state space of the reduced variable y, then use an aggregation with E_Z followed by a disaggregation with F_Z to "compress" the probability distribution. If this process is repeated one obtains an efficient algorithm to solve the CME using a fast solver for the original CME on a small subspace.

4. APPROXIMATING SOLUTIONS

Here we consider again the piecewise aggregation but will use a disaggregation which is based on a piecewise linear approximation. A related approach was first suggested by Ferm and Loetstedt in their paper [1] and we will consider an error analysis of a simplified variant of their proposal. The motivation for this approach is the often one finds that the probability distribution varies slowly over the states x. For the following let E_h be as defined in equation (2) and F_h defined by

$$(F_h\overline{p})_x = \frac{\overline{p}_{\lfloor x/h \rfloor}}{h} + (x - yh - (h-1)/2)\,\frac{\overline{p}_{\lfloor x/h \rfloor + 1} - \overline{p}_{\lfloor x/h \rfloor - 1}}{2h^2}, \tag{5}$$

where $\overline{p}_{M/h} = \overline{p}_0$ and $\overline{p}_{-1} = \overline{p}_{M/h-1}$. It turns out that this disaggregation satisfies the consistency condition $E_h F_h = I$ and furthermore, it recovers "linear parts" of the original probability distribution exactly:

Proposition 3. *Let the aggregation matrix $E_h \in \mathbb{R}^{M/h,M}$ be defined as in equation 2 and the disaggregation matrix be defined as in equation 5. Then*

(i) $E_h F_h = I$

(ii) if $p_x = \alpha x + \beta$ *for some* $x = k_0 h, \ldots, k_1 h$ *then*

$$(F_h E_h p)_x = p_x, \quad \text{for } x = (k_0+1)h, \ldots, (k_1 - 1)h$$

Proof.

(i) Let $(\tilde{p})_y = F_h\overline{p}$ for some probability distribution p. Then

$$\begin{aligned}(E_h\tilde{p})_y &= \sum_{x=yh}^{yh+h-1} \tilde{p}_x \\ &= \overline{p}_y + \sum_{x=yh}^{yh+h-1} (x - yh - (h-1)/2)\,(\overline{p}_{y+1} - \overline{p}_{y-1})/(2h^2).\end{aligned}$$

As $\sum_{x=yh}^{yh+h-1}(x - yh - (h-1)/2) = 0$ one then gets $(E_h\tilde{p})_y = \overline{p}_y$ and thus $E_h F_h \overline{p} = \overline{p}$.

(ii) Let $\overline{p} = E_h p$. One then has

$$\overline{p}_y = \sum_{x=yh}^{yh+h-1} (\alpha x + \beta) = \alpha h(hy + (h+1)/2) + h\beta, \quad \text{for } y = k_0, \ldots, k_1$$

and with $x = yh + i$, for $i = 0, \ldots, h-1$ and $y = k_0 + 1, \ldots, k_1 - 1$ one gets for $\tilde{p} = F_h \overline{p}$:

$$\begin{aligned} \tilde{p}_x &= \overline{p}_y/h + (x - yh - (h-1)/2)\,(\overline{p}_{y+1} - \overline{p}_{y-1})/(2h^2) \\ &= \alpha(hy + (h-1)/2) + \beta + (x - yh - (h-1)/2)\alpha \\ &= \alpha x + \beta. \end{aligned}$$

as $\overline{p}_{y+1} - \overline{p}_{y-1} = 2\alpha h^2$.

□

For a bound on the approximation error of $F_h E_h p$ one rewrites the aggregation and disaggregation matrices using cyclic matrices and sampling. First let the *cyclic shift matrix* $S \in \mathbb{R}^{M,M}$ be defined by $(Sp)_x = p_{x+1}$ where, as usual, $p_M = p_0$. Furthermore, let the *sampling matrix* $Q_h \in \mathbb{R}^{M/h,M}$ be defined by $(Q_h p)_y = p_{yh}$. Furthermore, define the two polynomials $e_h(z)$ and $f_h(z)$ by $e_h(z) = \sum_{k=0}^{h-1} z^k$ and

$$f_h(z) = \frac{e_h(z^{-1})}{h} + \frac{z^h - z^{-h}}{2h^2} \sum_{k=0}^{h-1} (k - (h-1)/2) z^{-k}. \tag{6}$$

One can then show the lemma

Lemma 1. *Let S by the cyclic shift matrix, Q_h the sampling matrix and e_h and f_h be suitably defined polynomials. Then*

(i) for $e_h(z) = \sum_{k=0}^{h} z^k$ one has

$$E_h = Q_h e_h(S)$$

(ii) for $f_h(z)$ as defined in equation (6)

$$F_h = f_h(S) Q_h^T$$

(iii) and with $D_h = \mathrm{diag}(\exp(-2\pi i k h/M))_{k=1}^{M}$ one has

$$Q_h^T Q_h = \frac{1}{h} \sum_{j=0}^{h-1} D_h^j$$

The proof of this lemma is obtained by inserting the definitions. Note the third statement which amounts to a matrix version of the *sampling theorem*. Using this lemma and the triangle inequality one then gets the error bound

Proposition 4. *The error norm of the aggregation-disaggregation approximation $F_h E_h p$ is bounded by*

$$\|F_h E_h p - p\| \leq \|f_h(S) e_h(S) p - p\| + \frac{1}{h} \sum_{j=1}^{h-1} \|f_h(S) D_h^j e_h(S) p\| \tag{7}$$

In signal processing terminology, the first term in the bound is the approximation error of the the combination of the two filters while the second term takes into account the aliasing which occurs because of the sampling. Both errors can be bounded theoretically and we found that both grow like $O(h^2)$. This growth is confirmed experimentally where the error of an approximation of the probability distribution of a Poisson process was determined, see table 1.

Table 1: error norms of the piecewise linear approximation for a Poisson distribution with $\lambda t = 6 \cdot 10^4$.

h	l_1-error	ratio	l_2-error	ratio
1	0	–	0	–
2	$3.85 \cdot 10^{-8}$	–	$1.19 \cdot 10^{-9}$	–
4	$8.07 \cdot 10^{-6}$	209.00	$2.45 \cdot 10^{-7}$	207.00
8	$3.24 \cdot 10^{-5}$	4.02	$1.13 \cdot 10^{-6}$	4.59
16	$1.34 \cdot 10^{-4}$	4.14	$4.68 \cdot 10^{-6}$	4.16
32	$5.66 \cdot 10^{-4}$	4.21	$1.94 \cdot 10^{-5}$	4.15
64	$2.57 \cdot 10^{-3}$	4.55	$8.60 \cdot 10^{-5}$	4.43
128	$1.34 \cdot 10^{-2}$	5.21	$4.39 \cdot 10^{-4}$	5.10
256	$7.48 \cdot 10^{-2}$	5.58	$2.37 \cdot 10^{-3}$	5.39
512	0.37	4.90	$1.04 \cdot 10^{-2}$	4.39
1,024	0.66	1.80	$1.69 \cdot 10^{-2}$	1.63
2,048	1.26	1.91	$2.62 \cdot 10^{-2}$	1.55
4,096	1.81	1.43	$3.05 \cdot 10^{-2}$	1.16
8,192	1.86	1.03	$3.29 \cdot 10^{-2}$	1.08
16,384	1.93	1.04	$3.28 \cdot 10^{-2}$	1.00

5. CONCLUSION

Aggregation and disaggregation is related to restriction and prolongation of multigrid methods, compression and decompression in image and signal processing and sampling and interpolation of function approximation. These three views led to three different styles of algorithms, a multigrid method for the solution of the time-discretised CME, an efficient method for the solution of the reduced system using a generic solver for the original system and a piecewise linear approximation of smooth probability distributions which preserve total probability.

Many challenges have not been touched in this short paper. This includes in particular the solution of problems with many reactions. The most promising approaches suggested previously include sparse grids [3] and the finite state projection method [4]. Both approaches are based on aggregation and disaggregation. In comparison to traditional numerical techniques the functions to be approximated here depend on (vector) integer variables. A rigorous analysis of errors requires the application of Fourier space techniques and is currently under development.

REFERENCES

[1] Lars Ferm and Per Lötstedt. Adaptive solution of the master equation in low dimensions. *Appl. Numer. Math.*, 59(1):187–204, 2009.

[2] Moshe Haviv. Aggregation/disaggregation methods for computing the stationary distribution of a markov chain. *SIAM Journal on Numerical Analysis*, 24(4):952–966, 1987.

[3] 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.

[4] 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.