

加速随机模拟化学反应系统的 “最后所有可能的步进”方法*

周文^{1,2}, 彭新俊¹, 刘祥¹, 闫正楼¹, 王翼飞¹

(1. 上海大学 数学系, 上海 200444;

2. 安徽师范大学 数学与计算机科学学院, 安徽 芜湖 241000)

(郭兴明推荐)

摘要: 提出了一种改进的模拟化学反应系统的加速方法, 称为“最后所有可能的步进”(FAPS)方法. 此方法不仅可以用较少的模拟次数得到整个运行过程中任何时刻物种的统计特性, 而且还可以与 leap 算法相结合, 有效地进行大规模的化学反应系统的模拟. 数值模拟的结果表明提出的方法可以广泛地应用于多种化学反应系统, 并且与已有的方法相比其运行效率有显著的改进.

关键词: “最后所有可能的步进”方法; 随机模拟算法; 化学反应系统

中图分类号: O644; O242.1 **文献标识码:** A

引言

由活细胞形成的系统中, 某些关键的反应物的分子数目可能较小, 而离散性和随机性则起着很重要的作用, 于是化学反应系统的随机模拟研究引起了学者们的浓厚兴趣^[1-4]. 有微观物理基础的化学主方程(CME)刻画了均匀混合的化学反应系统的化学动力学的随机性态. 由于一般难以得到化学主方程的解析解, Gillespie^[5-6]提出了随机模拟算法(SSA)模拟由化学主方程所保证的化学反应系统的精确的统计性质. 然而在 SSA 中, 每步只允许 1 个反应发生, 模拟速度较慢. 尽管最近对 SSA 有了重大的提速^[7-8], 但对于一些实际问题仍然不适用.

为了进一步加速离散的随机模拟方法, Gillespie 提出了一种近似的方法(τ -leap 方法)^[9]. 之后很多学者也提出了一些改进 τ -leap 的方法, 如 Gillespie 和 Petzold^[10]改进了原始的选取 τ 的方案, 其中的 τ 都是由 leap 条件(每一步的状态改变足够小使得任何倾向函数的值在宏观上都没有很大的改变)确定, 将仅用倾向函数的平均值满足 leap 条件改进为由其平均值和方差共同满足 leap 条件; Rathinam 等人^[11]引入了一种隐式的 τ -leap 来解决刚性问题; Tian 与 Burrage^[12]两人和 Chatterjee 等人^[13]各自引入了二项式 τ -leap 方法以避免负的物种数目的产

* 收稿日期: 2007-08-03; 修订日期: 2008-01-29

基金项目: 国家自然科学基金资助项目(30571059); 国家高技术研究发展(S63)计划资助项目(AA0272190)

作者简介: 周文(1980—), 女, 安徽人, 讲师, 博士生(E-mail: zhouw83@shu.edu.cn);
王翼飞(1948—), 男, 教授(联系人, Tel: + 86-21-66134331; E-mail: yifei_wang@staff.shu.edu.cn).

生; Peng 等人^[14]改进了二项式 τ -leap 方法以提高精确性, 扩大应用范围. 最近 Cai 和 Xu^[15] 提出了 K -leap 方法, 几乎同时 Auger 等人^[16] 获得了 R -leap 方法. 这两个方法均是限制在一步跳跃期间总的反应次数是 K (或 R), 从而更好地满足 leap 条件, 提高了模拟的精确性.

然而, 上述这些方法都是关注于算法本身的改进, 而没有对输出数据进行有效的分析. 最近, Lipshtat^[17] 提出了作用于整个时间过程的“所有可能的步进”(APS)方法, 减少了对各阶矩进行满足要求的估计所需要的运行次数. 这个方法得到的结果是物种在整个时间过程中的统计特性. 但是, 我们需要的常常是物种在某个时刻的统计特征^[5, 6, 9-20].

而且, APS 方法在每一步都要考虑所有可能的状态, 如果将其应用到 leap 方法中, 我们就必须考虑所有可能的运行轨迹而不仅仅是 leap 方法中所有可能的步骤, 这是难以实现的.

在本文中, 我们提出了基于 SSA 的“最后所有可能的步进”(FAPS)方法. FAPS 方法是在模拟到最后时刻的前一步根据反应发生的概率考虑所有可能的状态. 这样, 我们用较少的模拟次数得到了最后时刻所有物种的理想统计特性. 如果我们需要某一时刻的统计结果, 只要令这个时刻是“最后时刻”来运行程序即可. 进而, FAPS 方法可以得到任何时刻的信息. 由于 FAPS 方法没有改变系统的演化, 而只是利用了所有可能的概率, 所以比 SSA 使用较少的模拟次数却获得了较为精确的结果. 并且, FAPS 方法可以应用于几乎所有的 leap 方法, 得到 FAPS leap 方法. 对于模拟有很多反应通道的真实系统而言, 由于 leap 方法需要的运行时间比 SSA 少得多, 所以 FAPS leap 方法更适用. 下面, 我们将分别简单介绍 SSA, τ -leap 方法, “所有可能的步进”方法.

1) SSA

我们考虑 $N \geq 1$ 种均匀混合的分子 $\{S_1, \dots, S_N\}$ 通过 $M \geq 1$ 个反应通道 $\{R_1, \dots, R_M\}$ 进行化学反应, 使用状态向量 $X(t) = (X_1(t), \dots, X_N(t))$ 描述化学系统的状态. 其中 $X_i(t)$ ($i = 1, \dots, N$), 表示物种 S_i 在 t 这个时刻的分子数, 且分子数 $X_i(t)$ 是由反应通道所确定的随机变量. 反应通道 R_j 的动力学特征由其倾向函数 a_j 与状态改变向量 $V_j = (v_{1j}, \dots, v_{Nj})$ 决定. v_{ij} 是由 R_j 反应导致的 S_i 分子数目的改变量. 若 $X(t) = x$, 在下一个无限小的时间间隔 $[t, t + dt)$ 内 R_j 会发生的概率为 $a_j(x)dt$. 因此系统中的下一个反应将在无限小的时间间隔 $[t + \tau, t + \tau + d\tau)$ 内发生且是 R_μ 的概率为

$$P(\tau, \mu | x, t) = a_\mu(x) e^{-a_0(x)\tau}, \quad \tau \geq 0; \mu = 1, \dots, M, \quad (1)$$

其中

$$a_0(x) = \sum_{\mu=1}^M a_\mu(x). \quad (2)$$

在 SSA 中, Gillespie 表明下一次反应发生的时间 τ 和相应的反应编号 μ 分别有下列的概率密度函数:

$$p(\tau) = a_0(x) e^{-a_0(x)\tau}, \quad \tau > 0 \quad (3)$$

$$\text{与} \quad p(\mu) = a_\mu(x) / a_0(x), \quad \mu = 1, \dots, M. \quad (4)$$

则 τ 和 μ 可以分别由方程(3)和(4)所确定的两个独立的一致随机变量产生, 系统的状态更新为 $X(t + \tau) = x + v_\mu$. 重复这一过程, 直到达到某一结束的时间或者条件为止. SSA 精确地模拟每一反应事件, 且每步只有 1 个反应发生, 这不仅很耗时而且还不能用于一些真实的系统.

2) τ -leap 方法

τ leap 方法允许每步可能有一些反应通道激活几次, 加速了随机模拟的运行. 令

$$K_j(\tau; \mathbf{x}, t) \stackrel{\Delta}{=} \text{给定 } X(t) = \mathbf{x}, \text{ 反应通道 } R_j \text{ 在时间间隔 } [t, t + \tau) \text{ 内激活的次数,} \\ j = 1, \dots, M. \quad (5)$$

当 τ 足够小, 使得满足 leap 条件 ($[t, t + \tau)$ 区间内的状态改变很微小, 使得期间的倾向函数的值没有很大的改变) 时, 每个 $K_j(\tau; \mathbf{x}, t)$ 可以近似地由 $P_j(a_j(\mathbf{x}), \tau)$ 给出. 其中, $P_j(a_j(\mathbf{x}), \tau)$ 是均值为 $a_j\tau$ 的独立的 Poisson 随机变量. 因此, Gillespie 的 τ leap 方法在每一步跳跃 τ 之后将状态更新为

$$X(t + \tau) \approx \mathbf{x} + \sum_{j=1}^M \nu_j P_j(a_j(\mathbf{x}), \tau). \quad (6)$$

为了确定符合 leap 条件的 τ 的最大值, Gillespie 作了下列限制来满足 leap 条件:

$$|\Delta a_j(\tau, \mathbf{x})| \leq \varepsilon a_0(\mathbf{x}), \quad \forall j = 1, \dots, M, \quad (7)$$

其中, $\Delta a_j(\tau, \mathbf{x}) \stackrel{\Delta}{=} a_j(X(t + \tau)) - a_j(\mathbf{x})$, ε 是满足 $0 < \varepsilon \ll 1$ 的一个预设的误差控制参数.

由于 $\Delta a_j(\tau, \mathbf{x})$ 是一个随机变量, 所以很难直接从方程(7) 解得 τ 的值. Gillespie^[9] 最初提出使用 $\Delta a_j(\tau, \mathbf{x})$ 的一阶 Taylor 展开作为它的近似值, 用 $\varepsilon a_0(\mathbf{x})$ 作为这个 $\Delta a_j(\tau, \mathbf{x})$ 的近似值的绝对平均值的边界. 后来, Gillespie 和 Petzold^[10] 表明 τ 的最大值应该由 $\varepsilon a_0(\mathbf{x})$ 作为 $\Delta a_j(\tau, \mathbf{x})$ 的近似值的绝对平均值和标准方差的边界来确定. K -leap 方法^[15] 限定在一步跳跃期间总的反应次数是根据 leap 条件计算得到的一个常数 K . 当在一步跳跃期间每个反应激活的次数由一个合适的数作为边界时, 可以更好地满足 leap 条件, 从而提高了模拟的精确性. 总之, 上述这些方法都关注于算法的直接提高, 而没有对数据进行分析.

3) “所有可能的步进”方法

Lipshtat^[21] 提出了“所有可能的步进”(APS)方法来获取物种在整个时间过程中的统计性质. 该方法通过分析输出数据, 计算了运行 SSA 的整个过程中所有可能的访问状态的概率而并没有实际地去访问它们. 假设在第 i 步, 系统的状态是 $\mathbf{x} = X(t_i)$. APS 方法考虑下一状态 $\mathbf{x} + \nu_\mu$ 的所有可能的反应, 更新其(非正则化的) 概率 $q(\mathbf{x} + \nu_\mu)$ 为 $q(\mathbf{x} + \nu_\mu) + (a_\mu/a_0)\tau_\mu$ ($\mu = 1, \dots, M$). 然后运行 SSA, 即随机地选择一个反应 R_μ 且相应地将状态 \mathbf{x} 更新为 $\mathbf{x} + \nu_\mu$. Lipshtat^[21] 同时还将该方法推广到了对于时间过程的模拟. 对落在时间过程 $[t_i, t_i + \tau_\mu)$ 中的每个时刻 t_i , 系统在这个时刻的概率 $q(\mathbf{x}, t_i)$ 在下一状态 $\mathbf{x} + \nu_\mu$ 将被更新为 $q(\mathbf{x}, t_i) + a_\mu/a_0$ ($\mu = 1, \dots, M$).

我们的算法将在第 1 节提出. 在第 2 节中, 我们对两个化学反应系统进行数值模拟, 例证了 FAPS 方法和 FAPS leap 方法的实现情况. 最后我们进行了讨论和总结.

1 “最后所有可能的步进”方法

我们知道对各阶矩的估计是随机模拟方法的典型目标. 由于有很多方法可以将最终的轨线转化为矩, 我们有必要找到其中最好的一种方法. “所有可能的步进”方法是基于 SSA 来获得一个时间过程中的各阶矩的加速了的方法. 实际问题中, 我们常常想估计在某一时刻的矩. 正如 Gillespie 所说的^[5], “对耦合的化学反应问题的通常的随机方法……关注于概率函数 $P(X_1, X_2, \dots, X_N; t) \stackrel{\Delta}{=} \text{在 } V \text{ 中 } t \text{ 时刻有 } X_1 \text{ 个 } S_1 \text{ 分子, } X_2 \text{ 个 } S_2 \text{ 分子, } \dots, X_N \text{ 个 } S_N \text{ 分子的概率, 它的矩}$

$$X_i^{(k)}(t) \equiv \sum_{X_1=0}^{\infty} \dots \sum_{X_N=0}^{\infty} X_i^k P(X_1, X_2, \dots, X_N; t) \quad (i = 1, \dots, N; k = 1, 2, \dots)$$

是对随机过程中从时刻

0 到时刻 t 的一些重复运行的平均, 且每次从相同的初始状态 $\{X_i^{(0)}\}$ 开始”.

虽然也可以使用其他的方法来计算各阶矩, 根据概率分布得到的结果是精确的. 这里我们提出一种改进的方法来计算最后时刻的概率, 称为“最后所有可能的步进”(FAPS)方法. 该方法在最后时刻考虑所有可能的状态而不去实际地运行它们, 从而提高了模拟效率. 在到达最后时刻的前一步, 我们假设在 SSA 中系统的状态是 $X(t) = \mathbf{x}$, 有 M 个可能的状态分别以概率 $a_\mu(\mathbf{x})/a_0(\mathbf{x})$ ($\mu = 1, \dots, M$) 发生. 然而 SSA 仅根据点概率 $a_\mu(\mathbf{x})/a_0(\mathbf{x})$ 随机地选取下一反应 R_j , 然后更新 $\mathbf{x} + \mathbf{v}_j$ 的概率为

$$q(\mathbf{x} + \mathbf{v}_j) \rightarrow q(\mathbf{x} + \mathbf{v}_j) + 1. \quad (8)$$

在 FAPS 方法中, 我们考虑虚拟的所有可能的反应, 更新概率如下:

$$q(\mathbf{x} + \mathbf{v}_\mu) \rightarrow q(\mathbf{x} + \mathbf{v}_\mu) + a_\mu(\mathbf{x})/a_0(\mathbf{x}), \quad \mu = 1, \dots, M, \quad (9)$$

所有需要的运行次数结束后, 对概率进行正则化.

基于 SSA 的 FAPS 方法总结为如下的算法:

算法 1 (FAPS)

- 1) 初始化 $t_0 = 0, t_{\text{final}} = T, \mathbf{x} = \mathbf{x}(t_0), p(\mathbf{x}) = 0, q(\mathbf{x}) = 0$;
- 2) 重复运行步骤 3) ~ 5), 直至最后时刻 T 的前一步;

(SSA)

- 3) 计算倾向函数 $a_\mu(\mathbf{x})$ ($\mu = 1, \dots, M$) 与 $a_0(\mathbf{x})$;
- 4) 依据均值为 $a_0^{-1}(\mathbf{x})$ 的指数分布产生 τ , 时间 t 更新为 $t + \tau$;
- 5) 由点概率 $a_\mu(\mathbf{x})/a_0(\mathbf{x})$ 选择下个反应为 R_μ , 系统的状态由 \mathbf{x} 更新为 $\mathbf{x} + \mathbf{v}_\mu$;

(FAPS)

- 6) 根据方程(9), 更新所有可能的状态的概率;
- 7) 运行 L 步后, 将 $q(\mathbf{x})$ 归一化为 $p(\mathbf{x}) = q(\mathbf{x}) / \sum_{\mathbf{x}} q(\mathbf{x})$.

步骤 1) ~ 6) 在每一次运行中都执行. 在每次运行期间, FAPS 算法比 SSA 只是需要多执行几步数学上的概率计算, 因此不需要更多的运行时间. 运行次数 L 由要求的精确性决定. 在相同的精度要求下, FAPS 方法中的 L 比 SSA 中的运行次数小几倍.

由于 SSA 模拟每个反应事件而且每次只执行一个反应, 所以对于包含有大数目的物种的系统是无效的. 有必要考虑加速 FAPS 算法. FAPS 算法仅在最后时刻执行, 所以可以结合到一些 leap 算法, 如 τ -leap 方法^[9-11], 二项式 τ -leap 方法^[12-14] 及 K -leap 方法^[15] 等. 并且, 我们不需要像结合到 APS 方法中那样考虑所有可能的轨线, 而只需在 t 到达最后时刻的前一步使用 FAPS 方法即可. FAPS leap 算法如下:

算法 2 (FAPS leap)

- 1) 初始化 $t_0 = 0, t_{\text{final}} = T, \mathbf{x} = \mathbf{x}(t_0), p(\mathbf{x}) = 0, q(\mathbf{x}) = 0$;
- 2) 运行 leap 算法产生 leap 时间 τ , 更新时间 t 与状态 $\mathbf{x}(t)$ 直至 $t_1 + \tau > T$;
- 3) 令 $t = t_1, \mathbf{x} = \mathbf{x}(t_1)$;
- 4) 运行 FAPS 算法中的步骤 2) ~ 7).

在步骤 2) 中, leap 算法可以是上述提到的任何算法^[9-16]. 步骤 1)、2)、3) 及步骤 4) 中的第 2) ~ 6) 步发生于每次运行中.

2 结果与讨论

现在我们以两个化学反应系统为例来验证 FAPS 方法的精确性和有效性. 在很难求解化

学主方程的情况下,为得到不同方法的精确性和有效性,我们首先画出经过一系列重复的 SSA 运行所得的最后时刻的分子数目直方图.然后分别使用 leap 方法, FAPS leap 算法和 FAPS 算法,对相同的时间间隔模拟相同的化学反应系统.最后,我们计算精确的 SSA 和其他的方法所得结果间的直方图距离^[17],比较相同的直方图距离情况下的运行次数和 CPU 时间.如果“自身距离很小”,直方图距离提供了 1 个误差的测度.全部模拟是在 1.99 GHz 处理器, 256 Mb 内存, WINDOWSXP 平台, MATLAB 环境下完成的.

2.1 独立双通道系统

这个简单的例子为



用这个例子来测试不同的模拟方法的精确性和有效性.如果 S_1 分子的初始数目为 $X_1(0) = x_0$, 可以从化学主方程解出 $X_1(t)$ 的概率密度函数为^[5]

$$p[X_1(t) = x | x_0, 0] = \frac{x_0!}{x!(x_0 - x)!} [e^{-ct}]^x [1 - e^{-ct}]^{x_0 - x}, \quad x = 0, \dots, x_0. \quad (11)$$

这样,我们可以计算由 SSA, FAPS 方法, τ -leap 算法, FAPS leap 算法得到的 $X_1(2)$ 的直方图与方程(11)中 $X_1(2)$ 的概率密度函数间的直方图距离,然后比较相同的精度时他们的运行次数.在系统运行到最后时刻的前一步时,若 S_1 分子的数目是 x_1 , 我们更新 S_1 分子下一步所有可能的状态为

$$q(x_1 - 1) \rightarrow q(x_1 - 1) + a_1/a_0, \quad q(x_1) \rightarrow q(x_1) + a_2/a_0. \quad (12)$$

然而,在 SSA 中,仅仅 S_1 分子的下一状态发生的概率 $q(x_1)$ (或者 $q(x_1 - 1)$) 更新为

$$q(x_1) \rightarrow q(x_1) + 1 \quad \text{或者} \quad q(x_1 - 1) \rightarrow q(x_1 - 1) + 1. \quad (13)$$

我们模拟这个系统在时间间隔 $[0, 2]$ 内的演化,其中初始条件为 $X(0) = (3\,000, 3\,000, 10^4, 0)$, 反应率常数 $c = (1, 10^{-4})$.

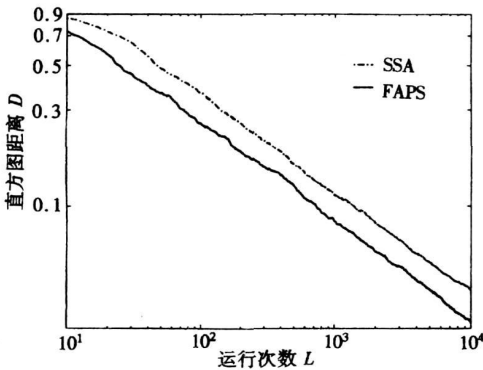


图 1 独立双通道系统(10)的 $X_1(2)$ 的直方图距离

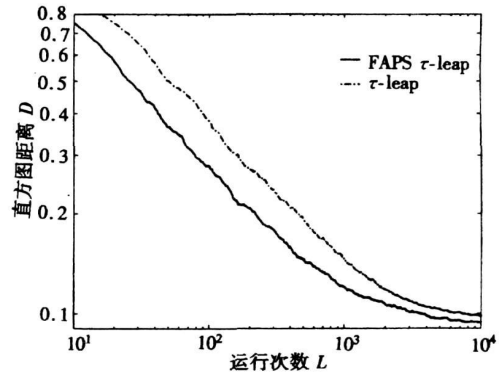


图 2 独立双通道系统(10)的 $X_1(2)$ 的直方图距离

图 1 表示的是使用 SSA 和 FAPS 算法在 $t = 2$ 时刻 $X_1(t)$ 的直方图距离.我们注意到在每种情况下直方图距离随着运行次数的增加而减少.由于对于任意给定的精度, SSA 需要比 FAPS 算法多将近 2 倍的运行次数,所以 FAPS 算法胜过 SSA.例如,为得到误差 0.1,使用 SSA 平均需要 1 388 次运行,而应用 FAPS 算法平均仅需 735 次.这是因为 S_1 分子的数目 x_1 有 2 个

可能的状态: $x_1 - 1$ 和 x_1 . 且这 2 种可能的状态的发生概率是不同的, 因此并不能达到确切的 2 倍关系. 此特征也适用于 leap 算法. 图 2 表示了使用 Gillespie 和 Petzold 的 τ -leap 算法^[10] 及相应的 FAPS τ -leap 算法在 $\varepsilon = 0.01$ 时 $X_1(2)$ 的直方图距离对运行次数的图形. 为进一步表明 FAPS leap 算法的优势, 我们在表 1 中列出了使用 Gillespie 和 Petzold 的 τ -leap 算法^[10] 及相应的 FAPS τ -leap 算法在取不同的 ε 时, $X_1(2)$ 的直方图距离及对应的运行次数和 CPU 时间. 其中, FAPS τ -leap 算法和 τ -leap 算法均使用 Gillespie 和 Petzold 的文章(参考文献[10]) 中的方程(6) 计算 τ , 分别应用上述公式(12) 和(13) 计算 $t = 2$ 时的概率. 从这张表我们可以看出 FAPS τ -leap 算法所需的 CPU 时间也比 τ -leap 算法的少很多.

表 1 独立双通道系统(10) 在各种 ε 取值时获得相应的 $X_1(2)$ 的直方图距离所需的运行次数 L 和 CPU 时间 t

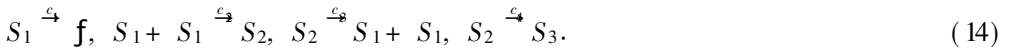
方法	参 数								
	$\varepsilon = 0.04$			$\varepsilon = 0.02$			$\varepsilon = 0.01$		
	次数 L	时间 t/s	直方图距离 D	次数 L	时间 t/s	直方图距离 D	次数 L	时间 t/s	直方图距离 D
τ -leap	2 116	148.6	0.335	2 526	230.8	0.181	7 205	982.0	0.100
FAPST-leap	1 327	108.1	0.335	1 596	159.9	0.181	3 562	519.1	0.100

表 2 使用不同的参数值 ε 得到降解—二聚化模型中的 $X_1(15)$ 的直方图距离平均所需要的 CPU 时间 t 和运行次数 L

方法	参 数								
	$\varepsilon = 0.04$			$\varepsilon = 0.03$			$\varepsilon = 0.02$		
	次数 L	时间 t/s	直方图距离 D	次数 L	时间 t/s	直方图距离 D	次数 L	时间 t/s	直方图距离 D
τ -leap	3 994	4 018	0.060	4 573	5 958	0.055	7 737	15 251	0.100
FAPST-leap	1 744	1 813	0.060	1 760	2 315	0.055	2 590	5 321	0.100

2.2 降解——二聚化模型

该例子已在早期的文章^[9, 10, 15] 中使用过来检验他们的方法, 降解——二聚化模型包含下述 4 个反应:



在我们的模拟中, 使用下列的反应率常数及初始分子数目

$$c_1 = 0.1, c_2 = 0.002, c_3 = 0.5, c_4 = 0.04. \quad (15)$$

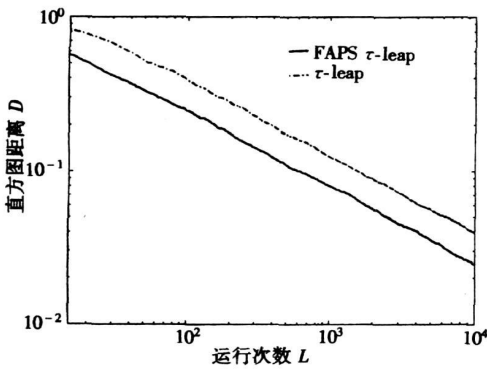
$$X_1(0) = 1\,000, X_2(0) = 1\,000, X_3(0) = 0. \quad (16)$$

我们运行 4×10^5 次 SSA, 每次从 $t_0 = 0$ 开始到 $t_{\text{final}} = 15$ 结束, 得到最后时刻 $t_{\text{final}} = 15$ 的直方图. 在到达最后时刻的前一步, 使用 FAPS 方法, 这里我们仅考虑 S_1 分子的数目 x_1 和 S_2 分子的数目 x_2 的所有可能的概率更新:

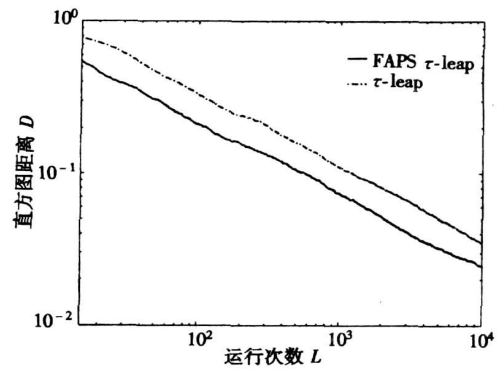
$$\begin{cases} q(x_1 - 1) \rightarrow q(x_1 - 1) + a_1/a_0, & q(x_1 - 2) \rightarrow q(x_1 - 2) + a_2/a_0, \\ q(x_1) \rightarrow q(x_1) + a_4/a_0, & q(x_1 + 2) \rightarrow q(x_1 + 2) + a_3/a_0, \\ q(x_2 - 1) \rightarrow q(x_2 - 1) + (a_3 + a_4)/a_0, & q(x_2) \rightarrow q(x_2) + a_1/a_0, \\ q(x_2 + 1) \rightarrow q(x_2 + 1) + a_2/a_0. \end{cases} \quad (17)$$

最后,我们计算 SSA 和 leap 算法和 FAPS leap 算法之间的直方图距离.

图 3(a) 和图 3(b) 分别表示 $\varepsilon = 0.03$ 时 $X_1(15)$ 和 $X_2(15)$ 的直方图距离对应运行次数, 其中的 τ 均由 Gillespie 和 Petzold 的参考文献[10] 文中的方程(6) 来确定. 可以看出, 对 1 个特定的精度, τ -leap 算法比 FAPS τ -leap 算法对 S_1 大约需要将近 3 倍的模拟次数, 对 S_2 大约需要 2 倍多的运行次数. 例如, 对于直方图距离 $D = 0.1$ 时的分子 $S_1(S_2)$, 使用 τ -leap 算法平均需要 1 650(1 151) 次模拟, 而 FAPS τ -leap 算法只需 665(533) 次模拟就足够了. 这均是 2 倍多的提高. 因为 S_1 有可能不同的发生概率的 4 个可能出现的状态, S_2 有可能不同的发生概率的 3 个可能出现的状态. 表 2 是对参数 ε 的不同值时使用 τ -leap 算法和 FAPS τ -leap 算法所得的 $X_1(15)$ 的直方图距离, 平均需要的 CPU 时间和运行次数. 其中的 τ 均由 Gillespie 和 Petzold 的参考文献[10] 文中的方程(6) 来确定. 我们可以看出相同的直方图距离时, 使用 FAPS τ -leap 算法大大地节约了 CPU 时间. 因此, FAPS τ -leap 算法大大地提高了模拟效率.



(a) $X_1(15)$ (分子 S_1)



(b) $X_2(15)$ (分子 S_2)

图 3 $\varepsilon = 0.03$ 时得到的 $X_1(15)$ 和 $X_2(15)$ 的直方图距离

3 结 论

SSA^[4,5], 加速的 SSA^[6,7] 和 leap 算法^[9,16] 都必须进行大量的模拟来获得一个较高的精度. 这里我们提出了一个修正的方法, 称作 FAPS 方法, 大大地减少了模拟次数. 该方法基于 SSA, 在达到最后时刻的前一步应用所有可能发生的反应的概率. 而且, 由于 FAPS 方法仅在最后一步有变化, 我们可以将其结合到 leap 方法中, 得到 FAPS leap 算法. 对两个不同的反应集的模拟测试验证了我们的新算法大大地提高了模拟效率和精度.

由于在 FAPS (leap) 算法中要求使用 SSA 或 leap 方法模拟的次数随着反应集中反应物种的所有可能的状态的数目的增加而减少, 由我们的新算法所获得的效率提高对于物种有多个可能的状态的系统将是最明显的.

FAPS (leap) 方法可以推广到获得在模拟时间间隔内任何时刻的信息. 为此, 我们仅需要在此时刻应用 FAPS (leap) 算法, 而不改变其他的模拟过程. 故我们的 FAPS (leap) 方法用较少的模拟次数得到所有物种在任何时刻的理想统计性质. FAPS (leap) 方法也可以用于有延滞的随机模拟算法, 我们将在另一篇文章中进行讨论. 进一步的研究是要建立 FAPS (leap) 方法的理论.

[参 考 文 献]

- [1] McAdams H H, Arkin A. Stochastic mechanisms in gene expression[J]. Proc Natl Acad Sci U S A, 1997, **94**(3): 814-819.
- [2] Arkin A, Ross J, McAdams H H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ infected E coli cells[J]. Genetics, 1998, **149**(4): 1633-1648.
- [3] Fedoroff N, Fontana W. Small numbers of large molecules[J]. Science, 2002, **297**(5584): 1129-1130.
- [4] Van Kampen N G. Stochastic Process in Physics and Chemistry [M]. Amsterdam: North-Holland, 1981.
- [5] Gillespie D T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions[J]. J Comput Phys, 1976, **22**(4): 403-434.
- [6] Gillespie D T. Exact stochastic simulation of coupled chemical reactions[J]. J Chem Phys, 1977, **81**(25): 2340-2361.
- [7] Gibson M, Bruck J. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems[J]. J Chem Phys, 2000, **104**(9): 1876-1889.
- [8] McCollum J M, Peterson G D, Cox C D, et al. The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior[J]. Comput Biol Chem, 2006, **30**(1): 39-49.
- [9] Gillespie D T. Approximate accelerated stochastic simulation of chemically reacting systems[J]. J Chem Phys, 2001, **115**(4): 1716-1733.
- [10] Gillespie D T, Petzold L R. Improved leap-size selection for accelerated stochastic simulation[J]. J Chem Phys, 2003, **119**(16): 8229-8234.
- [11] Rathinam M, Petzold L R, Cao Y, et al. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method[J]. J Chem Phys, 2003, **119**(24): 12784-12794.
- [12] Tian T, Burrage K. Binomial leap methods for simulating stochastic chemical kinetics[J]. J Chem Phys, 2004, **121**(21): 10356-10364.
- [13] Chatterjee A, Vlachos D G, Katsoulakis M A. Binomial distribution based τ -leap accelerated stochastic simulation[J]. J Chem Phys, 2005, **122**(2): 024112.
- [14] Peng X J, Zhou W, Wang Y F. Efficient binomial leap method for simulating chemical kinetics[J]. J Chem Phys, 2007, **126**(22): 1-9.
- [15] Cai X D, Xu Z Y. K-leap method for accelerating stochastic simulation of coupled chemical reactions [J]. J Chem Phys, 2007, **126**(7): 1.
- [16] Auger A, Chatelain P, Koumoutsakos P. R-leaping: Accelerating the stochastic simulation algorithm by reaction leaps[J]. J Chem Phys, 2006, **125**(8): 084103.
- [17] Cao Y, Gillespie D T, Petzold L R. Efficient step size selection for the tau-leaping simulation method [J]. J Chem Phys, 2006, **124**(4): 044109.
- [18] Cao Y, Petzold L R. Accuracy limitations and the measurement of errors in the stochastic simulation of chemically reacting systems[J]. J Comput Phys, 2006, **212**(1): 6-24.
- [19] Cao Y, Li H, Petzold L R. Efficient exact stochastic simulation algorithm for chemically reacting systems[J]. J Comput Phys, 2004, **121**(9): 4059-4067.
- [20] Cao Y, Gillespie D T, Petzold L R. Avoiding negative populations in explicit poisson tau-leaping[J]. J Chem Phys, 2005, **123**(5): 054104.
- [21] Lipshtat A. "All possible steps" approach to the accelerated use of Gillespie's algorithm[J]. J Chem Phys, 2007, **126**(18): 184103.

“Final All Possible Steps” Approach for Accelerating Stochastic Simulation of Coupled Chemical Reactions

ZHUO Wen^{1,2}, PENG Xi-jun¹, LIU Xiang¹,
YAN Zheng-lou¹, WANG Yi-fei¹

(1. Department of Mathematics, Shanghai University,

Shanghai 200444, P. R. China;

2. Department of Mathematics, Anhui Normal University, Wuhu 241000, P. R. China)

Abstract: A modified accelerated stochastic simulation method for chemically reacting systems, called the “final all possible steps” (FAPS) method is developed. The reliable statistics of all species were obtained in any time during the time course with fewer simulation times. Moreover, the FAPS method can be incorporated into the leap methods, which makes the simulation of larger systems more efficient. Numerical results indicate that the proposed methods can be applied to a wide range of chemically reacting systems with a high-precision level and a significant improvement can be obtained on efficiency over the existed methods.

Key words: “final all possible steps” approach; stochastic simulation algorithm; chemically reacting system