

化学反应随机动力学的数值模拟

胡煜成

导师 李铁军

合作者 闵斌, Assyr Abdulle

博士论文答辩

2010年5月28日

<http://dsec.pku.edu.cn/~huyc/>

I. 背景介绍

- 化学反应系统及其数值模拟

II. Tau-leaping 算法

- 误差分析
- 高精度格式
- 刚性问题

III. 多尺度系统

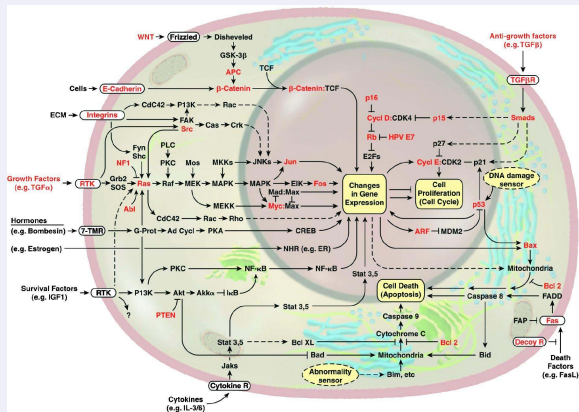
- Boosted 混合算法

IV. 应用

- 模拟癌症动力学

我们的问题是什么？

我们关心化学反应系统随时间的演化规律



- 随机性
- 大规模
- 多尺度

研究背景和动机

能否从化学反应规律出发理解生命活动的规律？



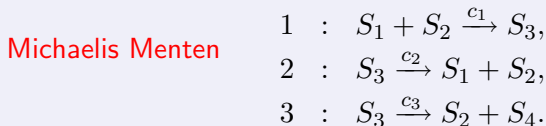
- 美国的CBB实验室 (<https://bioinformatics.cs.vt.edu/>)
- 日本的E-Cell计划 (<http://www.e-cell.org/>)
- 瑞士的SIB研究所 (<http://www.isb-sib.ch/>)

- Endy, D. and Brent, R., *Modelling cellular behavior*, Nature, 2001.
- M. Elowitz et al. *Stochastic gene expression in a single cell*, Science, 2002.
- N. Guido et al. *A bottom-up approach to gene regulation*, Nature, 2006.

潜在应用: 药物开发

化学反应系统

假设一个化学反应系统中有 N 种反应物和 M 个反应:



- 分子数 $\mathbf{X}(t) = (X_1(t), \dots, X_N(t))$
- 反应速率函数 $a_j(\mathbf{x})$, 表征反应发生的快慢.
- 状态改变向量 $\boldsymbol{\nu}_j$, $\boldsymbol{\nu}_1 = (-1, -1, 1, 0)$.
- 在一个时间微分 dt 内, 反应 j 发生的概率为 $a_j(\mathbf{X}(t))dt$; 一旦第 j 个反应发生, 状态改变 $\mathbf{X}(t) \rightarrow \mathbf{X}(t) + \boldsymbol{\nu}_j$.
- $\mathbf{X}(t)$ 是一个 Markov 跳过程.

SSA: Stochastic Simulation Algorithm

化学反应主方程 (CME)

$$\frac{\partial}{\partial t} p(\mathbf{x}, t) = \sum_{j=1}^M a_j(\mathbf{x} - \boldsymbol{\nu}_j) p(\mathbf{x} - \boldsymbol{\nu}_j, t) - \sum_{j=1}^M a_j(\mathbf{x}) p(\mathbf{x}, t).$$

精确格式: SSA (Gillespie, 1976)

- 1 给定 t 时刻的状态 $\mathbf{X}(t)$, 计算反应速率函数 $a_j(\mathbf{X}(t))$ 以及 $a_0 = \sum_{j=1}^M a_j(\mathbf{X}(t))$.
- 2 生成参数为 a_0 的指数分布随机变量 τ , 为下一个反应的等待时间.
- 3 生成 $[0, 1]$ 区间上均匀分布随机变量 u , 找到 k 使得,

$$\sum_{j=1}^{k-1} a_j(\mathbf{X}(t)) \leq ua_0 < \sum_{j=1}^k a_j(\mathbf{X}(t)).$$

- 4 更新时间 and 系统状态 $\mathbf{X}(t + \tau) = \mathbf{X}(t) + \boldsymbol{\nu}_k$.

II. Tau-leaping 算法

- 误差分析

Yucheng Hu, Tiejun Li and Bin Min *Local truncation error analysis of tau-leaping methods: revisited*, submitted to Commun. Math. Sci.

- 高精度格式

Yucheng Hu and Tiejun Li, *Highly Accurate Tau-leaping Methods with Random Corrections*, J. Chem. Phys., 130, 12., 2009.

Yucheng Hu, Tiejun Li and Bin Min, *A Weak Second Order Tau-leaping Method for Chemical Reaction Systems*, In preparation.

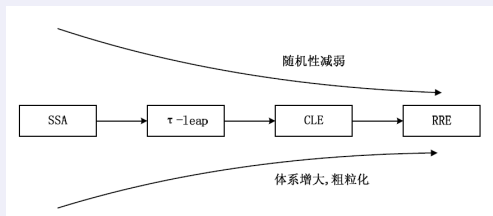
- 刚性问题

Assyr Abdulle, Yucheng Hu and Tiejun Li, *Chebyshev methods with discrete noise: the τ -ROCK methods*, J. Comp. Math., 28, 195-217, 2010.

泊松噪声驱动的 SDE

$$d\mathbf{X}(t) = \sum_{j=1}^M \nu_j \mathcal{P}_j(a_j(\mathbf{X}_{t-}) dt).$$

tau-leaping: $\mathbf{X}_{n+1} = \mathbf{X}_n + \sum_{j=1}^M \nu_j \mathcal{P}(a_j(\mathbf{X}_n) \tau).$



Tau-leaping 的收敛性 (Rathinam et al., 2005; T. Li, 2006)

- **强半阶收敛**: $\mathbb{E}|\mathbf{X}_n - \mathbf{X}_{t_n}|^2 \leq C\tau$, $\tau = \max \delta t$.
- **弱一阶收敛**: $|\mathbb{E}g(\mathbf{X}_n) - \mathbb{E}g(\mathbf{X}_{t_n})| \leq C\tau$, $\tau = \max \delta t$.

其它改进 tau-leaping 格式

- 中点 tau-leaping 格式 (Gillespie, 2001)

$$\begin{cases} \mathbf{Y} = \mathbf{X}_n + \frac{\tau}{2} \sum_{j=1}^M \nu_j a_j(\mathbf{X}_n), \\ \mathbf{X}_{n+1} = \mathbf{X}_n + \sum_{j=1}^M \nu_j \mathcal{P}(a_j(\mathbf{Y})\tau). \end{cases}$$

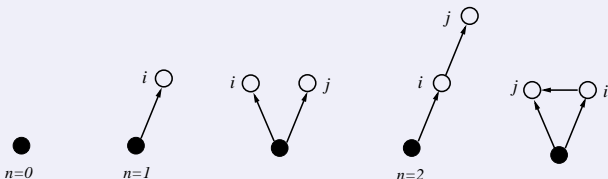
- 泊松 Runge-Kutta 格式 (Burrage and Tian, 2003)
- ...

弱 Taylor 展开

无穷小生成元: $\mathcal{L}g(\mathbf{x}) = \sum_{j=1}^M a_j(\mathbf{x}) \left(g(\mathbf{x} + \boldsymbol{\nu}_j) - g(\mathbf{x}) \right)$.

弱 Taylor 展开: $\mathbb{E}f(\mathbf{X}(t+h)) = \sum_{n=0}^{\infty} \frac{h^n}{n!} \mathbb{E}\mathcal{L}^n f(\mathbf{X}(t))$.

基于有向图的展开项对应 (ODE, 有根树, SDE, S-树)



$$\mathcal{L}^{(2)} f = a_i a_j f^{ij} + a_j a_i^j f^i + a_j a_i^j f^{ij}.$$

相容性分析

- 中点 tau-leaping 格式对于期望近似二阶相容.
- s -级泊松 Runge-Kutta 格式对期望最多 2 阶相容, 而对高阶矩最多 1 阶相容.

Yucheng Hu, Tiejun Li and Bin Min, submitted to Commun. Math. Sci.

是否能够设计高阶精度的格式?

- 随机校正格式, 对期望和方差二阶相容.

Yucheng Hu and Tiejun Li, J. Chem. Phys., 130, 12., 2009.

- 梯形校正格式, 对各阶矩二阶相容.

Yucheng Hu, Tiejun Li and Bin Min, In preparation.

随机校正格式

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \sum_{j=1}^M \boldsymbol{\nu}_j (r_j + \tilde{\mathbf{r}}_j).$$

高斯随机校正 tau-leaping

① 产生泊松随机变量 $r_j = \mathcal{P}(a_j(\mathbf{X}_n)\tau)$.

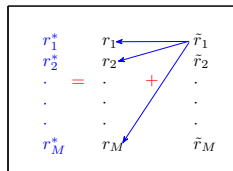
② 产生高斯随机变量 $\tilde{\mathbf{r}}_j$,

$$\mathbb{E}_{\mathbf{r}}[\tilde{\mathbf{r}}_j] = \frac{\tau}{2} \sum_{k=1}^M r_k \boldsymbol{\eta}_{jk} + \frac{\tau}{2} \sum_{\boldsymbol{\eta}_{jk} < 0} \left(\frac{a_k}{a_j} r_j - \tau a_k \right) \boldsymbol{\eta}_{jk},$$

$$\text{Var}_{\mathbf{r}}[\tilde{\mathbf{r}}_j] = \frac{\tau^2}{2} \sum_{k=1}^M a_k |\boldsymbol{\eta}_{jk}|;$$

③ 更新时间 $t_n + \tau$ 和系统状态

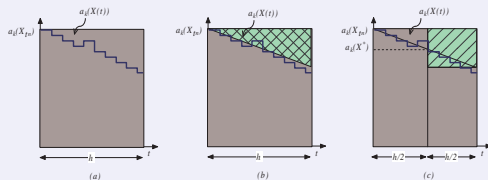
$$\mathbf{X}_{n+1} = \mathbf{X}_n + \boldsymbol{\nu} \cdot (\mathbf{r} + \tilde{\mathbf{r}}).$$



对期望和方差
二阶相容。

梯形校正格式

借鉴了 D. Anderson and J. Mattingly, *A weak trapezoidal method for a class of stochastic differential equations*, 2009.

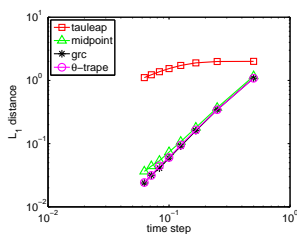


梯形校正 tau-leaping

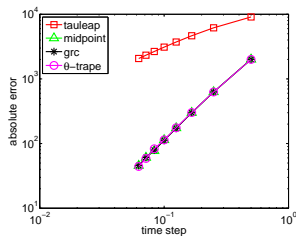
对于固定的 $\theta \in (0, 1)$, 令 $\alpha_1 = \frac{1}{2(1-\theta)\theta}$, $\alpha_2 = \frac{(1-\theta^2)+\theta^2}{2(1-\theta)\theta}$. 给定 \mathbf{X}_n , 按如下方式计算 \mathbf{X}_{n+1} ,

- $$\mathbf{X}^* = \mathbf{X}_n + \sum_{j=1}^M \nu_j \mathcal{P}(a_j(\mathbf{X}_n)\theta\tau),$$
- $$\mathbf{X}_{n+1} = \mathbf{X}^* + \sum_{j=1}^M \nu_j \mathcal{P}\left(\left(\alpha_1 a_j(\mathbf{X}^*) - \alpha_2 a_j(\mathbf{X}_n)\right)^+ (1-\theta)\tau\right).$$

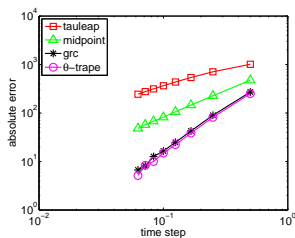
数值比较— 系统 1: $S \rightarrow 2S$



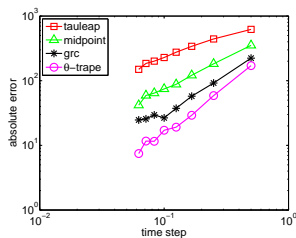
(a) 分布



(b) 期望

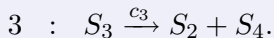
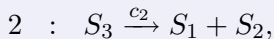
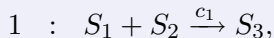


(c) 方差



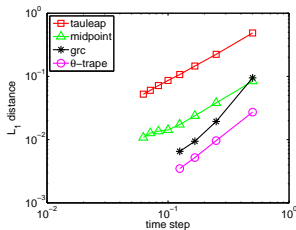
(d) 三阶中心矩

Michaelis-Menten

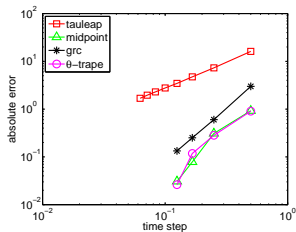


- $\mathbf{c} = (1 \times 10^{-4}, 0.5, 0.5)$.
- $\mathbf{a}(\mathbf{x}) = (c_1 x_1 x_2, c_2 x_3, c_3 x_3)$.
- $\mathbf{X}_0 = (1000, 200, 2000, 0)$.
- 模拟的时间区间为 $[0, 6]$.

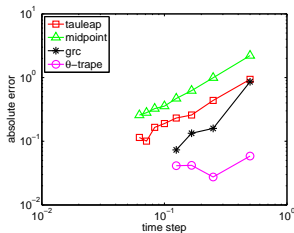
数值比较— 系统 2, X_1



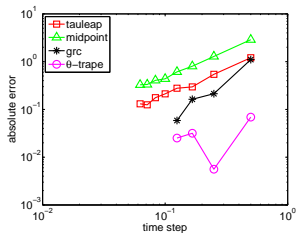
(e) 分布



(f) 期望

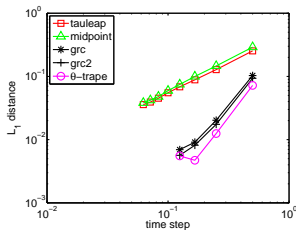


(g) 方差

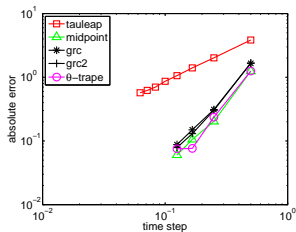


(h) 四阶中心矩

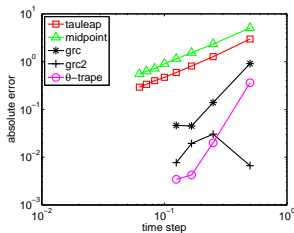
数值比较— 系统 2, X_2



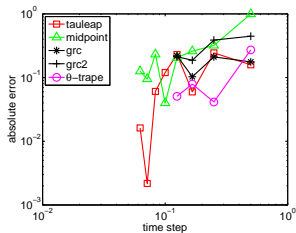
(i) 分布



(j) 期望



(k) 方差



(l) 三阶中心矩

刚性问题与隐式 tau-leaping

$$\begin{aligned}d\mathbf{X}(t) &= \sum_{j=1}^M \nu_j a_j(\mathbf{X}(t))dt + \sum_{j=1}^M \nu_j \left(\mathcal{P}(a_j(\mathbf{X}(t))dt) - a_j(\mathbf{X}(t))dt \right) \\ &= \mathbf{f}(\mathbf{X}(t))dt + d\mathbf{Q}(t).\end{aligned}$$

隐式 tau-leaping (Rathinam et al., 2003)

- ① 给定 t_n 时刻的状态 $\mathbf{X}(t_n)$, 计算反应速率函数 $a_j(\mathbf{X}(t_n))$.
- ② 产生泊松分布随机变量 $\mathcal{P}(a_j(\mathbf{X}_n)\tau), j = 1, 2, \dots, M$ 并计算

$$\mathbf{Q}(\mathbf{X}_n, \tau) = \sum_{j=1}^M \nu_j \left(\mathcal{P}(a_j(\mathbf{X}_n)\tau) - a_j(\mathbf{X}_n)\tau \right)$$

- ③ 求解隐式方程

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \mathbf{f}(\mathbf{X}_{n+1})\tau + \mathbf{Q}(\mathbf{X}_n, \tau).$$

ODE : Lebedev, 1994; Abdulle and Medovikov, 2001.

SDE : Abdulle and Cirilli, 2008; Abdulle and Li, 2008.

Tau-ROCK (逆向 tau-ROCK)

给定 t_n 时刻的状态 $\mathbf{X}(t_n)$, 选定一个比较大的步长 δt .

迭代的计算 \mathbf{K}_m :

$$\mathbf{K}_0 = \mathbf{X}_n, \quad \mathbf{K}_1 = \mathbf{K}_0 + \tau \frac{\omega_1}{\omega_0} \mathbf{f}(\mathbf{K}_0),$$

$$\mathbf{K}_j = 2\tau\omega_1 \frac{T_{j-1}(\omega_0)}{T_j(\omega_0)} \mathbf{f}(\mathbf{K}_{j-1}) + 2\omega_0 \frac{T_{j-1}(\omega_0)}{T_j(\omega_0)} \mathbf{K}_{j-1} - \frac{T_{j-2}(\omega_0)}{T_j(\omega_0)} \mathbf{K}_{j-2},$$

$$j = 2, \dots, m-1,$$

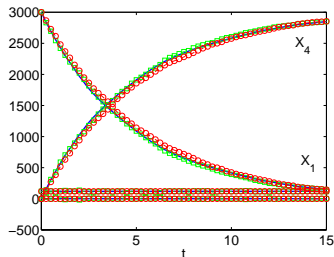
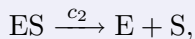
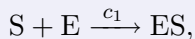
$$\mathbf{K}_m = 2\tau\omega_1 \frac{T_{m-1}(\omega_0)}{T_m(\omega_0)} \mathbf{f}(\mathbf{K}_{m-1}) + 2\omega_0 \frac{T_{m-1}(\omega_0)}{T_m(\omega_0)} \mathbf{K}_{m-1} - \frac{T_{m-2}(\omega_0)}{T_m(\omega_0)} \mathbf{K}_{m-2}$$

$$+ \sum_{j=1}^M \nu_j \left(\mathcal{P}(a_j(\mathbf{K}_{m-1})\tau) - a_j(\mathbf{K}_{m-1})\tau \right).$$

更新系统时间 $t_n + \tau$ 和状态 $\mathbf{X}_{n+1} = \mathbf{K}_m$. (Assyr Abdulle, Yucheng Hu and Tiejun Li, J. Comp. Math., 2010.)

数值结果

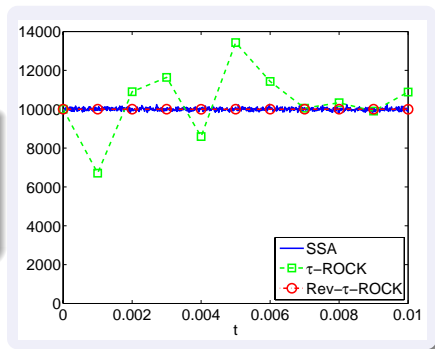
系统 1



	tau-leaping		tau-ROCK		
	δt	随机变量数目	$\tau_{eff,d}$	$\tau_{eff,s}$	随机变量数目
$c_3 = 10$	0.15	300	0.0833	0.25	180
$c_3 = 100$	0.05	9000	0.0313	0.25	180
$c_3 = 10^3$	10^{-5}	4.5×10^6	0.0086	0.25	180
$c_3 = 10^4$	2.5×10^{-6}	1.8×10^7	0.0024	0.25	180

数值结果

系统 2



	SSA	tau-ROCK	逆向 tau-ROCK
$X_1(0.01)$	72.95	1422.20	23.70
$X_2(0.01)$	2.22	2.18	2.19

III. 多尺度问题

- 系统划分
- Boosting
- 混合算法
- Boosted 混合算法

Yucheng Hu, Assyr Abdulle and Tiejun Li, *Boosted Hybrid Method for Solving Chemical Reaction System with Multiscale in Time and Population*, In preparation.

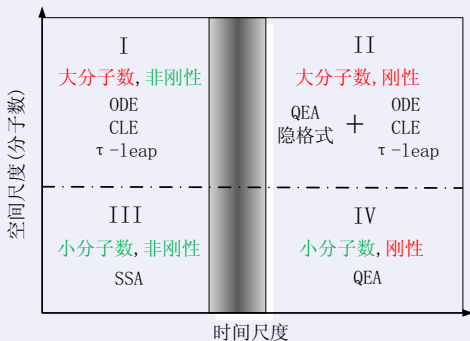
多尺度问题

反应速率

慢反应: 秒, 分钟
快反应: 毫秒, 微秒

反应物分子数

DNA, RNA: 1, 10
蛋白质: 10^3
小分子: 10^9



混合算法: 改进了 Haseltine & Rawlings 的算法 (JCP, 2002).

拟平衡态近似: Slow-scale SSA (Cao, Petzold and Gillespie, JCP, 2005),
nested-SSA (E Weinan, Liu, D and Vanden-Eijnden, JCP, 2005), boosting.

$$\begin{cases} \frac{dx}{dt} = f(x, y), \\ \frac{dy}{dt} = \frac{1}{\epsilon}(y - \phi(x)). \end{cases} \implies \frac{dx}{dt} = f(x, \phi(x)) \triangleq F(x).$$

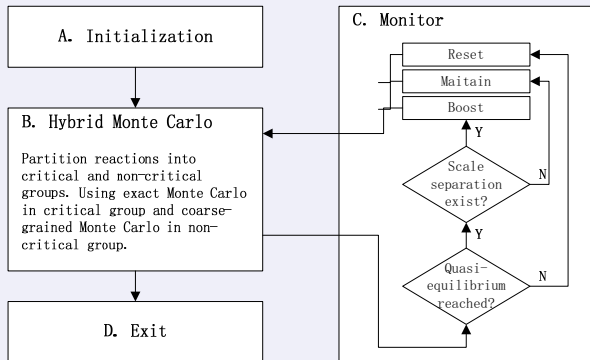
Boosting (Eric Vanden-Eijnden, Comm. Math. Sci., 2007):
 $\epsilon = 10^{-6} \rightarrow \epsilon = 10^{-3}$

$$\begin{aligned} S_1 &\rightleftharpoons S_2, \\ S_1 &\rightleftharpoons S_3. \end{aligned} \quad c'_j = \kappa c_j, \quad j \text{ 为快反应}, \quad \kappa \leq 1.$$

$$\frac{1}{\tau} \int_0^\tau a_j(\mathbf{X}(s)) ds \approx \frac{1}{T} \int_0^T a_j(\mathbf{X}(s)) ds = \frac{1}{\lambda T} \int_0^{\lambda T} a_j(\mathbf{X}(s/\lambda)) ds$$

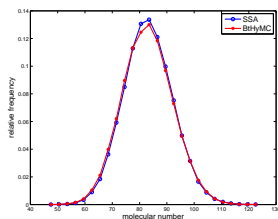
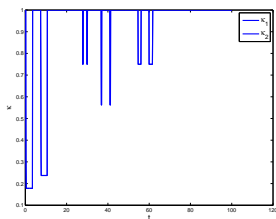
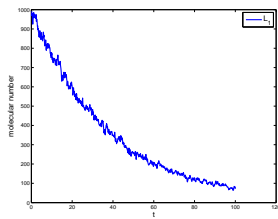
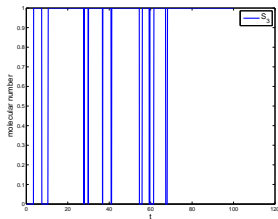
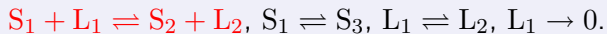
跟 Slow-scale SSA 和 Nested SSA 相比, Boosting **实现更方便**.

Boosted 混合算法

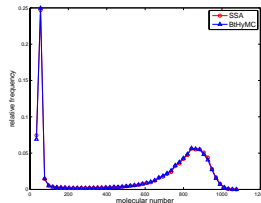
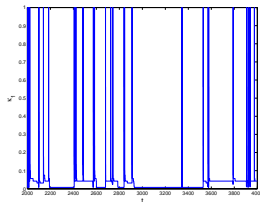
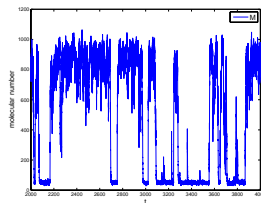
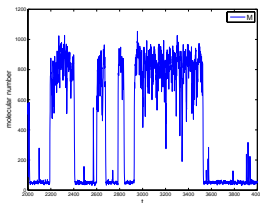
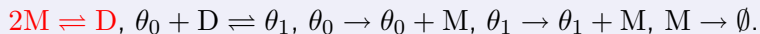


难点: 判断尺度分离和拟平衡

数值结果— 系统 1



数值结果— 系统 2

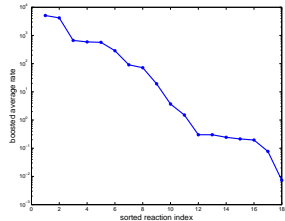
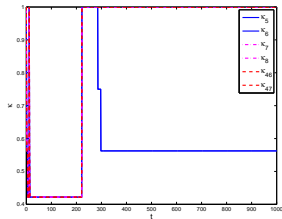
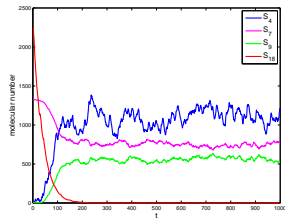
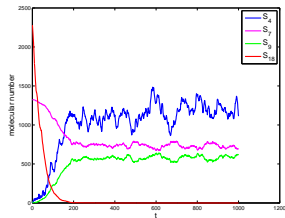


SSA: 2 个小时, Boosted 混合算法: 23s.

数值结果— 复杂系统

$$\begin{aligned} S_1 + S_2 &\stackrel{c_1}{=} S_3 \\ S_1 + S_4 &\stackrel{c_2}{=} S_5 \\ S_1 + S_6 &\stackrel{c_3}{=} S_7 \\ S_4 + S_{14} &\stackrel{c_4}{=} S_{15} \\ S_{14} + S_{16} &\stackrel{c_5}{=} S_{17} \\ S_3 + S_6 &\stackrel{c_6}{=} S_8 \\ S_5 + S_6 &\stackrel{c_7}{=} S_9 \\ S_3 + S_{10} &\stackrel{c_8}{=} S_{12} \\ S_5 + S_{11} &\stackrel{c_9}{=} S_{13} \\ S_{15} + S_{18} &\stackrel{c_{10}}{=} S_{19} \\ \emptyset &\stackrel{c_{11}^{s_{13}}}{=} S_{22} \\ \emptyset &\stackrel{c_{12}^{s_{22}}}{=} S_{14} \\ S_{17} &\stackrel{c_{13}}{=} S_{16} \\ S_{15} &\stackrel{c_{14}}{=} S_4 \\ S_{19} &\stackrel{c_{15}}{=} S_4 + S_{18} \\ S_{28} &\stackrel{c_{16}}{=} S_4 + S_{27} \\ \emptyset &\stackrel{c_{17}^{s_{13}}}{=} S_{23} \\ \emptyset &\stackrel{c_{18}^{s_{23}}}{=} S_{18} \\ S_{19} &\stackrel{c_{19}}{=} S_{15} \\ \emptyset &\stackrel{c_{20}^{s_{12}}}{=} S_{25} \\ \emptyset &\stackrel{c_{21}^{s_{25}}}{=} S_4 \\ S_{19} &\stackrel{c_{22}}{=} S_{14} + S_{18} \\ S_{21} &\stackrel{c_{23}}{=} S_{20} \\ S_{28} &\stackrel{c_{24}}{=} S_{14} + S_{27} \\ \emptyset &\stackrel{c_{25}^{s_{13}}}{=} S_{24} \\ \emptyset &\stackrel{c_{26}^{s_{24}}}{=} S_{20} \\ S_{21} &\stackrel{c_{27}}{=} S_4 \\ S_4 + S_{20} &\stackrel{c_{28}}{=} S_{21} \\ \emptyset &\stackrel{c_{29}^{s_{13}}}{=} S_{26} \\ \emptyset &\stackrel{c_{30}^{s_{26}}}{=} S_{27} \\ S_{28} &\stackrel{c_{31}}{=} S_{15} \\ S_{15} + S_{27} &\stackrel{c_{32}}{=} S_{28} \\ S_5 &\stackrel{c_{33}}{=} S_1 \\ S_{13} &\stackrel{c_{34}}{=} S_1 + S_{11} \\ S_9 &\stackrel{c_{35}}{=} S_7 \\ S_{15} &\stackrel{c_{36}}{=} S_{14} \\ S_{19} &\stackrel{c_{37}}{=} S_{14} + S_{18} \\ S_{20} &\stackrel{c_{38}}{=} S_{14} + S_{27} \\ S_{21} &\stackrel{c_{39}}{=} S_{20} \end{aligned}$$

数值结果— 复杂系统



待解决的问题

$$\emptyset \rightarrow A,$$

$$A + E \rightarrow EA,$$

$$EA \rightarrow B + E,$$

$$B \rightarrow \emptyset.$$

$$\emptyset \rightarrow A,$$

$$A \rightarrow B,$$

$$B \rightarrow \emptyset.$$

Yucheng Hu, Assyr Abdulle and Tiejun Li, In preparation.

IV. 癌症的群体演化动力学

合作者: 朱天琪, 李铁军, 杨子恒, 马志明, 张德兴

- 背景介绍
- 建立化学反应的模型
- 运用混合算法模拟

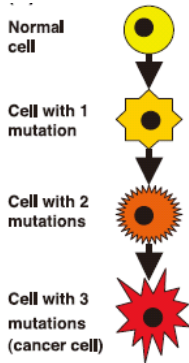
Tianqi Zhu, Yucheng Hu , Ziheng Yang, Dexing Zhang, Tiejun Li and Zhiming Ma, *Efficient Simulation under a Population Genetics Model of Carcinogenesis*, In preparation.

应用：癌症的群体演化动力学

背景

- 癌症是一种**基因**疾病。正常细胞的生长受机体的调控，而癌细胞则会无限复制。
- **大量** ($10^6 \sim 10^9$) 细胞群体长时间 (几十年) 的演化结果。
- 演化过程由**突变**和**选择**驱动，跟达尔文的群体演化过程类似。

- Beerenwinkel et al., *Genetic progression and the waiting time to cancer*, PLoS Comput. Biol., 2007.
- Durrett et al., *A waiting time problem arising from the study of multi-stage carcinogenesis*, Ann. Appl. Prob., 2009.



Ashkenazi et al., Neoplasia Vol. 10, No. 11, 2008.

建立模型

演化模型 (Moran 模型)

考虑 N 个细胞组成的群体, 按照突变数目将细胞分为 0-类, 1-类, \dots , m -类, 初始时全为 0-类细胞.

细胞存活参数为 1 的指数分布时间, 然后被新细胞取代, 新细胞的类型从当前的群体中以选择系数 s_i 为权重随机生成.

i -类细胞以速率 μ_i 突变为 $(i+1)$ -类.

当群体中出现第一个 m 类细胞时癌症产生.

化学反应模型的描述

系统状态:

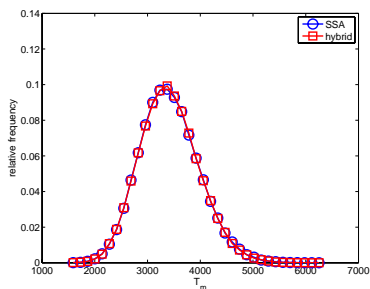
$\mathbf{X}(t) = \{x_0, x_1, \dots, x_m\}$, x_i 表示 i -类细胞的数目,
 $\mathbf{X}(0) = \{N, 0, \dots, 0\}$

反应: j -类细胞被 j' -类细胞替换,
 $a_{jj'} = x_j s_{j'} x_{j'} / \sum_{l=0}^m s_l x_l$,
 $\nu_{jj'} = \xi_{j'} - \xi_j$.

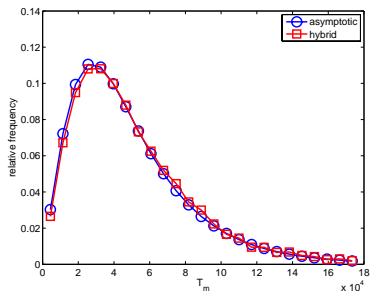
反应: i -类细胞突变, $a_i = \mu_i x_i$,
 $\nu_i = \xi_{i+1} - \xi_i$.

$T_m = \inf_t \{x_m(t) > 0\}$.

数值结果: 精度



(m)



(n)

等待时间 T_m 的统计直方图比较, 左: $N = 10^3$, 右: $N = 10^5$.

数值结果: 效率

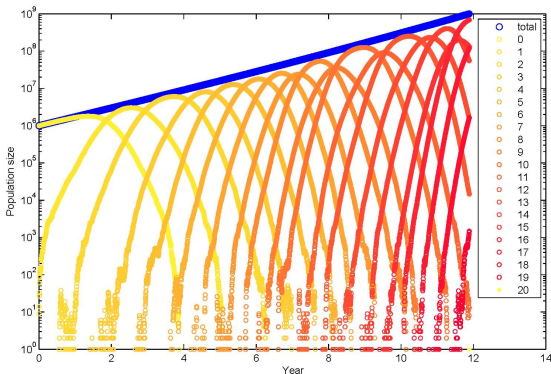
消耗时间

	$N = 10^3$	$N = 10^4$	$N = 10^5$	$N = 10^6$
SSA (sec.)	6	111	1595	19414
Hybrid (sec.)	5	19	46	83

平均步长

	$N = 10^3$	$N = 10^4$	$N = 10^5$	$N = 10^6$
SSA	3.13E-3	3.67E-4	4.46E-5	5.45E-6
Hybrid	0.0108	0.0104	0.0102	0.0101

数值结果



具有选择优势的突变类型逐渐取代现有类型，群体中类型的比例波浪式的随时间推进。

总结

- 发展一套对 tau-leaping 方法的**误差分析工具**.
- 设计了**高斯随机校正**和**梯形随机校正**两类高精度格式.
- 尝试了用 **Chebyshev 方法**求解**刚性**问题.
- 开发了具有**自适应**特点的 **Boosted 混合算法**, 能高效求解部分多尺度问题.
- 用混合算法高效模拟了**大规模的癌症群体演化动力学模型**.

谢谢!

论文和 ppt 可以从 <http://dsec.pku.edu.cn/~huyc/> 上下载.