Simulation methods for stochastically modeled reaction networks

David F. Anderson

anderson@math.wisc.edu
Department of Mathematics
University of Wisconsin - Madison

The Isaac Newton Institute, Cambridge

March 23rd, 2010
Overview

1. Notation and terminology: biochemical reaction networks.

2. Discuss one type of numerical approximation method: leaping algorithms.


4. Discuss diffusion approximations
   ▶ Accuracy in approximating the jump process.
   ▶ Methods of approximating the diffusion approximation. (Joint work with Jonathan Mattingly)
   ▶ Structural issues.
Stochastic models of (bio)chemical reactions

- Standard notation for chemical reactions:

\[ A + B \rightarrow C \]

is interpreted as “a molecule of A combines with a molecule of B to give a molecule of C.”

- Each instance of the reaction \( A + B \rightarrow C \) changes the state of the system by the vector:

\[
\nu'_k - \nu_k = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}.
\]

1. Other types of reactions:

\[ A \rightarrow B, \quad 2A \rightarrow B, \quad B \rightarrow 2A, \quad \emptyset \rightarrow A, \quad A + B + C \rightarrow * \]
Markov chain models

- We consider a network of reactions involving $d$ chemical species, $S_1, \ldots, S_d$:
  \[
  \sum_{i=1}^{d} \nu_{ik} S_i \rightarrow \sum_{i=1}^{d} \nu'_{ik} S_i
  \]

- The state of the system, $X(t) \in \mathbb{Z}^{d}_{\geq 0}$, gives the number of molecules of each species in the system at time $t$.

- $\nu_k$: number of molecules of each chemical species consumed in the $k$th reaction.

- $\nu'_k$: number of molecules of each chemical species created in the $k$th reaction.

- If $k$th reaction occurs at time $t$, the new state becomes
  \[
  X(t) = X(t^-) + \nu'_k - \nu_k.
  \]

- The waiting times for the reactions are exponentially distributed with rate (intensity/propensity) functions $\lambda_k : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$. 
Markov chain models

This model is a continuous time Markov chain in $\mathbb{Z}_{\geq 0}^d$ with generator

$$(Af)(x) = \sum_k \lambda_k(x)(f(x + \nu'_k - \nu_k) - f(x)).$$

Kolmogorov’s forward equation ("Chemical Master Equation") describes the evolution of the distribution of the state of the system

$$\frac{d}{dt} P(x, t) = \sum_k \lambda_k(x - \nu'_k + \nu_k) P(x - \nu'_k + \nu_k, t) - \sum_k \lambda_k(x) P(x, t),$$

where $P(x, t)$ is probability $X(t) = x$. 
The number of times the $k$th reaction occurs by time $t$ is a counting process with intensity $\lambda_k(X(t))$ and can be written as

$$R_k(t) = Y_k \left( \int_0^t \lambda_k(X(s))ds \right),$$

where the $Y_k$ are independent, unit-rate Poisson process.

This representation is equivalent to the "chemical master equation" (Kolmogorov's forward equation) representation found in biology/chemistry literature.
The number of times the $k$th reaction occurs by time $t$ is a counting process with intensity $\lambda_k(X(t))$ and can be written as

$$R_k(t) = Y_k \left( \int_0^t \lambda_k(X(s)) ds \right),$$

where the $Y_k$ are independent, unit-rate Poisson process.

Therefore, a path-wise representation is given by random time-changes of Poisson processes

$$X(t) = X(0) + \sum_k R_k(t)(\nu'_k - \nu_k)$$

$$= X(0) + \sum_k Y_k \left( \int_0^t \lambda_k(X(s)) ds \right) (\nu'_k - \nu_k).$$

This representation is equivalent to the “chemical master equation” (Kolmogorov’s forward equation) representation found in biology/chemistry literature.
Mass-action kinetics

The standard intensity function chosen is mass-action kinetics:

$$\lambda_k(x) = \kappa_k \left( \prod_i \nu_{ik}! \right) \left( \frac{x}{\nu_k} \right) = \kappa_k \prod_i \frac{x_i!}{(x_i - \nu_{ik})!}.$$  

Rate is proportional to the number of distinct subsets of the molecules present that can form inputs for the reaction. (This assumes vessel is “well-stirred.”)

Example: If $S_1 \rightarrow$ anything, then $\lambda_k(x) = \kappa_k x_1$.

Example: If $S_1 + S_2 \rightarrow$ anything, then $\lambda_k(x) = \kappa_k x_1 x_2$.

Example: If $S_1 + 2S_2 \rightarrow$ anything, then $\lambda_k(x) = \kappa_k x_1 x_2(x_2 - 1)$.
Methods of investigation: numerical simulation

\[ X(t) = X(0) + \sum_{k} Y_k \left( \int_{0}^{t} \lambda_k(X(s)) \, ds \right) (\nu'_k - \nu_k), \]

(GOOD NEWS) There are a number of numerical methods that produce statistically exact sample paths:

1. Gillespie’s algorithm.
2. The next reaction method.
3. All just discrete event simulation.

For each time-step one must find:

(i) the amount of time that passes until the next reaction takes place: \( \Delta_n \).
   (the minimum of exponential RVs).

(ii) which reaction takes place at that time.
Methods of investigation: numerical simulation

\[ X(t) = X(0) + \sum_k Y_k \left( \int_0^t \lambda_k(X(s))ds \right) (\nu'_k - \nu_k), \]

(GOOD NEWS) There are a number of numerical methods that produce statistically exact sample paths:

1. Gillespie’s algorithm.
2. The next reaction method.
3. All just discrete event simulation.

For each time-step one must find:

(i) the amount of time that passes until the next reaction takes place: \( \Delta_n \).
   (the minimum of exponential RVs).

(ii) which reaction takes place at that time.

(BAD NEWS) If \( \sum_k \lambda_k(X(t)) \gg 1 \), then

\[ \mathbb{E}\Delta_n = \frac{1}{\sum_k \lambda_k(X(t))} \ll 1 \]

and the time needed to produce a single exact sample path over an interval \([0, T]\) can be prohibitive. Quite common.
Standard “τ-leaping” \(^1\) was developed by Dan Gillespie in an effort to overcome the problem that \(Δ_n\) may be prohibitively small.

**Tau-leaping** is essentially an Euler approximation of \(\int_0^t \lambda_k(X(s))ds\):

\[
Z(\tau) = Z(0) + \sum_k Y_k \left( \int_0^\tau \lambda_k(Z(s)) \, ds \right) (\nu'_k - \nu_k) \\
\approx Z(0) + \sum_k Y_k \left( \lambda_k(Z(0)) \tau \right) (\nu'_k - \nu_k) \\
\overset{d}{=} Z(0) + \sum_k \text{Poisson} \left( \lambda_k(Z(0)) \tau \right) (\nu'_k - \nu_k).
\]

Idea is to “leap” over many reactions.

---

**Tau-leaping**

**Algorithm:** For each time step do the following:

1. For each reaction, calculate $P_{k,n} = \text{Poisson}(\lambda_k(Z(t_n)) \tau)$.

2. Update, $Z(t_{n+1}) = Z(t_n) + \sum_k P_{k,n} \times (\nu'_k - \nu_k)$.

A pathwise representation for $Z(t)$ generated by $\tau$-leaping is

$$Z(t) = X(0) + \sum_k Y_k \left( \int_0^t \lambda_k(Z(\eta(s)))ds \right) (\nu'_k - \nu_k),$$

where

$$\eta(s) = t_n, \quad \text{if} \quad t_n \leq s < t_{n+1} = t_n + \tau$$

is a step function giving left endpoints of time discretization.
An “approximate midpoint” tau-leaping method

For a time discretization $0 = t_0 < t_1 < \cdots < t_N = T$, with $\tau = t_n - t_{n-1}$, let

$$\rho(z) = z + \frac{1}{2} \tau \sum_k \lambda_k(z)(\nu_k' - \nu_k),$$

be a “deterministic” midpoint approximation

Algorithm: For each time step do the following:

1. Let $y_n^* = \rho(Z(t_n))$.

2. For each reaction, calculate $P_{k,n} = \text{Poisson}(\lambda_k(y_n^*) \tau)$.

3. Update, $Z(t_{n+1}) = Z(t_n) + \sum_k P_{k,n} \times (\nu_k' - \nu_k)$.

Note: still just require one random variable per reaction per step.
An “approximate midpoint” tau-leaping method

For a time discretization $0 = t_0 < t_1 < \cdots < t_N = T$, with $\tau = t_n - t_{n-1}$, let

$$\rho(z) = z + \frac{1}{2} \tau \sum_k \lambda_k(z)(\nu'_k - \nu_k),$$

be a “deterministic” midpoint approximation.

**Algorithm:** For each time step do the following:

1. Let $y^*_n = \rho(Z(t_n))$.

2. For each reaction, calculate $P_{k,n} = \text{Poisson}(\lambda_k(y^*_n) \tau)$.

3. Update, $Z(t_{n+1}) = Z(t_n) + \sum_k P_{k,n} \times (\nu'_k - \nu_k)$.

**Note:** still just require one random variable per reaction per step.

Then $Z(t)$ solves:

$$Z(t) = X(0) + \sum_k \int_0^t \lambda_k(\rho(Z \circ \eta(s))) ds \left(\nu'_k - \nu_k\right),$$

where $\eta(s) = t_n$, if $t_n \leq s < t_{n+1} = t_n + \tau$. 
Previous error analysis

Under the scaling $\tau \to 0$:

1. Rathinam, Petzold, Cao, and Gillespie$^2$ showed, among other things (implicit methods, etc.), that tau-leaping is first order accurate in a weak sense if the intensity functions $\lambda_k$ are linear.

2. Li$^3$ extended this result by showing that standard tau-leaping has a strong error (in the $L^2$ norm) of order $1/2$ and a weak error of order one:

$$\sup_{n \leq N} \sqrt{\mathbb{E} |Z(t_n) - X(t_n)|^2} \leq C \tau^{1/2}$$

$$|\mathbb{E}f(Z(T)) - \mathbb{E}f(X(T))| \leq C \tau,$$

where $0 = t_0 < t_1 < \cdots < t_N = T$ is a partition of $[0, T]$.

3. The midpoint method is no more accurate than standard tau-leaping.

---


Example

Consider $A$ as a prey and $B$ as a predator.

\[ A \xrightarrow{\kappa_1} 2A, \quad A + B \xrightarrow{\kappa_2} 2B, \quad B \xrightarrow{\kappa_3} \emptyset, \]

with $A(0) = B(0) = 1000$ and $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.

Letting $\tau = 1/20$ and simulating 30,000 sample paths with each method yields the following approximate distributions for $B(10)$:
Problem with the $\tau \to 0$ scaling

Recall, tau-leaping methods are only useful if $\tau \gg \Delta_n$, for otherwise an exact method would be performed. Therefore, we should require that

$$\tau \gg \frac{1}{\sum_k \lambda_k(X(t))} \approx \Delta_n$$

while

$$\sum_k \lambda_k(X(t)) \gg 1.$$

So $\tau \to 0$ does not seem to be an appropriate scaling to describe the accuracy of these methods.
What is an appropriate scaling?

I can not answer in generality. It will depend upon how the system satisfies

\[ \sum_k \lambda_k(X(t)) \gg 1 \]

Multiple possibilities:

1. Could be that \( X_i = \mathcal{O}(V^{\alpha_i}) \) for \( V \gg 1 \) with \( \kappa_k = \mathcal{O}(1) \).

2. Could be that \( \kappa_k = \mathcal{O}(V^{\beta_k}) \) for \( V \gg 1 \) with \( X_i = \mathcal{O}(1) \).

3. Could be multiple scales.

4. Infinite number of possibilities.
The “classical scaling”
We will make the following natural assumptions on our model:

(i) $X_i(0) = O(V)$ for some $V \gg 1$.

(ii) $\lambda_k(X(t)) = O(V)$.
The “classical scaling”

We will make the following natural assumptions on our model:

(i) $X_i(0) = \mathcal{O}(V)$ for some $V \gg 1$.

(ii) $\lambda_k(X(t)) = \mathcal{O}(V)$.

(iii) The rate constants of the reactions must then satisfy:

\[
S_1 \rightarrow \ast, \quad S_1 + S_2 \rightarrow \ast, \quad S_1 + S_2 + S_3 \rightarrow \ast,
\]

scale with $V$ like:

\[
\kappa_k = \mathcal{O}(1), \quad \hat{\kappa}_k = \mathcal{O}\left(\frac{1}{V}\right), \quad \tilde{\kappa}_k = \mathcal{O}\left(\frac{1}{V^2}\right).
\]

For example: $A + B \rightarrow \ast$

\[
\mathcal{O}(V) = \lambda_k(X) = \kappa_k V^2 x_1 x_2 \Rightarrow \kappa_k = \mathcal{O}\left(\frac{1}{V}\right).
\]

(iv) This has the effect that $\lambda_k(Vu) = VA_k^V(u)$ for any $u$.  

\[\]

\[\]
The “classical scaling”

We will make the following natural assumptions on our model:

(i) \( X_i(0) = O(V) \) for some \( V \gg 1 \).

(ii) \( \lambda_k(X(t)) = O(V) \).

(iii) The rate constants of the reactions must then satisfy:

\[
S_1 \rightarrow \ast, \quad S_1 + S_2 \rightarrow \ast, \quad S_1 + S_2 + S_3 \rightarrow \ast,
\]

scale with \( V \) like:

\[
\kappa_k = O(1), \quad \bar{\kappa}_k = O\left(\frac{1}{V}\right), \quad \tilde{\kappa}_k = O\left(\frac{1}{V^2}\right).
\]

For example: \( A + B \rightarrow \ast \)

\[
O(V) = \lambda_k(X) = \kappa_k V^2 x_1 x_2 \Rightarrow \kappa_k = O\left(\frac{1}{V}\right).
\]

(iv) This has the effect that \( \lambda_k(Vu) = V A_k^V(u) \) for any \( u \).

In the limit as \( V \to \infty \), \( X(V, t)/V \to c(t) \), which satisfies deterministic mass action kinetics.
Important to recognize

1. It is clear that the choice of scaling laid out above will not explicitly cover all cases of interest:
   
   i. Multiple scales
   
   ii. Low copy numbers of constituent molecules.

2. However, the purpose is to give a more accurate picture of how tau-leaping methods approximate the exact solution, both strongly and weakly, in at least one plausible setting.
Q: Given a system satisfying our scaling, what size $\tau$ would a user choose?

Partial answer: something small, $\tau < 1$.

Complete answer: something not too small $\tau \gg 1$.

That is, for some $\beta \in (0, 1)$,

$$\tau = 1 V^\beta.$$ 

Note that $\tau X_k^\lambda(X(t)) = V^{-\beta} O(V) = O(V^{1-\beta}) \gg 1$,

which says that this time step will "leap" over many reactions.
Time-step

Q: Given a system satisfying our scaling, what size $\tau$ would a user choose?

Partial answer: something small, $\tau < 1$. 

Complete answer: something not too small $\tau \gg 1$.
Q: Given a system satisfying our scaling, what size \( \tau \) would a user choose?

Partial answer: something small, \( \tau < 1 \).

Complete answer: something not too small

\[
\tau \gg \frac{1}{\sum_k \lambda_k(X(t))} \approx \Delta_n = \mathcal{O}\left(\frac{1}{V}\right).
\]
Question: Given a system satisfying our scaling, what size $\tau$ would a user choose?

Partial answer: something small, $\tau < 1$.

Complete answer: something not too small

$$\tau \gg \frac{1}{\sum_k \lambda_k(X(t))} \approx \Delta_n = \mathcal{O}\left(\frac{1}{V}\right).$$

That is, for some $\beta \in (0, 1)$,

$$\tau = \frac{1}{V^{\beta}}.$$ 

Note that

$$\tau \sum_k \lambda(X(t)) = V^{-\beta} \mathcal{O}(V) = \mathcal{O}(V^{1-\beta}) \gg 1,$$

which says that this time step will “leap” over many reactions.
Error analysis on scaled processes

We consider the scaled processes

\[ X^V(t) \overset{\text{def}}{=} \frac{X(t)}{V} , \quad Z^V(t) \overset{\text{def}}{=} \frac{Z(t)}{V} , \quad \mathcal{Z}^V(t) \overset{\text{def}}{=} \frac{Z(t)}{V} . \]
Error analysis on scaled processes

We consider the scaled processes

\[ X^V(t) \overset{\text{def}}{=} \frac{X(t)}{V}, \quad Z^V(t) \overset{\text{def}}{=} \frac{Z(t)}{V}, \quad \mathcal{Z}^V(t) \overset{\text{def}}{=} \frac{\mathcal{Z}(t)}{V}. \]

which satisfy

\[
X^V(t) = X^V(0) + \frac{1}{V} \sum_k Y_k \left( V \int_0^t A_k^V(X^V(s)) ds \right) (\nu'_k - \nu_k)
\]

\[
Z^V(t) = X^V(0) + \frac{1}{V} \sum_k Y_k \left( V \int_0^t A_k^V(Z^V \circ \eta(s)) ds \right) (\nu'_k - \nu_k)
\]

\[
\mathcal{Z}^V(t) = X^V(0) + \frac{1}{V} \sum_k Y_k \left( V \int_0^t A_k^V(\rho^V(\mathcal{Z}^V \circ \eta(s))) ds \right) (\nu'_k - \nu_k),
\]

where \( \rho^V(z) \overset{\text{def}}{=} z + \frac{1}{2} \tau \sum_k A_k^V(z)(\nu'_k - \nu_k), \)
Note:

1. The purpose of generating $Z^V$ and $Z^V$ is to approximate $X^V$ in the case when $V \gg 1$.

2. We formalize this by
   - taking a limit as $V \to \infty$ and
   - consider the relationship of the processes $Z^V(t)$ and $Z^V(t)$ to the original process $X^V(t)$.
Strong error of standard tau-leaping

**Theorem (Anderson, Ganguly, Kurtz – 2010)**

Let $X^V(t)$ and $Z^V(t)$ denote the scaled processes for $t \in [0, T]$ with $\tau = V^{-\beta}$. There exists a constant $C = C(T) > 0$ such that

$$\sup_{t \leq T} \mathbb{E}|X^V(t) - Z^V(t)| \leq \frac{C}{V^\beta} = C\tau.$$
A new coupling of $X^V(t)$ and $Z^V(t)$

To sharpen the result, let

$$X^V(t) = X^V(0) + \frac{1}{V} \sum_k \left[ Y_{k,1} \left( V \int_0^t A^V_k(X^V(s)) \wedge A^V_k(Z^V \circ \eta(s)) \, ds \right) 
+ Y_{k,2} \left( V \int_0^t A^V_k(X^V(s)) - A^V_k(X^V(s)) \wedge A^V_k(Z^V \circ \eta(s)) \, ds \right) \right] (\nu'_k - \nu_k)$$

$$Z^V(t) = X^V(0) + \frac{1}{V} \sum_k \left[ Y_{k,1} \left( V \int_0^t A^V_k(X^V(s)) \wedge A^V_k(Z^V \circ \eta(s)) \, ds \right) 
+ Y_{k,3} \left( V \int_0^t A^V_k(Z^V \circ \eta(s)) - A^V_k(X^V(s)) \wedge A^V_k(Z^V \circ \eta(s)) \, ds \right) \right] (\nu'_k - \nu_k),$$

where the $Y_{k,i}$ are mutually independent Poisson processes.

Note: distribution of $X^V(t)$ and $Z^V(t)$ are unchanged.
Exact asymptotics for Euler tau-leaping

Theorem (Anderson, Ganguly, Kurtz – 2010)
Let $X^V(t)$ and $Z^V(t)$ satisfy the new coupling for $t \in [0, T]$. Then as $V \to \infty$

$$V^\beta (X^V - Z^V) \to \mathcal{E},$$

where

- $\mathcal{E}(t)$ is deterministic and independent of $V$ and $\beta$.
- And where $f^V \rightarrow f$ means uniform convergence on bounded intervals:

$$\lim_{V \to \infty} P\{ \sup_{t \leq T} |f^V(t) - f(t)| > \varepsilon \} = 0,$$

$$\lim_{V \to \infty} \mathbb{E} \sup_{t \leq T} \left| f^V(t) - f(t) \right| = 0.$$
Idea of proof

Can show:

\[ X^V(t) - Z^V(t) \approx M^V(t) + \int_0^t DF^V(Z^V \circ \eta(s))(X^V(s) - Z^V(s))ds \]
\[ + \frac{V^{-\beta}}{2} \int_0^t DF^V(Z^V \circ \eta(s))F^V(Z^V \circ \eta(s))ds \]

where \( M^V(t) \) is a martingale and scales like \( V^{-(1+\beta)/2} \).

- \( M^V(t) \) scales like \( O(V^{-(1+\beta)/2}) \).
- Last term scales like \( O(V^{-\beta}) \).
- \((1 + \beta)/2 > \beta\), so done.
Weak error of Euler tau-leaping

Theorem (Anderson, Ganguly, Kurtz – 2010)

Let $X^V(t)$ and $Z^V(t)$ be as before. Then, for any continuously differentiable function $f$, as $V \to \infty$

$$V^\beta \left( \mathbb{E} f(X^V(t)) - \mathbb{E} f(Z^V(t)) \right) \to \mathcal{E}(t) \cdot \nabla f(x(t)).$$
Strong error of midpoint tau-leaping

Theorem (Anderson, Ganguly, Kurtz – 2010)

Let $X^V(t)$ and $Z^V(t)$ denote the scaled processes for $t \in [0, T]$. Then there exists a constant $C = C(T) > 0$ such that

$$\sup_{t \leq T} \mathbb{E}|X^V(t) - Z^V(t)| \leq \frac{C}{V^{\kappa(\beta)}}, \quad \text{where} \quad \kappa(\beta) = \min \left\{ 2\beta, \frac{1 + \beta}{2} \right\}.$$
Strong error of midpoint tau-leaping

Theorem (Anderson, Ganguly, Kurtz – 2010)

Let $X^V(t)$ and $Z^V(t)$ denote the scaled processes for $t \in [0, T]$. Then there exists a constant $C = C(T) > 0$ such that

$$
\sup_{t \leq T} \mathbb{E}|X^V(t) - Z^V(t)| \leq \frac{C}{V^{\kappa(\beta)}}, \quad \text{where} \quad \kappa(\beta) = \min\left\{2\beta, \frac{1 + \beta}{2}\right\}.
$$
A new coupling of \(X^V(t)\) and \(Z^V(t)\)

To sharpen the result, let

\[
X^V(t) = X^V(0) + \frac{1}{V} \sum_k \left[ Y_{k,1} \left( V \int_0^t A_k^V(X^V(s)) \wedge A_k^V(\rho^V \circ Z^V \circ \eta(s)) \, ds \right) 
\right. \\
\left. + Y_{k,2} \left( V \int_0^t A_k^V(X^V(s)) - A_k^V(X^V(s)) \wedge A_k^V(\rho^V \circ Z^V \circ \eta(s)) \, ds \right) \right](\nu'_k - \nu_k)
\]

\[
Z^V(t) = X^V(0) + \frac{1}{V} \sum_k \left[ Y_{k,1} \left( V \int_0^t A_k^V(X^V(s)) \wedge A_k^V(\rho^V \circ Z^V \circ \eta(s)) \, ds \right) 
\right. \\
\left. + Y_{k,3} \left( V \int_0^t A_k^V(\rho^V \circ Z^V \circ \eta(s)) - A_k^V(X^V(s)) \wedge A_k^V(\rho^V \circ Z^V \circ \eta(s)) \, ds \right) \right](\nu'_k - \nu_k),
\]

where the \(Y_{k,i}\) are mutually independent Poisson processes.
Theorem (Anderson, Ganguly, Kurtz – 2010)

Let $X^V(t)$ and $Z^V(t)$ satisfy the new coupling for $t \in [0, T]$. Then as $V \to \infty$

\[ V^{2\beta} (X^V - Z^V) \to \mathcal{E}_1, \quad \text{if} \quad \beta < 1/3. \]

\[ V^{2\beta} (X^V - Z^V) \Rightarrow \mathcal{E}_2, \quad \text{if} \quad \beta = 1/3. \]

\[ V^{(1+\beta)/2} (X^V - Z^V) \Rightarrow \mathcal{E}_3, \quad \text{if} \quad \beta > 1/3. \]

Where $\mathcal{E}_1$ is deterministic and $\mathcal{E}_2, \mathcal{E}_3$ are stochastic.
Idea of proof

Can show:

$$X^V(t) - Z^V(t) \approx M^V(t) + \int_0^t DF^V(Z^V \circ \eta(s))(X^V(s) - Z^V(s))ds + V^{-2\beta} H^V(t)$$

where $V^{(1+\beta)/2} M^V \Rightarrow M(t)$ and $H^V \rightarrow H$.

- $M^V(t)$ scales like $O(V^{-(1+\beta)/2})$.

- Last term scales like $O(V^{-2\beta})$.

- Now scaling goes with $\min\{2\beta, (1 + \beta)/2\}$.
Weak error of midpoint tau-leaping

Theorem (Anderson, Ganguly, Kurtz – 2010)

Let $X^V(t)$ and $Z^V(t)$ be as before. Then, for any $C^3$ function $f$, there exists a constant $C = C(f, T) > 0$ such that

$$\sup_{t \leq T} \left| \mathbb{E}f(X^V(t)) - \mathbb{E}f(Z^V(t)) \right| \leq \frac{C}{V^{2\beta}}.$$
Example

Consider $A$ as a prey and $B$ as a predator.

$$A \overset{\kappa_1}{\rightarrow} 2A, \quad A + B \overset{\kappa_2}{\rightarrow} 2B, \quad B \overset{\kappa_3}{\rightarrow} \emptyset,$$

with $A(0) = B(0) = 1000$ and $\kappa_1 = 2$, $\kappa_2 = .002$, $\kappa_3 = 2$.

$V = 1,000$ and $\tau = 1/20 = 1/V^{.434}$ (so $\beta = .434$). Simulating 30,000 sample paths yields the following approximate distributions for $B(10)$:
Diffusion Approximation

Under the classical scaling, can also use a diffusion approximation.

Let
\[ \tilde{Y}_k(u) := Y_k(u) - u, \]
denote the centered Poisson process. Then, \( X^V(t) \) satisfies
\[
X^V(t) = X^V(0) + \sum_k \int_0^t A^V_k(X^V(s)) \, ds (\nu' - \nu_k) \\
+ \sum_k \frac{1}{V} \tilde{Y}_k(V \int_0^t A^V_k(X^V(s)) \, ds)(\nu' - \nu_k). 
\]
Diffusion Approximation

Under the classical scaling, can also use a diffusion approximation.

Let
\[ \tilde{Y}_k(u) := Y_k(u) - u, \]
denote the centered Poisson process. Then, \( X^V(t) \) satisfies

\[
X^V(t) = X^V(0) + \sum_k \int_0^t A^V_k(X^V(s)) ds(\nu'_k - \nu_k) \\
+ \sum_k \frac{1}{V} \tilde{Y}_k(V \int_0^t A^V_k(X^V(s)) ds)(\nu'_k - \nu_k).
\]

Using that \( \tilde{Y}_k(Vu)/\sqrt{V} \approx W_k(u) \), yields

\[
X^V(t) \approx X^V(0) + \sum_k \int_0^t A^V_k(X^V(s)) ds(\nu'_k - \nu_k) \\
+ \sum_k \frac{1}{\sqrt{V}} W_k(\int_0^t A^V_k(X^V(s)) ds)(\nu'_k - \nu_k).
\]
Diffusion Approximation

This suggests a good approximation to $X^V$ will be the solution to the SDE

$$Z^V(t) = X^V(0) + \sum_k \int_0^t A^V_k(Z^V(s))ds (\nu'_k - \nu_k)$$

$$+ \frac{1}{\sqrt{V}} \sum_k W_k(\int_0^t A^V_k(Z^V(s))ds)(\nu'_k - \nu_k),$$

where the $W_k$ are independent Brownian motions.
This suggests a good approximation to $X^V$ will be the solution to the SDE

$$Z^V(t) = X^V(0) + \sum_k \int_0^t A_k^V(Z^V(s)) ds \left( \nu'_k - \nu_k \right)$$

$$+ \frac{1}{\sqrt{V}} \sum_k W_k(\int_0^t A_k^V(Z^V(s)) ds)(\nu'_k - \nu_k),$$

where the $W_k$ are independent Brownian motions.

An equivalent formulation is given by the Itô equations

$$dZ^V(t) = \sum_k A_k^V(Z^V(t)) \left( \nu'_k - \nu_k \right) dt + \frac{1}{\sqrt{V}} \sum_k (\nu'_k - \nu_k) \sqrt{A_k^V(Z^V(t))} dW_k(t),$$

(1)

System (1) is termed the chemical Langevin equation in the biology/chemistry literature.
Error of Diffusion Approximation

Accuracy:

- Pathwise\(^1\), the error is \( O \left( \frac{\log(V)}{V} \right) \).

- Pathwise, tau-leaping methods can be more or less accurate, depending upon choice of \( \beta \) and \( V \).

<table>
<thead>
<tr>
<th>( V )</th>
<th>Euler ( \tau )-leap more accurate if ( \beta &gt; )</th>
<th>Step-size (Euler ( \tau ))</th>
<th>Midpoint ( \tau )-leap more accurate if ( \beta &gt; )</th>
<th>Step-size (mdpt ( \tau ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>( \beta &gt; .6684 )</td>
<td>0.046</td>
<td>( \beta &gt; .3368 )</td>
<td>( \tau &lt; .212 )</td>
</tr>
<tr>
<td>1,000</td>
<td>( \beta &gt; .7202 )</td>
<td>0.0069</td>
<td>( \beta &gt; .4404 )</td>
<td>( \tau &lt; .0477 )</td>
</tr>
<tr>
<td>10,000</td>
<td>( \beta &gt; .7589 )</td>
<td>0.00092</td>
<td>( \beta &gt; .5178 )</td>
<td>( \tau &lt; .0085 )</td>
</tr>
<tr>
<td>100,000</td>
<td>( \beta &gt; .7878 )</td>
<td>0.00012</td>
<td>( \beta &gt; .5756 )</td>
<td>( \tau &lt; .0013 )</td>
</tr>
</tbody>
</table>

\(^1\)T. Kurtz, SPA, 6, 1978, 223 - 240.
Approximating the diffusion approximation

- If $-\mathcal{A}$ is the generator for the jump process
  $-\mathcal{B}$ is the generator for the diffusion.

  \[ \mathcal{A} f = \mathcal{B} f + \mathcal{O} \left( \frac{1}{V^2} \right). \]

  Suggests that for fixed $t$, weak error is $\mathcal{O}(1/V^2)$.

- For fixed $t$, even midpoint $\tau$-leap will never be as accurate.

- However, now you must **approximate the diffusion approximation**.
Consider SDEs of the form

\[ dX(t) = b(X(t))dt + \sum_k \sigma_k(X(t)) \nu_k \, dW_k(t), \]

where \( b: \mathbb{R}^d \rightarrow \mathbb{R}^d \), \( \sigma_k: \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0} \), \( \nu_k \in \mathbb{R}^d \), and \( \{ W_k(t) : k = 1, \ldots, m \} \) are independent, one-dimensional Wiener processes.

Thus, randomness is entering the system in fixed directions \( \nu_k \), but at variable rates \( \sigma_k(X(t)) \).

Consider SDEs of the form

$$dX(t) = b(X(t))dt + \sum_{k} \sigma_{k}(X(t)) \nu_{k} \, dW_{k}(t),$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma_{k} : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, $\nu_{k} \in \mathbb{R}^d$, and $\{W_{k}(t) : k = 1, \ldots, m\}$ are independent, one-dimensional Wiener processes.

- Thus, randomness is entering the system in fixed directions $\nu_{k}$, but at variable rates $\sigma_{k}(X(t))$.

- In fact, all uniformly elliptic SDEs can be represented in such a way\(^2\).

---

Approximating SDES with special structure

Let approximate path be \( Y_i = Y(t_i) \).

Algorithm \(^1\) Fixing a \( \theta \in (0, 1) \), we define

\[
\alpha_1 \overset{\text{def}}{=} \frac{1}{2} \frac{1}{\theta(1 - \theta)} \quad \text{and} \quad \alpha_2 \overset{\text{def}}{=} \frac{1}{2} \frac{(1 - \theta)^2 + \theta^2}{\theta(1 - \theta)}.
\] (2)

Fixing \( h \), we first compute a \( \theta \)-midpoint \( y^* \) and then the new value \( Y_i \): (\( \eta_{jk}^{(i)} \) are ind. normals)

**Step 1.** Set

\[
y^* = Y_{i-1} + b(Y_{i-1})\theta h + \sum_k \sigma_k(Y_{i-1}) \nu_k \eta_{1k}^{(i)} \sqrt{\theta h}
\]

**Step 2.** Set

\[
Y_i = y^* + (\alpha_1 b(y^*) - \alpha_2 b(Y_{i-1}))(1 - \theta)h
\]

\[
+ \sum_k \sqrt{[\alpha_1 \sigma_k^2(y^*) - \alpha_2 \sigma_k^2(Y_{i-1})]^+} \nu_k \eta_{2k}^{(i)} \sqrt{(1 - \theta)h}.
\]

Order of approximation

Theorem (One-step approximation\(^1\))

If \(b, \sigma_k \in C^6_0\) with \(\inf_x \sigma_k(x) > 0\), then for \(h\) sufficiently small and \(f \in C^6_0\):

\[
|E_x f(X(h)) - E_x f(Y_1)| \leq K \|f\|_6 h^3.
\]

Order of approximation

Theorem (One-step approximation\textsuperscript{1})
If $b, \sigma_k \in C_0^6$ with $\inf_x \sigma_k(x) > 0$, then for $h$ sufficiently small and $f \in C_0^6$:

$$|\mathbb{E}_x f(X(h)) - \mathbb{E}_x f(Y_1)| \leq K \|f\|_6 h^3.$$ 

Theorem (Global approximation\textsuperscript{1})
Under the same assumptions for any $T > 0$:

$$\sup_{0 \leq n \leq T/h} |\mathbb{E}_x f(X(nh)) - \mathbb{E}_x f(Y_n)| \leq C(T)\|f\|_6 h^2.$$ 

\textsuperscript{1} D. F. Anderson and J. C. Mattingly, to appear in Comm. Math. Sci., 2010
Why this works

System

\[ dX(t) = b(X(t))dt + \sum_{k} \sigma_k(X(t)) \nu_k \, dW_k(t), \]

is equivalent to

\[ X(t) = X(0) + \int_0^t b(X(s))ds + \sum_{k=1}^{\infty} \nu_k \int_0^\infty \int_0^t 1_{[0, \sigma_k^2(X(s)))}(u) \tilde{W}_k(ds \times du), \]

where the \( \tilde{W}_k \) are independent, space-time white noise processes:

- Random measure such that:
  1. for \( A \subset \mathbb{R}^2 \), \( \tilde{W}_k(A) \sim \text{Normal}(0, \text{Area}(A)) \).
  2. \( A, B \subset \mathbb{R}^2 \), with \( A \cap B = \emptyset \) \( \Rightarrow \tilde{W}_k(A) \) and \( \tilde{W}_k(B) \) are independent.

Challenge is in approximating diffusion term.
Why this works: $\theta = 1/2$

Region 1

$V = \sigma^2_k(y^*) - \sigma^2_k(X(0)) - \sigma^2_k(X(t))$

$|V| = \sigma^2_k(y^*) - \sigma^2_k(X(0)) - \sigma^2_k(X(t))$

Region 2

Region 3

Region 4

Region 5
Why this works: $\theta = 1/2$

\[
\sigma_k^2(X(0)) \quad \sigma_k^2(y^*)
\]

Region 1

\[
V = \sigma_k^2(y^*) - \sigma_k^2(X(0)) \quad \sigma_k^2(X(t))
\]

Region 2

\[
|V| = \sigma_k^2(Y) - \sigma_k^2(X(0)) \quad \sigma_k^2(X(t))
\]

Region 3

\[
\sigma_k^2(X(t))
\]

Region 4

\[
\sigma_k^2(X(t))
\]

Region 5

\[
\sigma_k^2(X(t))
\]

$t$
Why this works: $\theta = 1/2$

\[
\sigma_k^2(X(0)) \quad \sigma_k^2(y^*) \\
\sigma_k^2(X(t))
\]

Region 1

\[
V = \sigma_k^2(y^*) - \sigma_k^2(X(0)) \sigma_k^2(X(t))
\]

Region 2

\[
V = \sigma_k^2(y^*) - \sigma_k^2(X(0)) \sigma_k^2(X(t))
\]

Region 3

\[
V = \sigma_k^2(y^*) - \sigma_k^2(X(0)) \sigma_k^2(X(t))
\]

Region 4

\[
|V| = 2 |V|
\]

Region 5

\[
|V| = 2 |V|
\]
Example

Consider the following system

\[
\begin{bmatrix}
    dX_1(t) \\
    dX_2(t)
\end{bmatrix} = \begin{bmatrix}
    -X_2(t) \\
    X_1(t)
\end{bmatrix} dt + \sqrt{\frac{\sin^2(X_1(t) + X_2(t)) + 6}{t + 1}} \begin{bmatrix}
    1 \\
    0
\end{bmatrix} dW_1(t) \\
+ \sqrt{\frac{\cos^2(X_1(t) + X_2(t)) + 6}{t + 1}} \begin{bmatrix}
    0 \\
    1
\end{bmatrix} dW_2(t).
\]

where \( W_1(t) \) and \( W_2(t) \) are independent Wiener processes.

It is simple to show that

\[
\mathbb{E}|X(t)|^2 = \mathbb{E}X(0)^2 + 13 \log(1 + t).
\]  
(3)
Figure: Log-log plots. Approximating $\mathbb{E}|X(1)|^2$. The best fit lines have slopes 2.223, .952, and 1.098, for the Weak Trapezoidal Algorithm, Euler's method, and the midpoint drift method, respectively.
Figure: (a) # degenerate steps, first example, $h = 1/10$, (b) slopes 1.865, 1.996, 2.029, and 2.033 for $\theta = .05, .25, .50, .75$, respectively.

For the model,

$$dX(t) = \sqrt{X(t)^2 + 1} \, dW(t), \quad X(0) = 1$$
Outlook

1. These methods will rarely be used in isolation.

2. Need to begin to combine them into hybrid strategies
   2.1 Some pieces of the network (species and/or reactions) are treated discretely.
   2.2 Some treated diffusively.
   2.3 Some treated as absolutely continuous (ODE approximation depending on stochastic parts).

3. And, want to be able to dynamically change the treatment.
Outlook

1. These methods will rarely be used in isolation.

2. Need to begin to combine them into hybrid strategies
   2.1 Some pieces of the network (species and/or reactions) are treated discretely.
   2.2 Some treated diffusively.
   2.3 Some treated as absolutely continuous (ODE approximation depending on stochastic parts).

3. And, want to be able to dynamically change the treatment.

Work has begun on algorithms:

1. E. Haseltine and J. Rawlings or Salis & Kaznessis: hybrid discrete/Langevin.