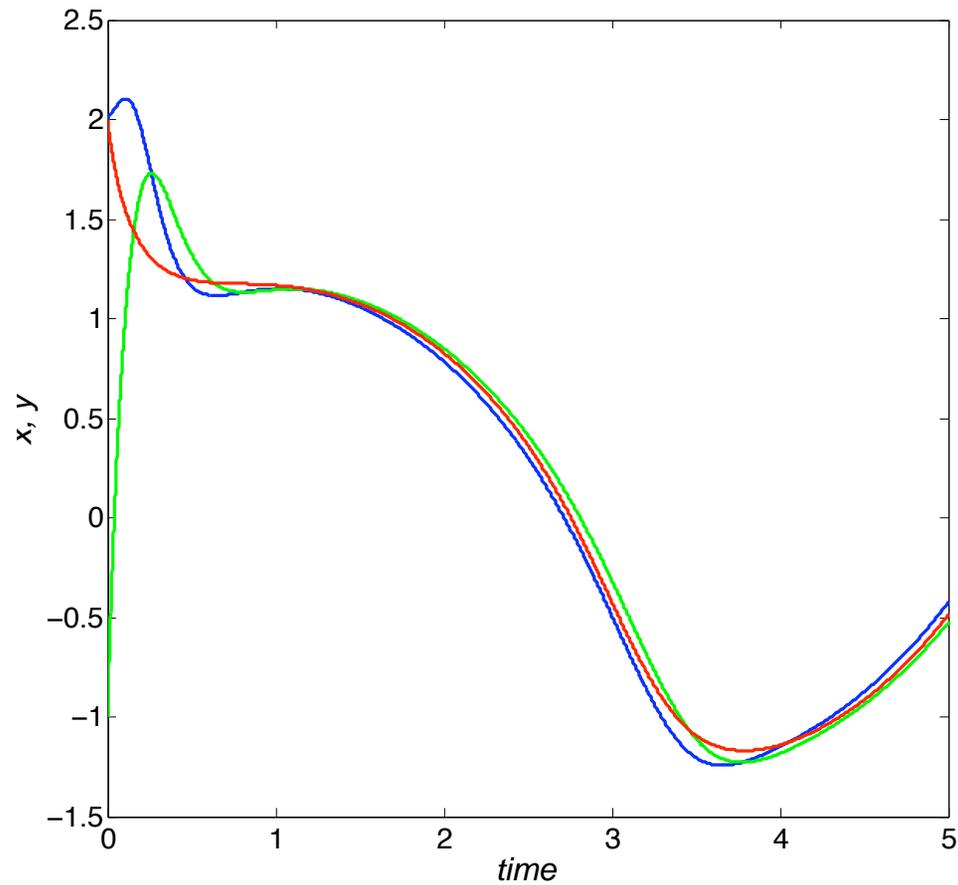


Asymptotic and numerical techniques for systems with multiple time-scales

Eric Vanden-Eijnden

Courant Institute

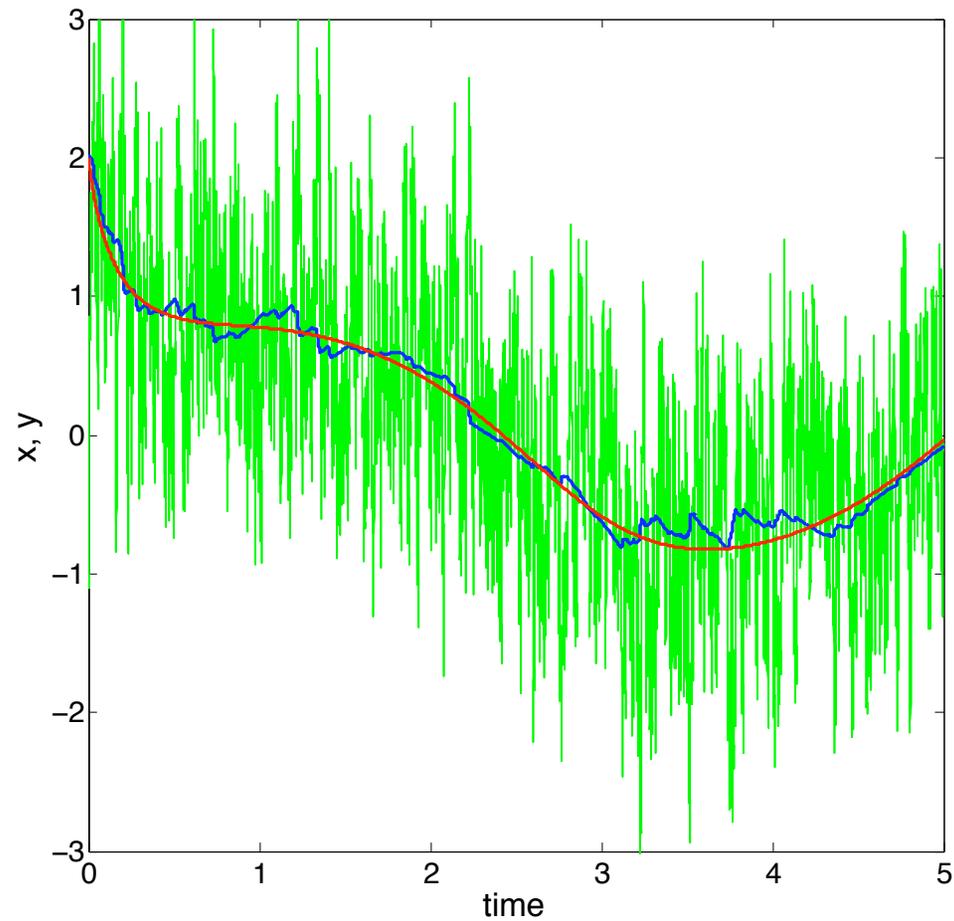


The solution of

$$\begin{cases} \dot{X} = -Y^3 + \cos(t) + \sin(\sqrt{2}t) \\ \dot{Y} = -\varepsilon^{-1}(Y - X) \end{cases}$$

when $\varepsilon = 0.1$ and we took $X(0) = 2$, $Y(0) = -1$. X is shown in blue, and Y in green. Also shown in red is the solution of the limiting equation

$$\dot{X} = -X^3 + \cos(t) + \sin(\sqrt{2}t)$$



The solution of

$$\begin{cases} \dot{X} = -Y^3 + \cos(t) + \sin(\sqrt{2}t) \\ dY = -\varepsilon^{-1}(Y - X)dt + \varepsilon^{-1/2}dW \end{cases}$$

when $\varepsilon = 0.01$ and $X(0) = 2, Y(0) = -1$. X is shown in blue, and Y in green. Also shown in red is the solution of the limiting equation

$$\dot{X} = -X^3 + X + \cos(t) + \sin(\sqrt{2}t)$$

Notice how noisy Y is.

Outline

Lecture 1: Theoretical aspects

Averaging principle for singularly perturbed Markov processes;
Homogenization

Lecture 2: Numerical Aspects

HHM-like multiscale integrators;
Seamless integrators (boosting method).

Lecture 3: Generalization to Markov jump processes

Gillespie stochastic simulation algorithm (SSA);
Nested SSA (nSSA).

Lecture 1: Singular perturbations techniques for Markov processes

Consider

$$\begin{cases} \dot{X} = f(X, Y), \\ dY = \varepsilon^{-1}g(X, Y)dt + \varepsilon^{-1/2}\sigma(X, Y)dW(t), \end{cases} \quad (\star)$$

Assume that: (i) the evolution Y at every fixed $X = x$ is ergodic with respect to the probability distribution

$$d\mu_x(y)$$

and (ii)

$$F(x) = \int_{\mathbb{R}^m} f(x, y)d\mu_x(y) \quad \text{exists}$$

Then in the limit as $\varepsilon \rightarrow 0$ the evolution for X solution of (\star) is governed by

$$\dot{X} = F(X)$$

In addition

$$F(x) = \int_{\mathbb{R}^m} f(x, y)d\mu_x(y) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(x, Y_t^x)dt$$

where

$$dY^x = \varepsilon^{-1}g(x, Y^x)dt + \varepsilon^{-1/2}\sigma(x, Y^x)dW(t)$$

Derivation 1. Consider the simpler equation

$$\dot{X} = f(X, g(t/\varepsilon))$$

Assume that

$$\forall y : |f(x_1, y) - f(x_2, y)| \leq K|x_1 - x_2|$$

and

$$\forall x : \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(x, g(s)) ds = F(x)$$

Then (assuming $X_{t=0} = x$)

$$\begin{aligned} X_t - x &= \int_0^t f(X_s, g(s/\varepsilon)) ds \\ &= \int_0^t f(x, g(s/\varepsilon)) ds + \int_0^t (f(X_s, g(s/\varepsilon)) - f(x, g(s/\varepsilon))) ds \\ &= t \left(\frac{\varepsilon}{t} \int_0^{t/\varepsilon} f(x, g(u)) du \right) + \delta(t, \varepsilon) \end{aligned}$$

Letting $\varepsilon/t \rightarrow 0$, the term in parenthesis converges to $F(x)$. On the other hand, for all ε , $|\delta(t, \varepsilon)| \leq Kt^2$. Thus, as $\varepsilon/t \rightarrow 0$, this equation converges to

$$X_t - x = tF(x) + O(t^2)$$

which is the forward Euler discretization of $\dot{X} = F(X)$.

Derivation 2. Let L be the infinitesimal generator of the Markov process generated by

$$\begin{cases} \dot{X} = f(X, Y), \\ dY = \varepsilon^{-1}g(X, Y)dt + \varepsilon^{-1/2}\sigma(X, Y)dW(t), \end{cases} \quad (*)$$

i.e.

$$\lim_{t \rightarrow 0^+} \frac{1}{t} (\mathbf{E}_{x,y} \phi(X_t, Y_t) - \phi(x, y)) = (L\phi)(x, y)$$

Then $L = L_0 + \varepsilon^{-1}L_1$, with

$$\begin{cases} L_0 = f(x, y) \cdot \nabla_x \\ L_1 = g(x, y) \cdot \nabla_y + (\sigma\sigma^T)(x, y) : \nabla_y \nabla_y \end{cases}$$

and

$$u(x, y, t) = \mathbf{E}_{x,y} \phi(X_t, Y_t)$$

satisfies the backward Kolmogorov equation

$$\frac{\partial u}{\partial t} = L_0 u + \varepsilon^{-1} L_1 u, \quad u|_{t=0} = \phi \quad (\text{BKE})$$

Look for a solution in the form of

$$u = u_0 + \varepsilon u_1 + O(\varepsilon^2)$$

so that $\lim_{\varepsilon \rightarrow 0} u = u_0$ (formally).

Inserting $u = u_0 + \varepsilon u_1 + O(\varepsilon^2)$ into (BKE) and equating equal powers in ε leads to the hierarchy of equations

$$\begin{cases} L_1 u_0 = 0, \\ L_1 u_1 = \frac{\partial u_0}{\partial t} - L_0 u_0, \\ L_1 u_2 = \dots \end{cases} \quad (**)$$

The first equation tells that u_0 belong to the null-space of L_1 .

The assumption that the evolution of Y at every fixed $X = x$ (i.e. the process with generator L_1) is ergodic with respect to the probability distribution $\mu_x(y)$ implies that for every x , the null-space of L_1 is spanned by functions constant in y , i.e.

$$u_0 = u_0(x, t)$$

Since the null-space of L_1 is non-trivial, the next equations each requires a solvability condition, namely that their right hand-side belongs to the range of L_1 .

To see what this solvability condition actually is, take the expectation of both sides of the second equation in $(\star\star)$ with respect to $d\mu_x(x)$. This gives

$$0 = \int_{\mathbb{R}^m} d\mu_x(y) \left(\frac{\partial u_0}{\partial t} - L_0 u_0 \right)$$

Explicitly, this equation is

$$\frac{\partial u_0}{\partial t} = F(x) \cdot \nabla_x u_0 \quad (\square)$$

where

$$F(x) = \int_{\mathbb{R}^m} f(x, y) d\mu_x(y)$$

(\square) is the backward Kolmogorov equation of the limiting equation

$$\dot{X} = F(X)$$

Remark: Computing the expectation wrt $\mu_x(y)$ in practice.

We have

$$(e^{L_1 t} \phi)(x, y) \rightarrow \int_{\mathbb{R}^m} \phi(x, y) d\mu_x(y) \quad \text{as } t \rightarrow \infty$$

In other words, if $v(x, y, t)$ satisfies

$$\frac{\partial v}{\partial t} = L_1 v, \quad v|_{t=0} = \phi$$

so that formally $v(x, y, t) = (e^{L_1 t} \phi)(x, y)$, then we have

$$\lim_{t \rightarrow \infty} v(x, y, t) = \int_{\mathbb{R}^m} \phi(x, y) d\mu_x(y)$$

But since $v(x, y, t) = \mathbf{E}_y \phi(Y_t^x)$ where

$$dY^x = \varepsilon^{-1} g(x, Y^x) dt + \varepsilon^{-1/2} \sigma(x, Y^x) dW(t)$$

it follows that

$$\begin{aligned} \int_{\mathbb{R}^m} \phi(x, y) d\mu_x(y) &= \lim_{t \rightarrow \infty} \mathbf{E}_y \phi(Y_t^x) \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \phi(Y_t^x) dt \end{aligned}$$

Example: the Lorenz 96 (L96) model

L96 consists of K slow variables X_k coupled to $J \times K$ fast variables $Y_{j,k}$ whose evolution is governed by

$$\begin{cases} \dot{X}_k = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F_x + \frac{h_x}{J} \sum_{j=1}^J Y_{j,k} \\ \dot{Y}_{j,k} = \frac{1}{\varepsilon} (-Y_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - Y_{j,k} + h_y X_k). \end{cases}$$

We will study (L96) with $F_x = 10$, $h_x = -0.8$, $h_y = 1$, $K = 9$, $J = 8$, and two values of ε : $\varepsilon = 1/128$ and $\varepsilon = 1/1024$.

Empirical way to check whether (L96) converges, as $\varepsilon \rightarrow 0$, towards

$$\dot{X}_k = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F_x + G_k(X)$$

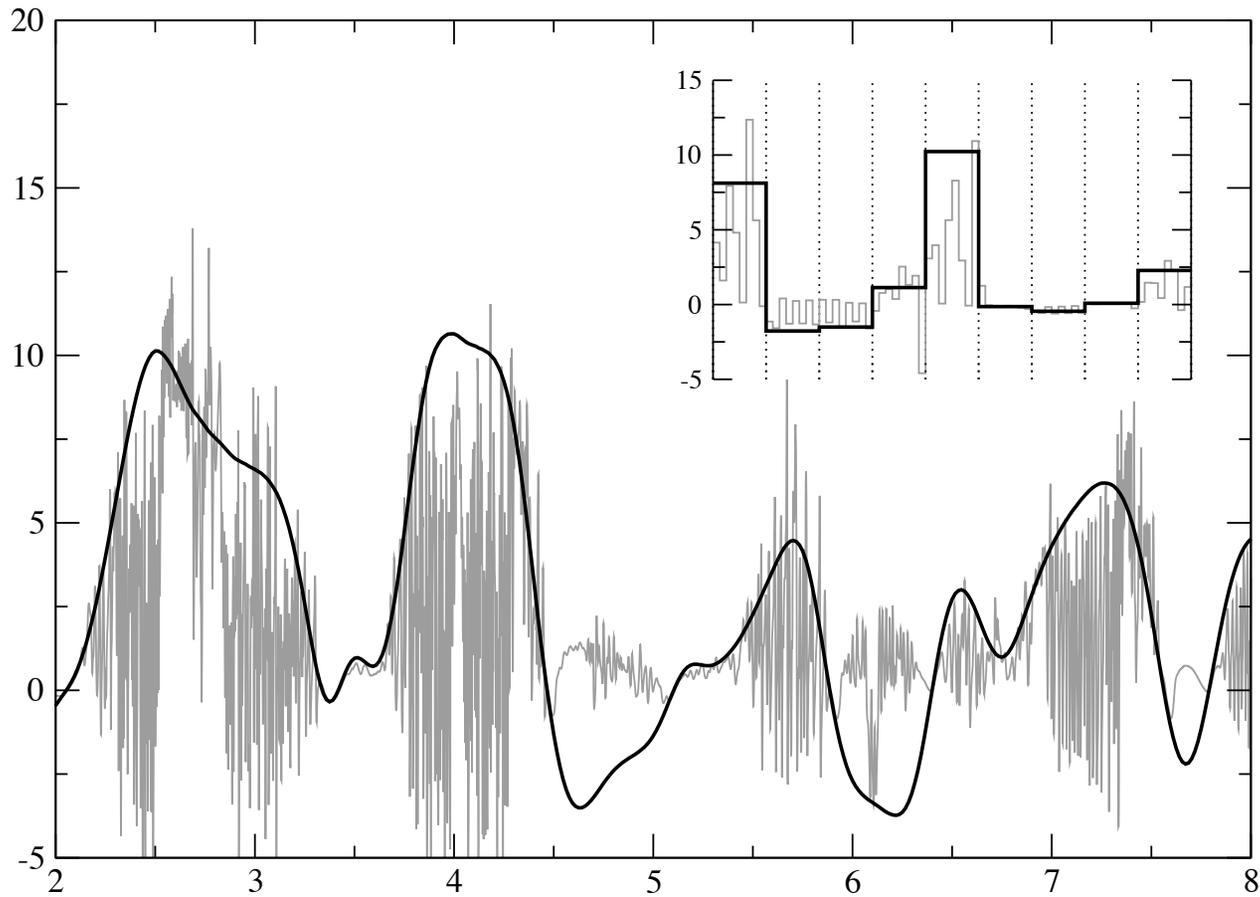
where

$$G_k(x) = \int_{\mathbb{R}^{KJ}} \left(\frac{h_x}{J} \sum_{j=1}^J y_{j,k} \right) d\mu_x(y)$$

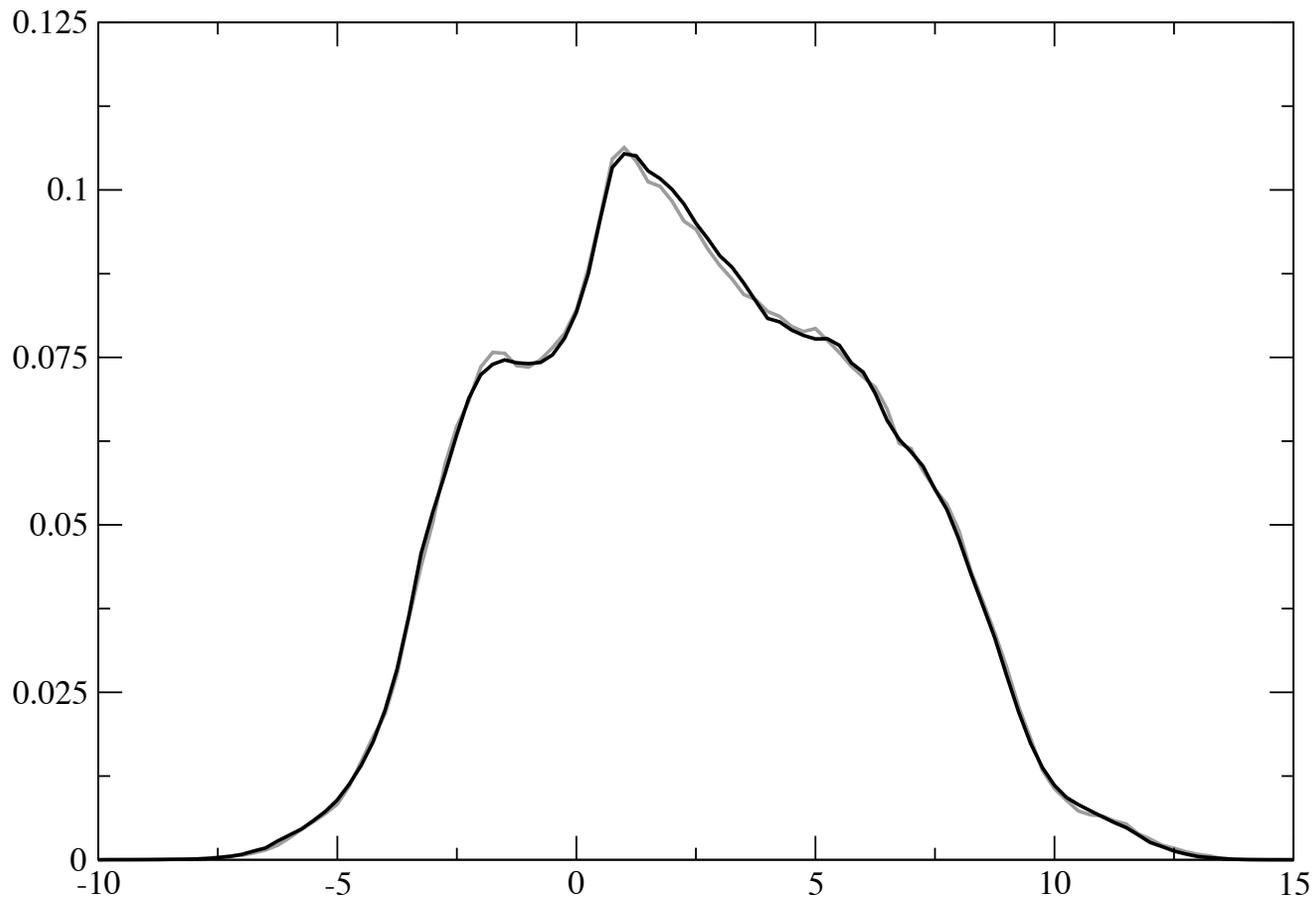
and $\mu_x(y)$ is the equilibrium measure of

$$\dot{Y}_{j,k}^x = \frac{1}{\varepsilon} (-Y_{j+1,k}^x(Y_{j+2,k}^x - Y_{j-1,k}^x) - Y_{j,k}^x + h_y x_k)$$

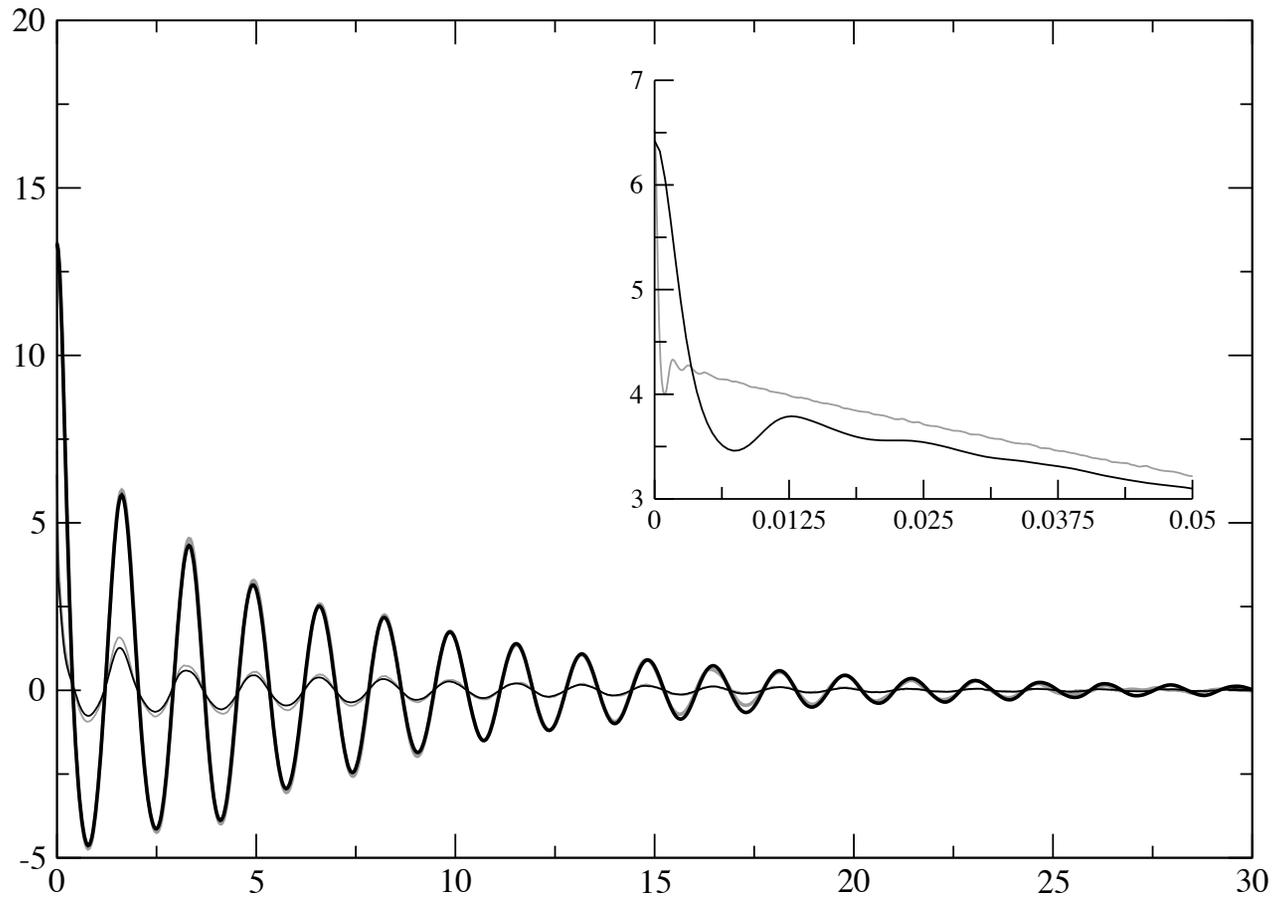
assuming that it exists for every x .



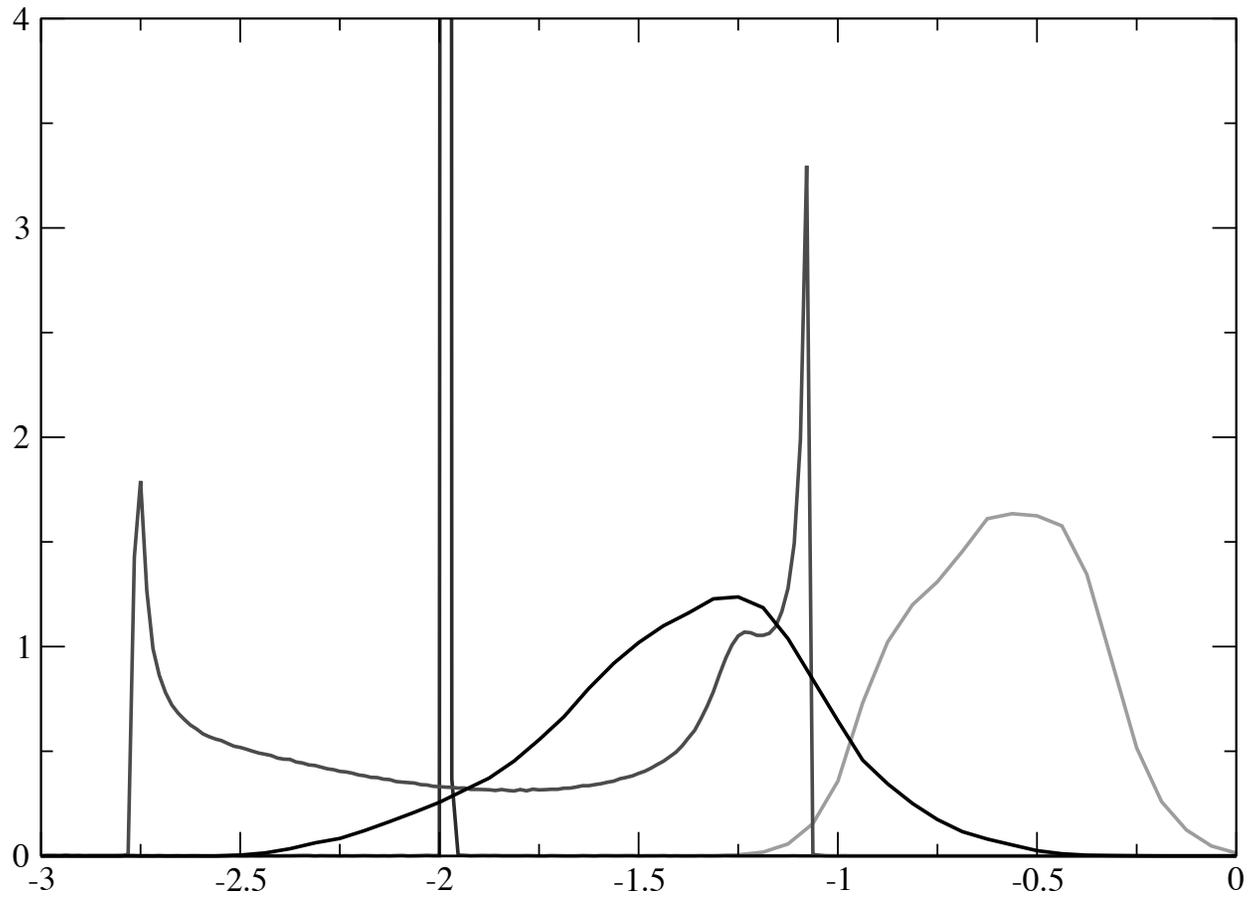
Typical time-series of the slow (black line) and fast (grey line) modes; $K = 9$, $J = 8$, $\varepsilon = 1/128$. The subplot displays a typical snapshot of the slow and fast modes at a given time.



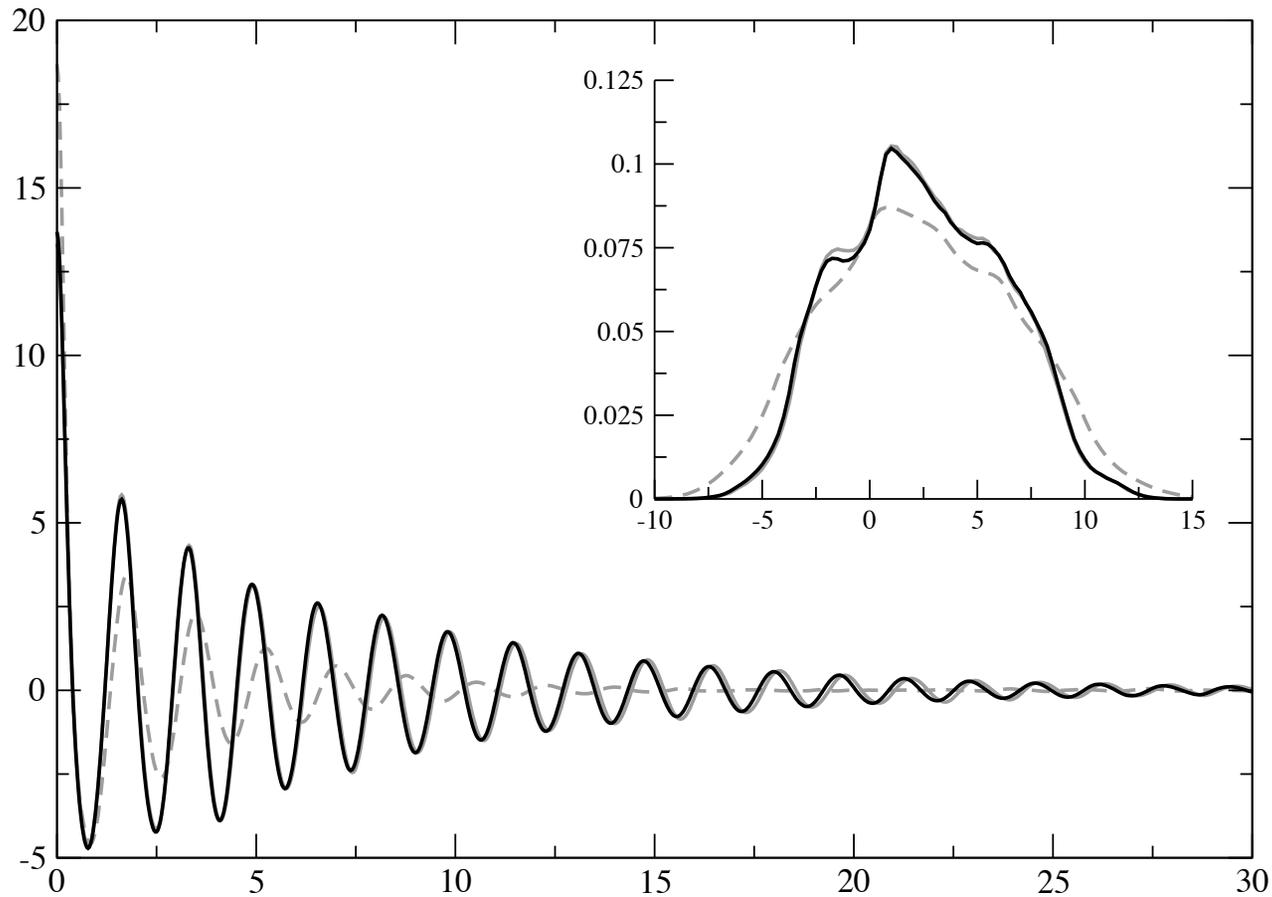
PDF of the slow variable; $K = 9$, $J = 8$, black line: $\varepsilon = 1/128$, grey line: $\varepsilon = 1/1024$. The insensitivity in ε of the PDFs indicates that the slow variables have already converged close to their limiting behavior when $\varepsilon = 1/128$.



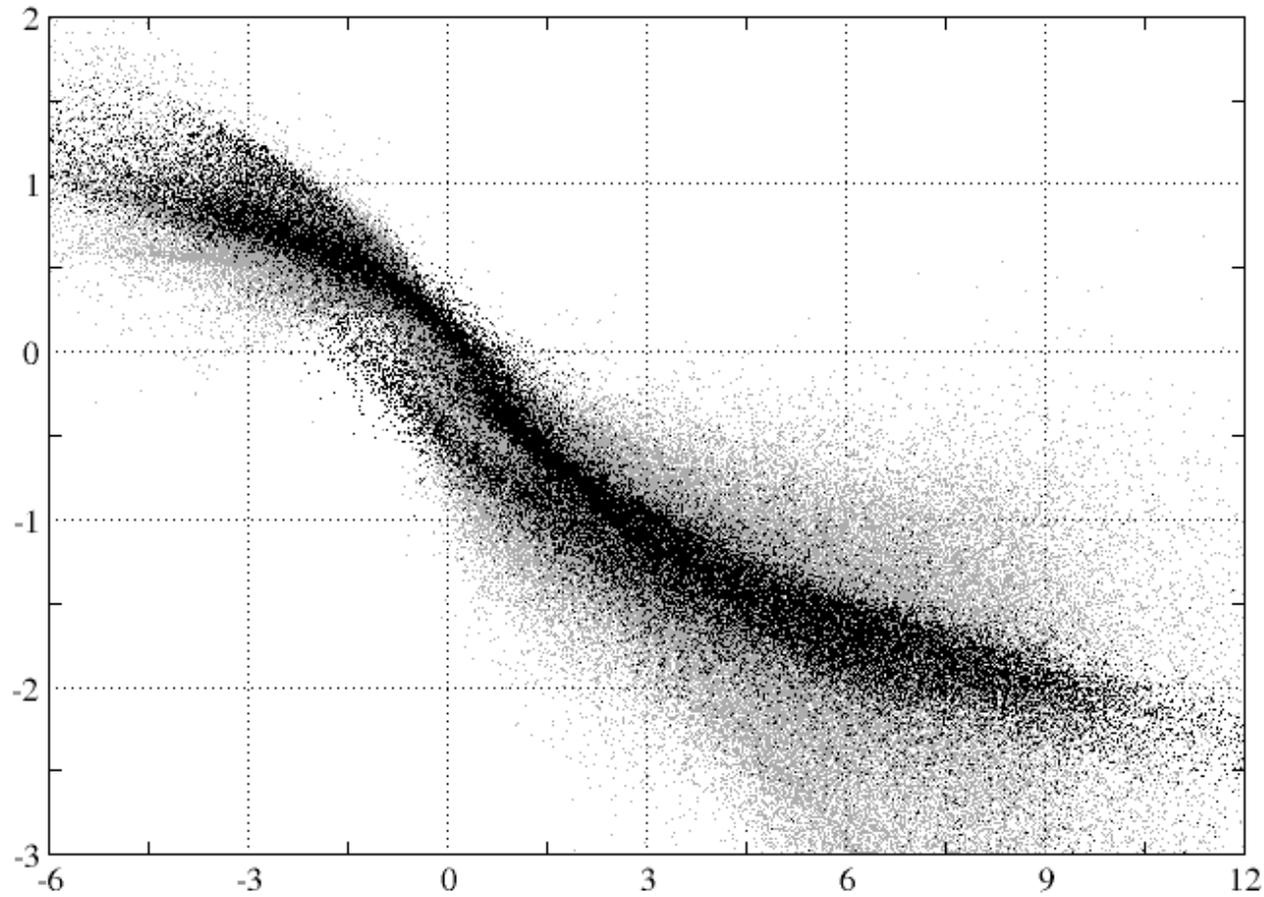
ACFs of the slow (thick line) and fast (thin line) variables; $\varepsilon = 1/128$, grey line: $\varepsilon = 1/1024$. The subplot is the zoom-in of the main graph which shows the transient decay of the ACFs of the fast modes becoming faster as ε is decreased: this is the only signature in the ACFs of the fact that the $Y_{j,k}$'s are faster.



Typical PDFs of the coupling term $(h_x/J) \sum_{j=1}^J Y_{j,k}^x$ for various values of x . These PDFs are robust against variations in the initial conditions indicating that the dynamics of the fast modes conditional on the slow ones being fixed is ergodic (?).



Comparison between the ACFs and PDFs; black line: limiting dynamics; full grey line: $\varepsilon = 1/128$. Also shown in dashed grey are the corresponding ACF and PDF produced by the truncated dynamics where the coupling of the slow modes X_k with the fast ones, $Y_{j,k}$, is artificially switched off.



Black points: scatterplot of the forcing $F(x)$ in the limiting dynamics.
Grey points: scatter plot of the bare coupling term $(h_x/J) \sum_{j=1}^J Y_{j,k}$
when $\varepsilon = 1/128$.

Diffusive time-scales (homogenization)

Suppose that

$$F(x) = \int_{\mathbb{R}^m} f(x, y) d\mu_x(y) = 0. \quad (\Delta)$$

Then the limiting equation on the $O(1)$ time-scale is trivial, $\dot{X} = 0$, and the interesting dynamics arises on the diffusive time-scale $O(\varepsilon^{-1})$.

Rescale time as $t \mapsto t/\varepsilon$ and consider then

$$\begin{cases} \dot{X} = \varepsilon^{-1} f(X, Y), \\ dY = \varepsilon^{-2} g(X, Y) dt + \varepsilon^{-1} \sigma(X, Y) dW_t. \end{cases}$$

Proceeding as before we arrive at the following limiting equation when $\varepsilon \rightarrow 0$:

$$dX = \bar{f}(X) dt + \bar{\sigma}(X) dB_t,$$

where

$$\begin{aligned} (\bar{L}\phi)(x) &\equiv \bar{f}(x) \cdot \nabla_x \phi(x) + \frac{1}{2} (\bar{\sigma} \bar{\sigma}^T)(x) : \nabla_x \nabla_x \phi(x) \\ &\equiv \int_0^\infty dt \int_{\mathbb{R}^m} d\mu_x(y) f(x, y) \cdot \nabla_x (\mathbf{E}_y f(x, Y_t^x) \cdot \nabla_x \phi(x)) \end{aligned}$$

and

$$dY^x = \varepsilon^{-2} g(x, Y^x) dt + \varepsilon^{-1} \sigma(x, Y^x) dW_t.$$

Derivation. The backward Kolmogorov equation for $u(x, y, t) = \mathbb{E}_{x,y} f(X)$ is now

$$\frac{\partial u}{\partial t} = \varepsilon^{-1} L_0 u + \varepsilon^{-2} L_1 u. \quad (\text{BKE})$$

Inserting the expansion $u = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + O(\varepsilon^2)$ (we will have to go one order in ε higher than before) in this equation now gives

$$\begin{cases} L_1 u_0 = 0, \\ L_1 u_1 = -L_0 u_0, \\ L_1 u_2 = \frac{\partial u_0}{\partial t} - L_0 u_1, \\ L_1 u_3 = \dots \end{cases}$$

The first equation tells that $u_0(x, y, t) = u_0(x, t)$.

The solvability condition for the second equation is satisfied by assumption because of (Δ) . Therefore this equation can be formally solved as

$$u_1 = -L_1^{-1} L_2 u_0.$$

Insert this expression in the third equation. The solvability condition for the resulting equation gives the limiting equation for u_0 :

$$\frac{\partial u_0}{\partial t} = \bar{L} u_0, \quad (\text{LBKE})$$

where

$$\bar{L} = \int_{\mathbb{R}^m} d\mu_x(y) L_0 L_1^{-1} L_0.$$

To see what this equation is explicitly, notice that $-L_1^{-1}g(y)$ is the steady state solution of

$$\frac{\partial v}{\partial t} = L_1 v + g(y).$$

The solution of this equation with the initial condition $v(y, 0) = 0$ can be represented by Feynman-Kac formula as

$$v(y, t) = \mathbb{E}_y \int_0^t g(Y_s^x) ds,$$

Therefore

$$-L_1^{-1}g(y) = \mathbb{E}_y \int_0^\infty g(Y_t^x) dt,$$

and the operator \bar{L} in the limiting backward Kolmogorov equation (LBKE) of the limiting equation

$$dX = \bar{f}(X)dt + \bar{\sigma}(X) dB_t,$$

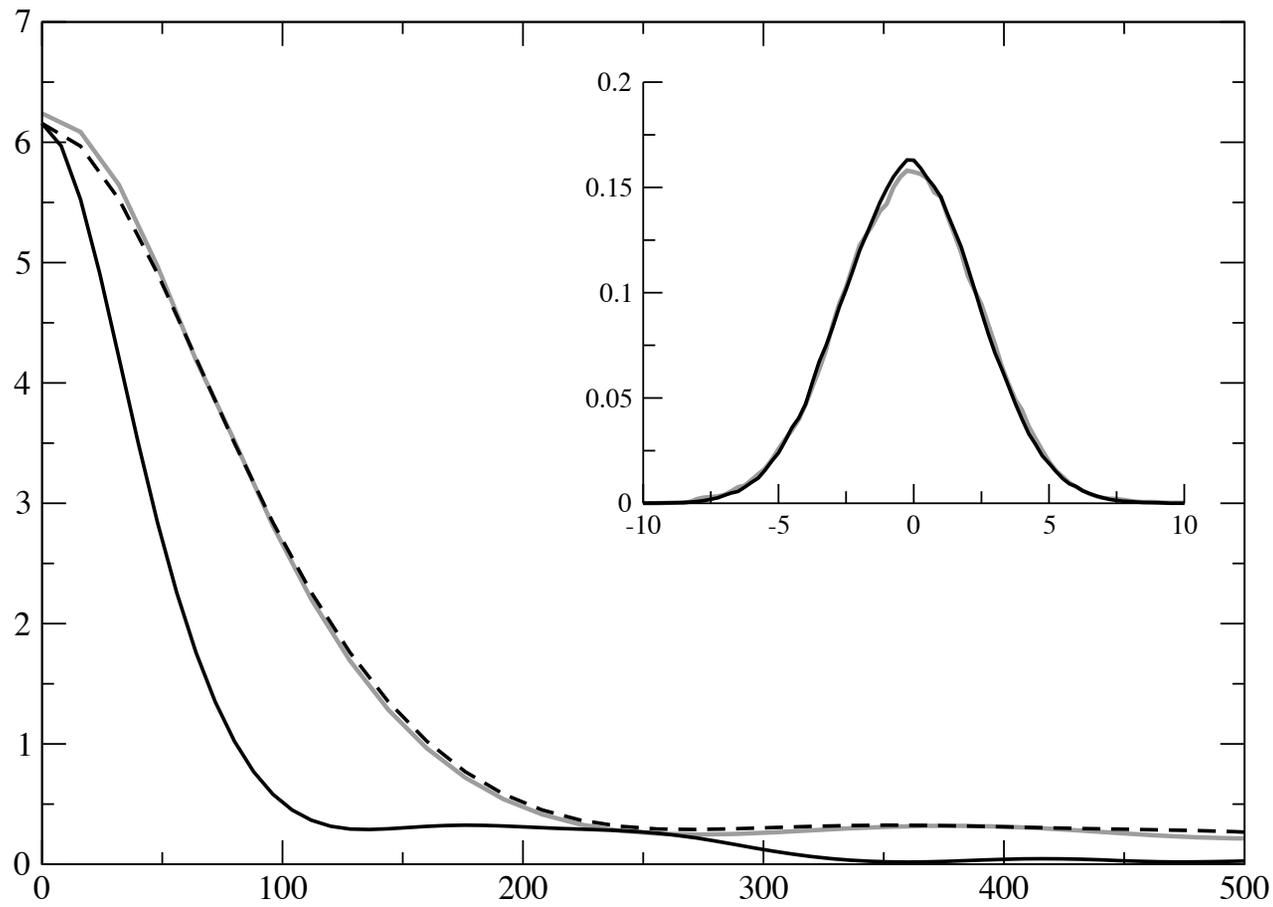
Example: the Lorenz 96 (L96) model

Consider the following modification of (I96)

$$\begin{cases} \dot{X}_k = -\varepsilon (X_{k-1}(X_{k-2} - X_{k+1}) + X_k) + \frac{h_x}{J} \sum_{j=1}^J (Y_{j,k+1} - Y_{j,k-1}) \\ \dot{Y}_{j,k} = \frac{1}{\varepsilon} (-Y_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - Y_{j,k} + F_y) + h_y X_k, \end{cases} \quad (\text{L96}')$$

By symmetry

$$0 = \int_{\mathbb{R}^{KJ}} \left(\frac{h_x}{J} \sum_{j=1}^J (y_{j,k+1} - y_{j,k-1}) \right) d\mu_x(y)$$



ACFs and PDFs (subplot) of the slow variable X_k evolving under (L96'). Grey line: $\varepsilon = 1/256$; full black line: $\varepsilon = 1/128$; dashed black line: $\varepsilon = 1/128$ with time rescaled as $t \rightarrow 2t$. The near perfect match confirms that evolution of X_k converges to some limiting dynamics on the $O(1/\varepsilon)$ time-scale.

Coupled truncated Burgers-Hopf (TBH). I. Stable periodic orbits

$$\left\{ \begin{array}{l} \dot{X}_1 = \frac{1}{\varepsilon} b_1 X_2 Y_1 + a Y_1 (R^2 - (X_1^2 + X_2^2)) - b X_2 (\alpha + (X_1^2 + X_2^2)), \\ \dot{X}_2 = \frac{1}{\varepsilon} b_2 X_1 Y_1 + a X_2 (R^2 - (X_1^2 + X_2^2)) + b X_1 (\alpha + (X_1^2 + X_2^2)), \\ \dot{Y}_k = -\operatorname{Re} \frac{ik}{2} \sum_{p+q+k=0} U_p^* U_q^* + \frac{1}{\varepsilon} b_3 \delta_{1,k} X_1 X_2, \\ \dot{Z}_k = -\operatorname{Im} \frac{ik}{2} \sum_{p+q+k=0} U_p^* U_q^*, \end{array} \right.$$

where $U_k = Y_k + iZ_k$.

Truncated system: stable periodic orbit (limit cycle)

$$(X_1(t), X_2(t)) = R (\cos \omega t, \sin \omega t) \quad \text{with frequency } \omega = b(\alpha + R^2).$$

NB: Truncated Burgers (Majda & Timofeyev, 2000)

Fourier-Galerkin truncation of the inviscid Burgers-Hopf equation,
 $u_t + \frac{1}{2}(u^2)_x = 0$:

$$\dot{U}_k = -\frac{ik}{2} \sum_{\substack{k+p+q=0 \\ |p|,|q|\leq\Lambda}} U_p^* U_q^*, \quad |k| < \Lambda$$

Features common with many complex systems. In particular:

- Display deterministic chaos.
- Ergodic on $E = \sum_k |U_k|^2$.
- Scaling law for the correlation functions with $t_k \approx O(k^{-1})$.

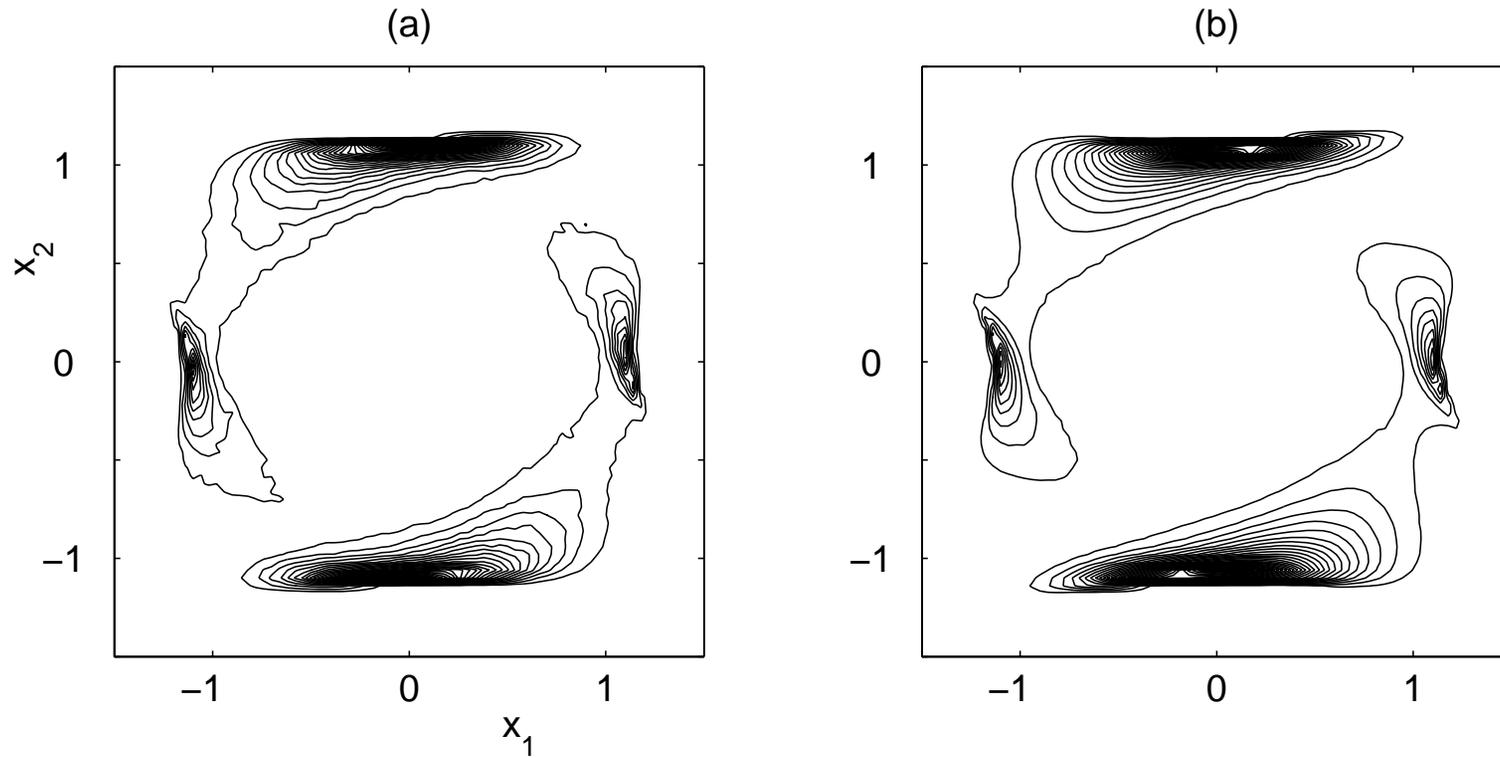
Here: *Used as a model for unresolved modes.*

Couple truncated Burgers with two resolved variables.

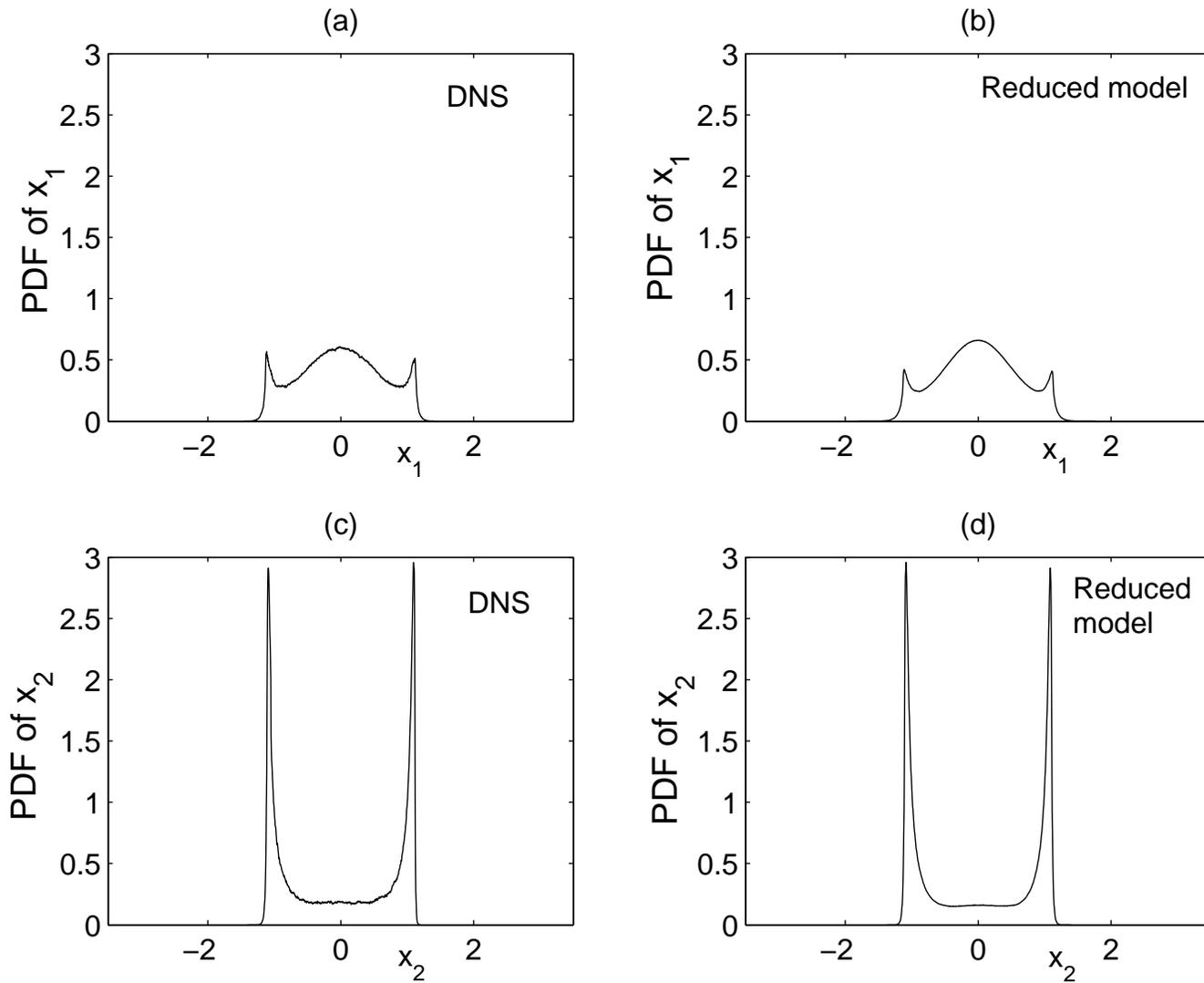
$$\left\{ \begin{array}{l} \dot{X}_1 = \frac{1}{\varepsilon} b_1 X_2 Y_1 + a X_1 (R^2 - (X_1^2 + X_2^2)) - b X_2 (\alpha + (X_1^2 + X_2^2)), \\ \dot{X}_2 = \frac{1}{\varepsilon} b_2 X_1 Y_1 + a X_2 (R^2 - (X_1^2 + X_2^2)) + b X_1 (\alpha + (X_1^2 + X_2^2)), \\ \dot{Y}_k = -\operatorname{Re} \frac{ik}{2} \sum_{p+q+k=0} U_p^* U_q^* + \frac{1}{\varepsilon} b_3 \delta_{1,k} X_1 X_2, \\ \dot{Z}_k = -\operatorname{Im} \frac{ik}{2} \sum_{p+q+k=0} U_p^* U_q^*, \end{array} \right.$$

Limiting SDEs:

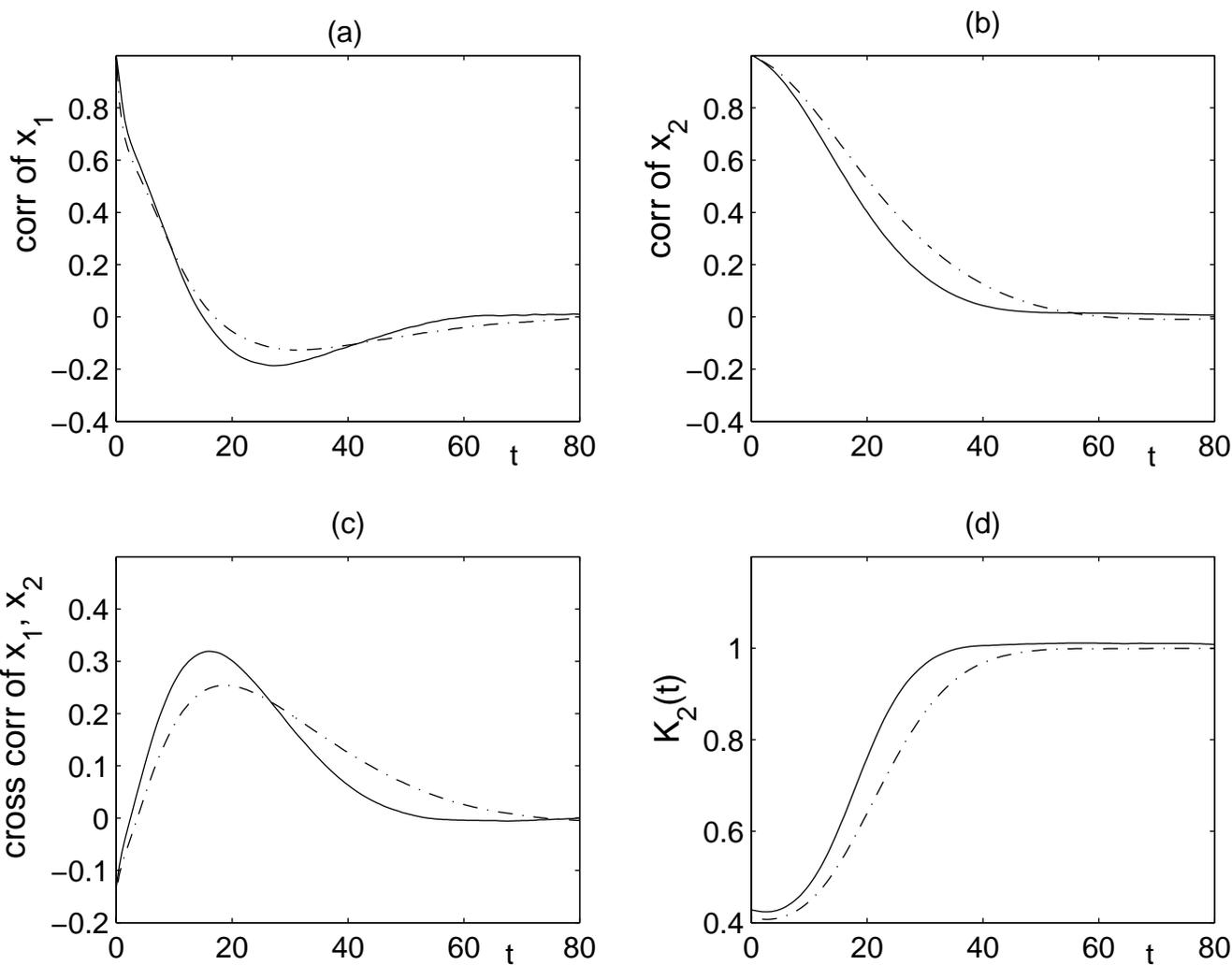
$$\left\{ \begin{array}{l} dX_1 = b_1 b_2 X_1 dt + N_1 X_1 X_2^2 dt + \sigma_1 X_2 dW(t) \\ \quad + a X_1 (R^2 - (X_1^2 + X_2^2)) dt - b X_2 (\alpha + (X_1^2 + X_2^2)) dt, \\ dX_2 = b_1 b_2 X_2 dt + N_2 X_2 X_1^2 dt + \sigma_2 X_1 dW(t) \\ \quad + a X_2 (R^2 - (X_1^2 + X_2^2)) dt + b X_1 (\alpha + X_1^2 + X_2^2) dt, \end{array} \right.$$



Contour plots of the joint probability density for the climate variables X_1 and X_2 ; (a) deterministic system with 102 variables; (b) limit SDE. There only remains a ghost of the limit cycle.



Marginal PDFs of X_1 and X_2 for the simulations of the full deterministic system with 102 variables (DNS) and the limit SDE.



Two-point statistics for X_1 and X_2 ; solid lines - deterministic system with 102 degrees of freedom; dashed lines - limit SDE; (a), (b) correlation functions of X_1 and X_2 , respectively; (c) cross-correlation function of X_1 and X_2 ; (d) normalized correlation of energy,

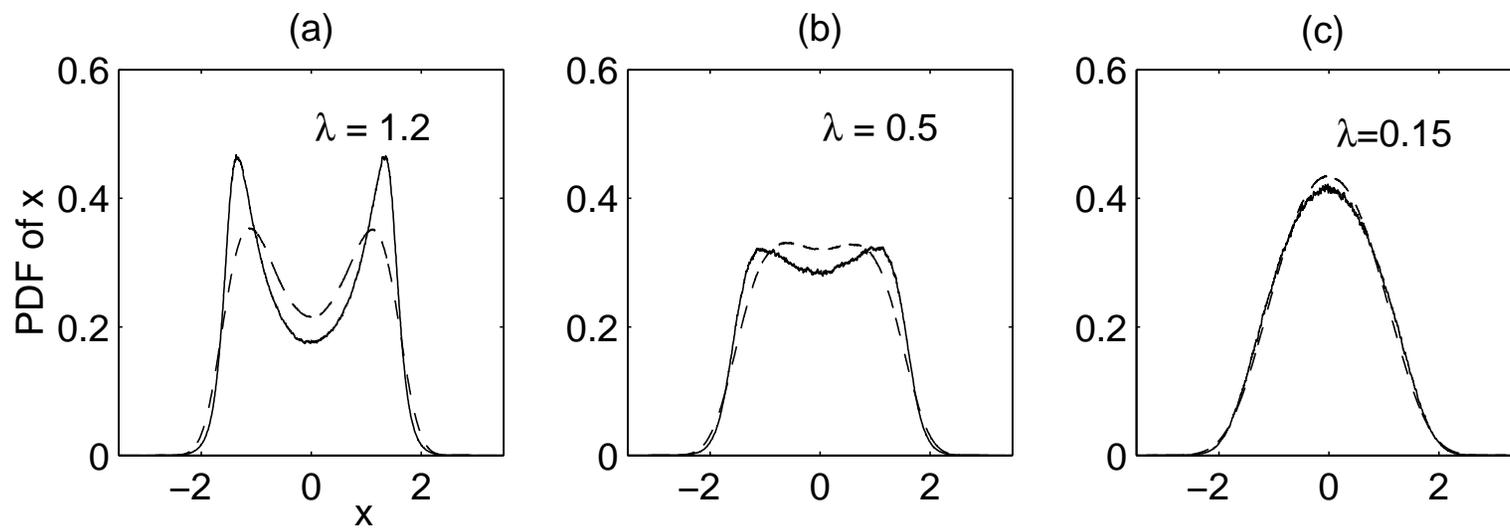
$$K_2(t) = \frac{\langle X_2^2(t)X_2^2(0) \rangle}{\langle X_2^2 \rangle^2 + 2\langle X_2(t)X_2(0) \rangle^2}$$

Coupled truncated Burgers-Hopf (TBH). II. Multiple equilibria

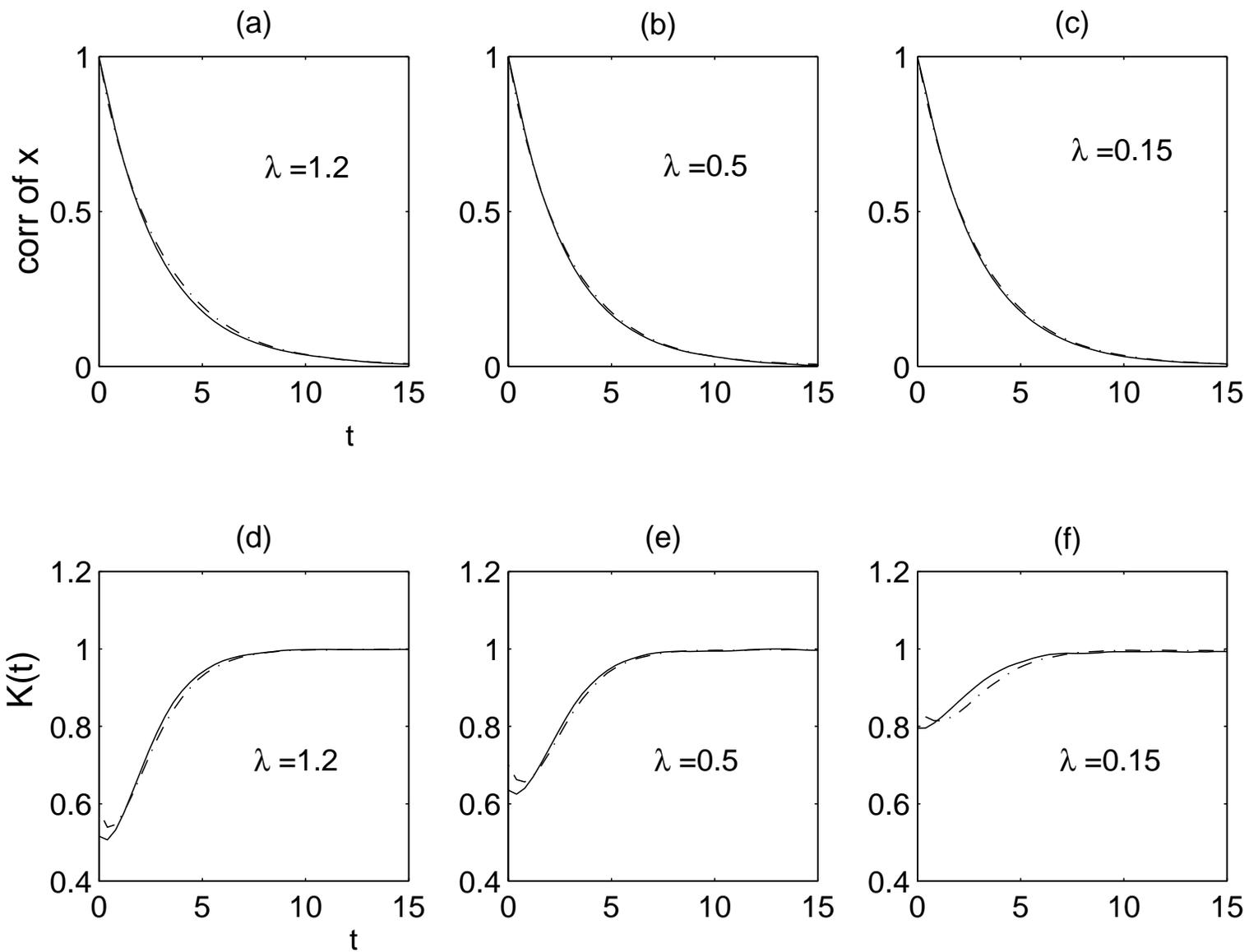
$$\left\{ \begin{array}{l} \dot{X} = \frac{1}{\varepsilon} b_1 Y_1 Z_1 + \lambda(1 - \alpha x^2)x, \\ \dot{Y}_k = -\operatorname{Re} \frac{ik}{2} \sum_{p+q+k=0} U_p^* U_q^* + \frac{1}{\varepsilon} b_2 \delta_{1,k} X Z_k, \\ \dot{Z}_k = -\operatorname{Im} \frac{ik}{2} \sum_{p+q+k=0} u_p^* u_q^* + \frac{1}{\varepsilon} b_3 \delta_{1,k} X Y_k, \end{array} \right.$$

Limiting SDE:

$$dX = -\gamma X dt + \sigma dW(t) + \lambda(1 - \alpha X^2)X dt$$



PDF of X for the simulations of the deterministic model (solid lines) and the limit SDE (dashed lines) in three regimes, $\lambda = 1.2, 0.5, 0.15$.



Two-point statistics of X in three regimes, $\lambda = 1.2, 0.5, 0.15$ for the deterministic equations (solid lines) and limit SDE (dashed lines) (a), (b), (c) correlation function of x ; (d), (e), (f) correlation of energy, $K(t)$, in x .

Generalizations

Let $Z_t \in S$ be the sample path of a continuous-time Markov process with generator

$$L = L_0 + \varepsilon^{-1} L_1$$

and assume that L_1 has several ergodic components indexed by $x \in S'$ with equilibrium distribution

$$\mu_x(z)$$

Then, as $\varepsilon \rightarrow 0$ there exists a limiting on S' with generator

$$\bar{L} = \mathbb{E}_{\mu_x} L_0$$

Similar results available on diffusive time-scales.

Can be applied to SDEs, Markov chains (i.e. discrete state-space), deterministic systems (periodic or chaotic), etc.

Other situations with limiting dynamics? Mori-Zwanzig projection

Consider

$$\begin{cases} \dot{X} = f(X, Y), \\ \dot{Y} = g(X, Y), \end{cases}$$

and denote by $\varphi(X_{[0,t]})$ the solution of the equation for Y at time t assuming that X is known on $[0, t]$

Observe that $\varphi(X_{[0,t]})$ is a functional of $\{X_s, s \in [0, t]\}$.

Then X satisfies the closed equation

$$\dot{X} = f(X, \varphi(X_{[0,t]}))$$

In general, X is not Markov!

It becomes Markov when Y is faster, or ...?

Other possibility: Weak coupling

The system

$$\begin{cases} \dot{X} = \frac{1}{N} \sum_{n=1}^N f(X, Y^n), \\ \dot{Y}^n = g(X, Y^n), \end{cases} \quad (\text{all the } Y^n \text{ coupled only via } X)$$

may have a limit behavior as $N \rightarrow \infty$

Example: Kac-Zwanzig model with Hamiltonian

$$H(q, p) = \sum_{j=0}^N p_j^2 / (2m_j) + V(q_0) + \sum_{j=1}^N \alpha_j (q_j - q_0)^2$$

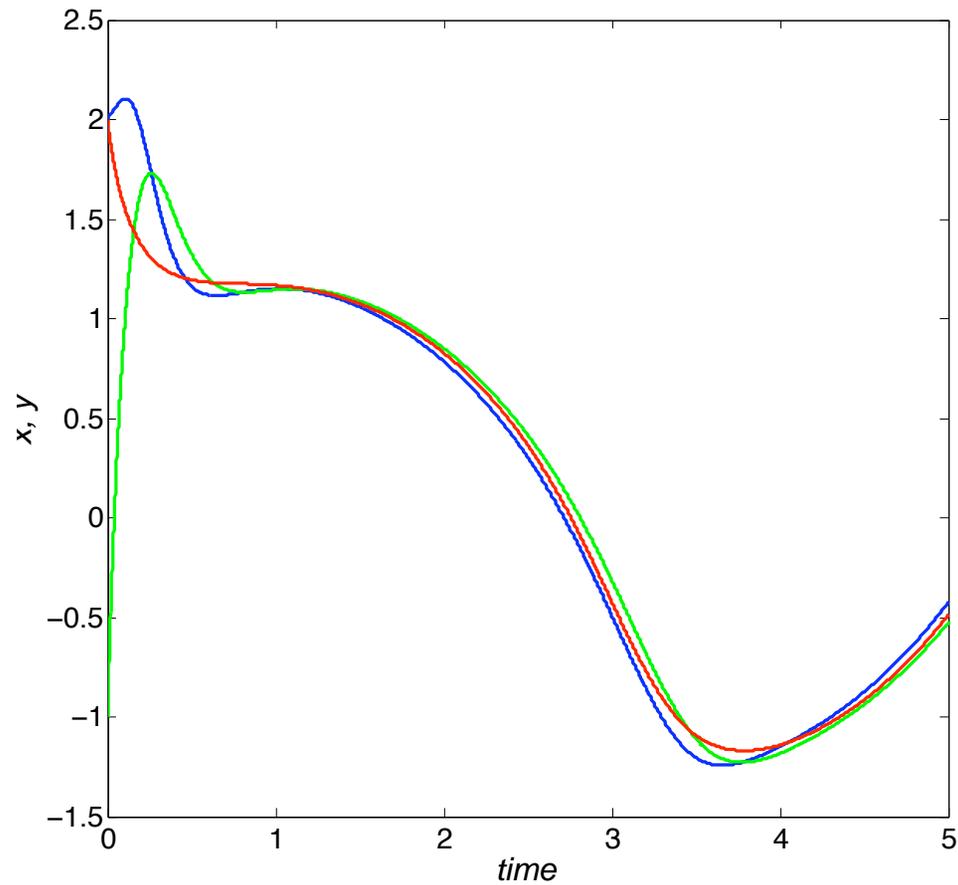
With specific choices of $\{m_j, \alpha_j\}_{j=1, \dots, N}$, there exists a closed SDE for $(q_0, \dot{q}_0, \ddot{q}_0)$ as $N \rightarrow \infty$

Lecture 2: Numerical aspects

Implicit schemes: why they work for stiff ODEs, why they don't for rapidly oscillatory and stochastic systems.

HMM-like integrators: based on averaging principle.

Boosting method: an new seamless extrapolation method for stochastic systems.



The solution of

$$\begin{cases} \dot{X} = -Y^3 + \cos(t) + \sin(\sqrt{2}t) \\ \dot{Y} = -\varepsilon^{-1}(Y - X) \end{cases}$$

when $\varepsilon = 0.1$ and we took $X(0) = 2$, $Y(0) = -1$. X is shown in blue, and Y in green. Also shown in red is the solution of the limiting equation

$$\dot{X} = -X^3 + \cos(t) + \sin(\sqrt{2}t)$$

Consider the stiff ODE

$$\begin{cases} \dot{X} = f(X, Y) \\ \dot{Y} = -\varepsilon^{-1}(Y - \phi(X)) \end{cases} \quad (\star)$$

If $\varepsilon \ll 1$, Y is very fast and it will adjust rapidly to the current value of X , i.e. after short $O(\varepsilon)$ transient we will have

$$Y = \phi(X) + O(\varepsilon) \text{ at all times.}$$

Then the equation for slow variables X reduces to

$$\dot{X} = f(X, \phi(X)) \quad (\star\star)$$

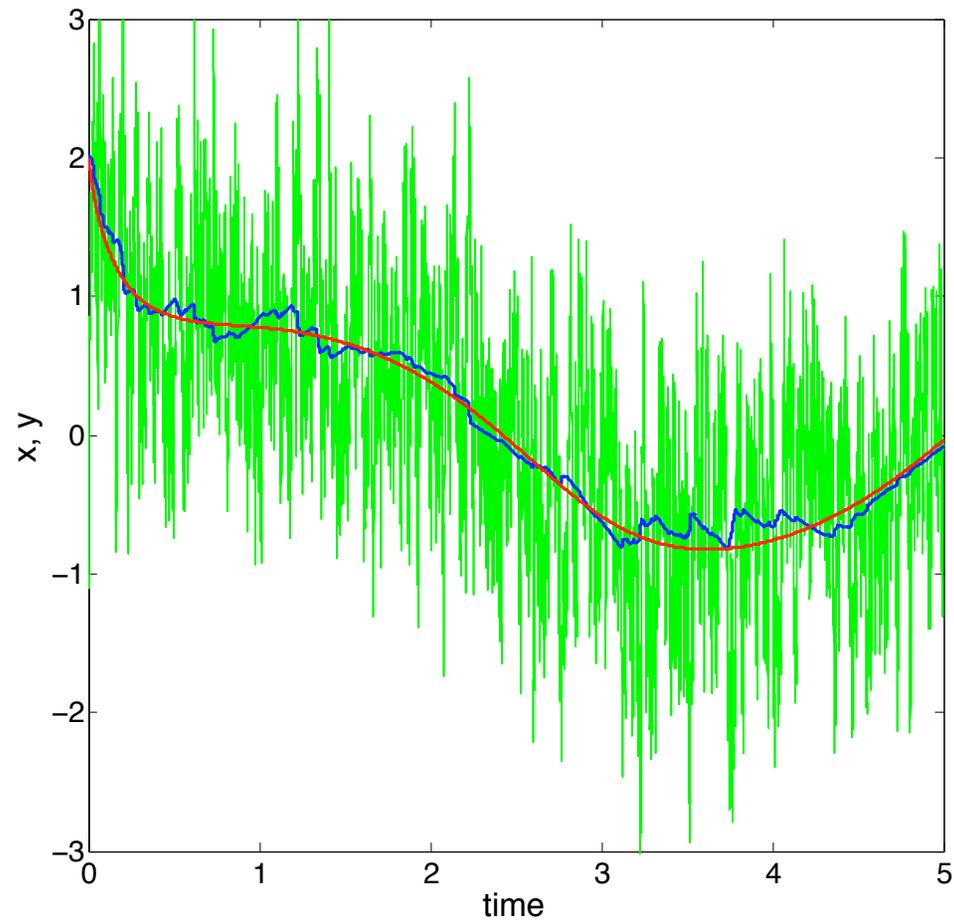
Can be integrated efficiently using an **implicit scheme**, like e.g.:

$$\begin{cases} X^{n+1} = X^n + \Delta t f(X^{n+1}, Y^{n+1}) \\ Y^{n+1} = Y^n - \frac{\Delta t}{\varepsilon}(Y^{n+1} - \phi(X^{n+1})) \end{cases}$$

When $\varepsilon \ll \Delta t \ll 1$:

$$\begin{cases} Y^{n+1} = \phi(X^n) + O(\Delta t) + O(\varepsilon), \\ X^{n+1} = X^n + \Delta t f(X^n, \phi(X^n)) + O(\Delta t^2) + O(\varepsilon) \end{cases}$$

However: *Implicit schemes are ill-suited for rapidly oscillatory or stochastic systems!*



The solution of

$$\begin{cases} \dot{X} = -Y^3 + \cos(t) + \sin(\sqrt{2}t) \\ dY = -\varepsilon^{-1}(Y - X)dt + \varepsilon^{-1/2}dW \end{cases}$$

when $\varepsilon = 0.01$ and $X(0) = 2, Y(0) = -1$. X is shown in blue, and Y in green. Also shown in red is the solution of the limiting equation

$$\dot{X} = -X^3 + X + \cos(t) + \sin(\sqrt{2}t)$$

Notice how noisy Y is.

Recall key limit thm: Consider

$$\begin{cases} \dot{X} = f(X, Y), \\ dY = \varepsilon^{-1}g(X, Y)dt + \varepsilon^{-1/2}\sigma(X, Y)dW(t), \end{cases} \quad (\star)$$

Assume that: (i) the evolution Y at every fixed $X = x$ is ergodic with respect to the probability distribution

$$d\mu_x(y)$$

and (ii)

$$F(x) = \int_{\mathbb{R}^m} f(x, y)d\mu_x(y) \quad \text{exists}$$

Then in the limit as $\varepsilon \rightarrow 0$ the evolution for X solution of (\star) is governed by

$$\dot{X} = F(X)$$

In addition

$$F(x) = \int_{\mathbb{R}^m} f(x, y)d\mu_x(y) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(x, Y_t^x)dt$$

where

$$dY^x = \varepsilon^{-1}g(x, Y^x)dt + \varepsilon^{-1/2}\sigma(x, Y^x)dW(t)$$

Basic HMM-like integrator

Use

$$X^{n+1} = X^n + \Delta t \tilde{F}^n, \quad X_0 = X(t=0)$$

Here \tilde{F}^n is an approximation of $F(X^n)$ obtained as

$$\tilde{F}^n = \frac{1}{M_T} \sum_{m=M}^{M+M_T} f(X^n, Y^{n,m})$$

where

$$Y^{n,m+1} = Y^{n,m} + \frac{\delta t}{\varepsilon} g(X^n, Y^{n,m}) + \sqrt{\frac{\delta t}{\varepsilon}} \sigma(X^n, Y^{n,m}) \xi^m,$$
$$Y_{0,0} = Y(t=0), \quad Y^{n+1,0} = Y^{n,M+M_T-1}$$

Why is this better?

Basically, because M and M_T are $O(1)$ in ε ! In other words, one can reach an $O(1)$ time-scale with a $O(1)$ number of steps.

In contrast, the direct scheme

$$\begin{cases} X^{p+1} = X^p + \delta t f(X^p, Y^p), \\ Y^{p+1} = Y^p + \frac{\delta t}{\varepsilon} g(X^p, Y^p), \end{cases}$$

takes $O(\varepsilon^{-1})$ steps to reach an $O(1)$ time-scale.

Error estimate (pessimistic):

Thm: For any $T > 0$, there exists a constant $C > 0$ such that

$$\mathbb{E} \left(\sup_{0 \leq n \leq T/\Delta t} |X(n\Delta t) - X^n| \right) \leq C \left(\sqrt{\varepsilon} + (\Delta t)^k + (\delta t/\varepsilon)^l + \sqrt{\frac{\varepsilon \Delta t}{M_T \delta t + 1}} \right)$$

Efficiency depends on how large M_T and M needs to be.

Recall that

$$\tilde{F}^n = \frac{1}{M} \sum_{m=M_T}^{M+M_T-1} f(X^n, Y^{n,m})$$

M tampers the relaxation errors;

M_T controls the size of the time-averaging window.

The nice surprise: the HMM-like multiscale integrator works even if $M_T = 1$ (no time averaging) provided only that

$$\varepsilon \ll \frac{\varepsilon \Delta t}{M \delta t} \ll 1$$

The factor $\lambda = \Delta t / M \delta t$ also gives the efficiency boost the HMM-like multiscale integrators over a direct scheme.

Why is time-averaging unnecessary (since in this case the approximation on \tilde{F}^n is very bad at each time step)?

A simple illustrative example:

$$\begin{cases} \dot{X} = -Y \\ dY = -\frac{1}{\varepsilon}(Y - X)dt + \frac{1}{\sqrt{\varepsilon}}dW(t) \end{cases}$$

Here

$$d\mu_x(y) = \frac{e^{-(y-x)^2}}{\sqrt{\pi}}dy$$

and so the limiting equation is

$$\dot{X} = -X$$

To understand the effect of (not) averaging, use

$$\begin{aligned} X^{n+1} &= X^n - \Delta t \left(X^n + \frac{\xi^n}{\sqrt{2}} \right), \\ \xi^n &= \text{independent } N(0, 1) \end{aligned}$$

Then

$$X^n = x(1 - \Delta t)^n + \frac{\Delta t}{\sqrt{2}} \sum_{j=1}^{n-1} (1 - \Delta t)^j \xi^{n-j}$$

and

$$\mathbb{E}|X^n - x(1 - \Delta t)^n|^2 = \frac{\Delta t^2}{2} \sum_{j=1}^{n-1} (1 - \Delta t)^{2j} = O(\Delta t) \quad [n = O(\Delta t^{-1})]$$

Same example, including relaxation errors:

$$\begin{cases} \dot{X} = -Y \\ dY = -\frac{1}{\varepsilon}(Y - X)dt + \frac{1}{\sqrt{\varepsilon}}dW(t) \end{cases} \quad (*)$$

Scheme:

1. Use:

$$Y^{n,m+1} = Y^{n,m} - \frac{\delta t}{\varepsilon}(Y^{n,m} - X^n) + \sqrt{\frac{\delta t}{\varepsilon}}\xi^{n,m}, \quad Y^{n,0} = Y^{n-1,M}$$

2. Use

$$X^{n+1} = X^n - \Delta t Y^{n,M}$$

3. Repeat.

Denoting $\lambda = \Delta t/M\delta t$, this integrator approximates

$$\begin{cases} \dot{X} = -Y \\ dY = -\frac{1}{\lambda\varepsilon}(Y - X)dt + \frac{1}{\sqrt{\lambda\varepsilon}}dW(t) \end{cases} \quad (**)$$

Hence, by the limit theorem, it approximates (*) and boost the efficiency by λ if

$$\varepsilon \ll \varepsilon\lambda \ll 1 \quad \text{or equivalently} \quad M\delta t \ll \Delta t \ll \varepsilon^{-1}M\delta t$$

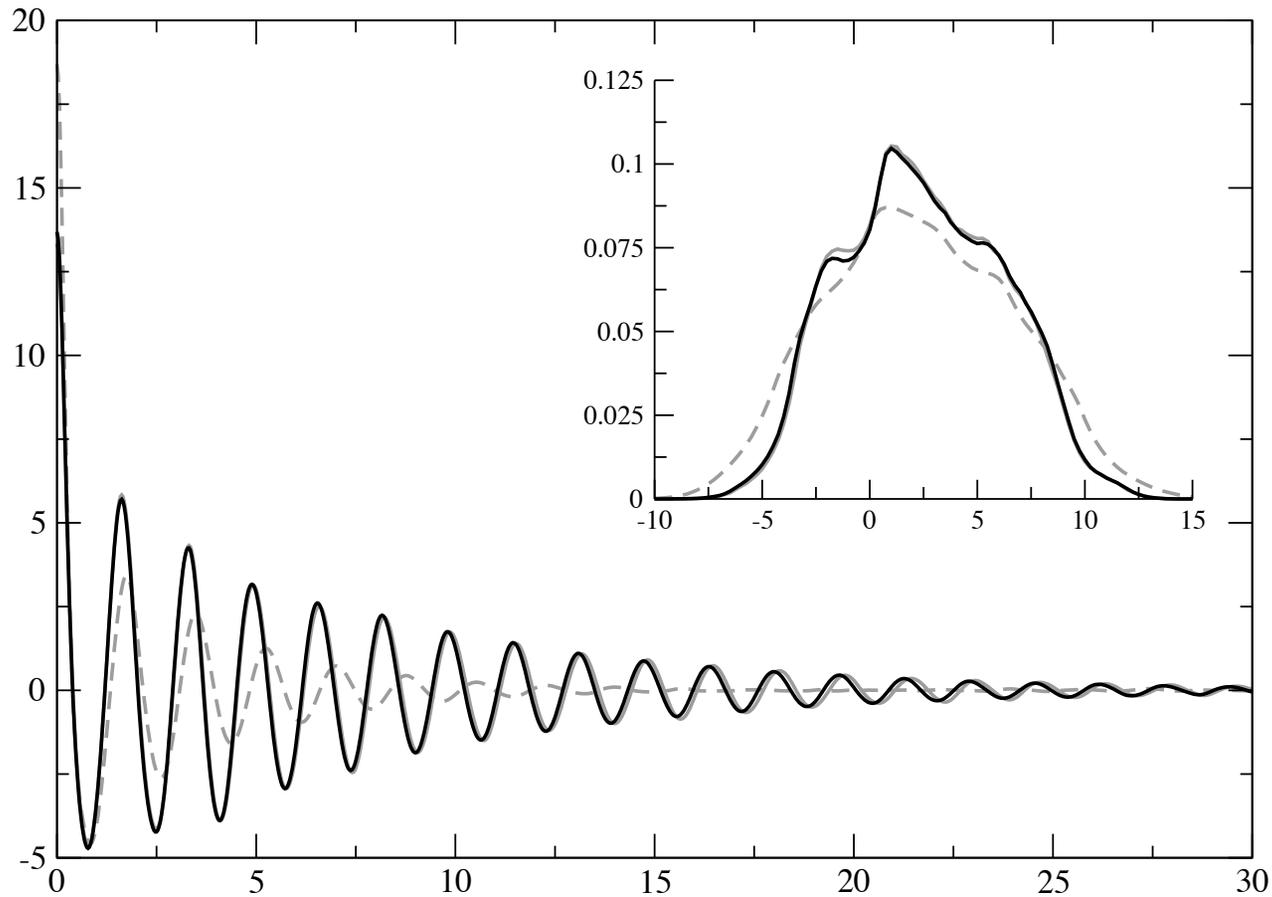
Example: the Lorenz 96 (L96) model

$$\begin{cases} \dot{X}_k = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F_x + \frac{h_x}{J} \sum_{j=1}^J Y_{j,k} \\ \dot{Y}_{j,k} = \frac{1}{\varepsilon} (-Y_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - Y_{j,k} + h_y X_k) . \end{cases} \quad (\text{L96})$$

with $F_x = 10$, $h_x = -0.8$, $h_y = 1$, $K = 9$, $J = 8$, and $\varepsilon = 1/128$.

M_T	R	Δt	Gain	ν	error(%)	ω	error(%)
$\varepsilon = 2^{-7}$		2^{-11}	—	0.135	—	3.81	—
<i>Truncated</i>		2^{-6}	—	0.287	113	3.57	6.3
1	4	2^{-4}	32	0.201	49	3.76	1.3
1	2	2^{-5}	32	0.171	27	3.89	2.1
1	1	2^{-6}	32	0.162	20	3.88	1.9
2	1	2^{-5}	32	0.162	20	3.88	1.9
4	1	2^{-4}	32	0.158	17	3.87	1.6
1	1	2^{-7}	16	0.137	2	3.84	0.8
2	1	2^{-7}	8	0.135	0	3.83	0.5

Gain in efficiency of the multiscale scheme over a direct solver for various values of the control parameters in the multiscale algorithm. The error and the gain are calculated relative to the simulation with $\varepsilon = 2^{-7}$. The parameters ν and ω are obtained by fitting by $C_0 \cos(\omega t)e^{-\nu t}$ the ACF produced by the simulations.

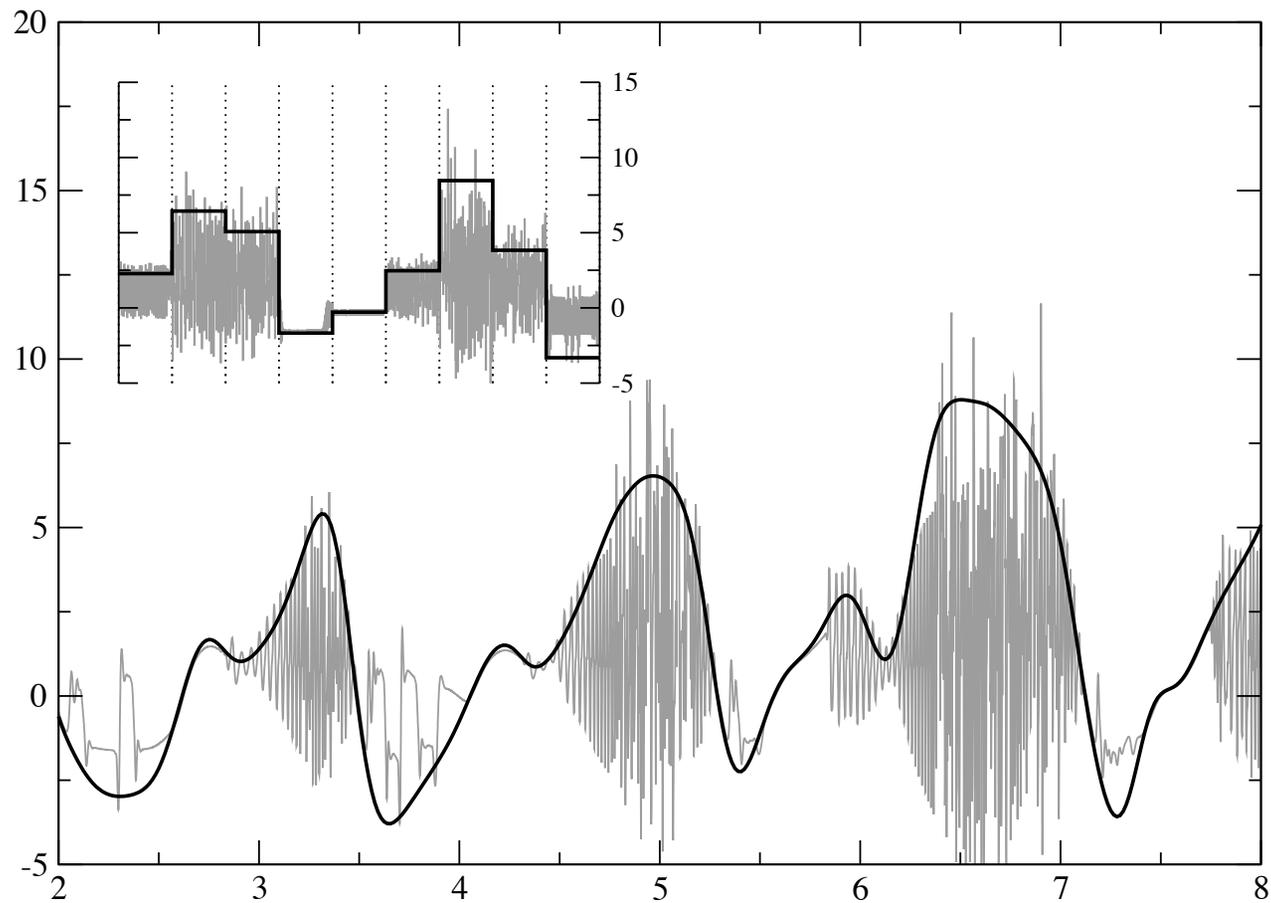


Comparison between the ACFs and PDFs; black line: multiscale solver; full grey line: direct solver with $\varepsilon = 1/128$. Efficiency gain: 16. Also shown in dashed grey are the corresponding ACF and PDF produced by the truncated dynamics where the coupling of the slow modes X_k with the fast ones, $Y_{j,k}$, is artificially switched off.

Example: L96 model with space-time scale separation

$$\begin{cases} \dot{X}_k = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F_x + \frac{h_x}{J} \sum_{j=1}^J Y_{j,k} \\ \dot{Y}_{j,k} = \frac{1}{\varepsilon} (-Y_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - Y_{j,k} + h_y X_k). \end{cases} \quad (\text{sL96})$$

with $F_x = 10$, $h_x = -0.8$, $h_y = 1$, $K = 9$, and $J = 1/\varepsilon = 128$.



Typical time-series of the slow (black line) and fast (grey line) modes; $K = 9$, $J = 1/\varepsilon = 128$. The subplot displays a typical snapshot of the slow and fast modes at a given time.

Working assumption: spatial interaction between the $Y_{j,k}$ are sufficiently weak and short-range on the average.

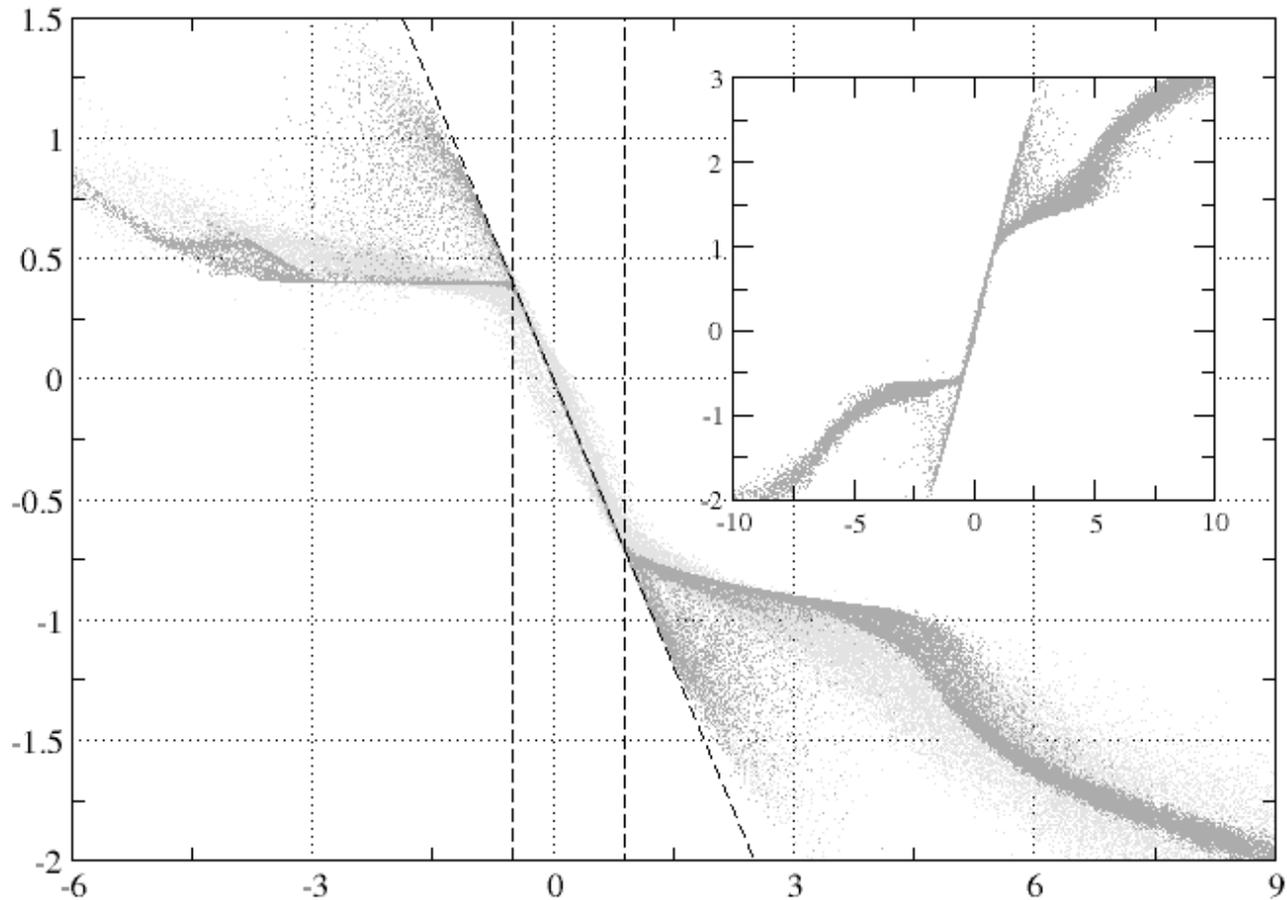
Then, at any given time, the term $(h_x/J) \sum_{j=1}^J Y_{j,k}$ self-averages in the limit as $J \rightarrow \infty$ (i.e. $\varepsilon \rightarrow 0$ since $J = \lfloor 1/\varepsilon \rfloor$) to a limit which, by the law of large numbers, satisfies

$$\lim_{\varepsilon \rightarrow 0} \varepsilon h_x \sum_{j=1}^{\lfloor 1/\varepsilon \rfloor} Y_{j,k} = h_x \mathbb{E}_{X_k} Y_{j,k} \equiv F(X_k),$$

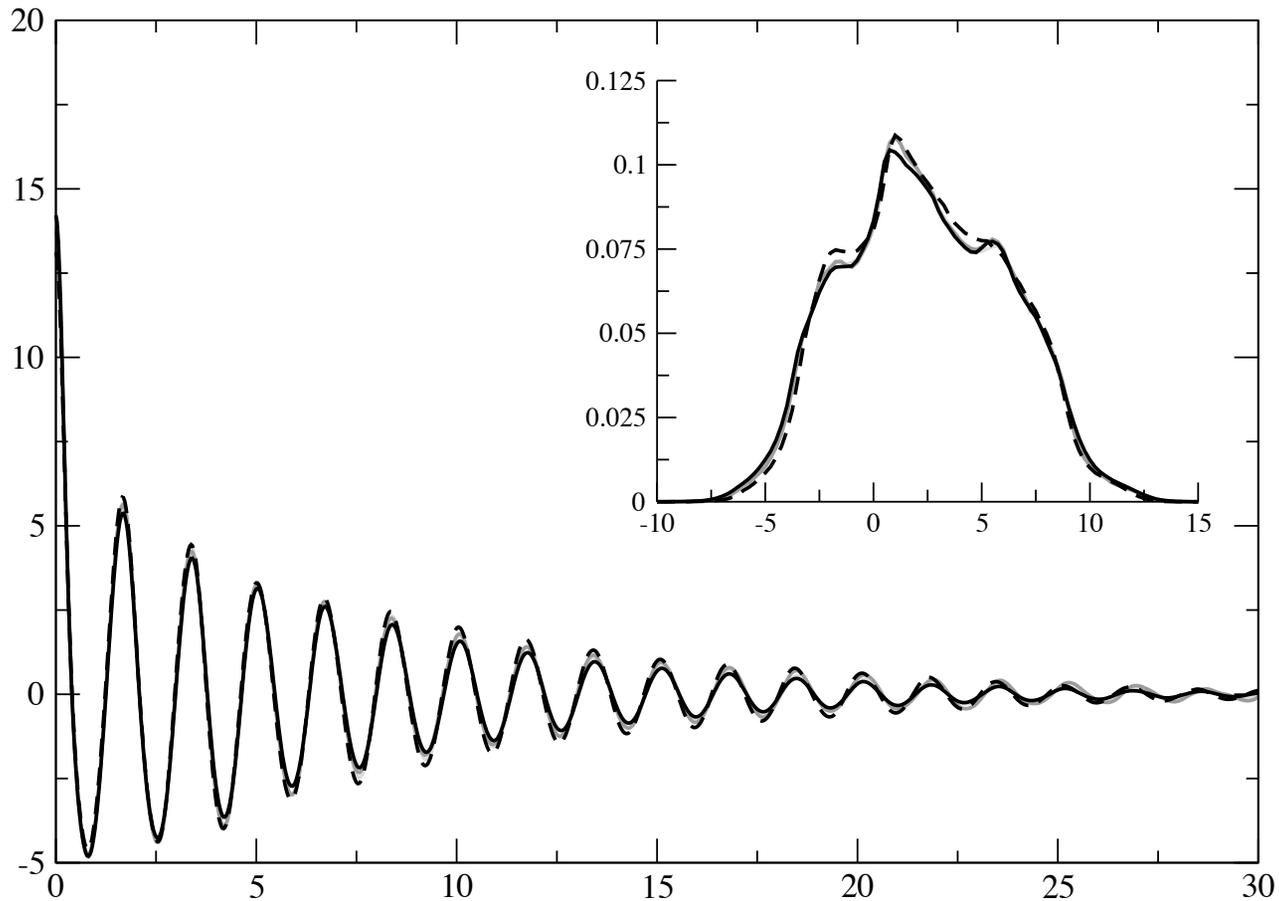
Here $\mathbb{E}_{X_k} Y_{j,k}$ is the conditional average of any given $Y_{j,k}$ at fixed X_k .

Notice that this implies $F(X_k)$ instead of $F(X_1, \dots, X_K)$.

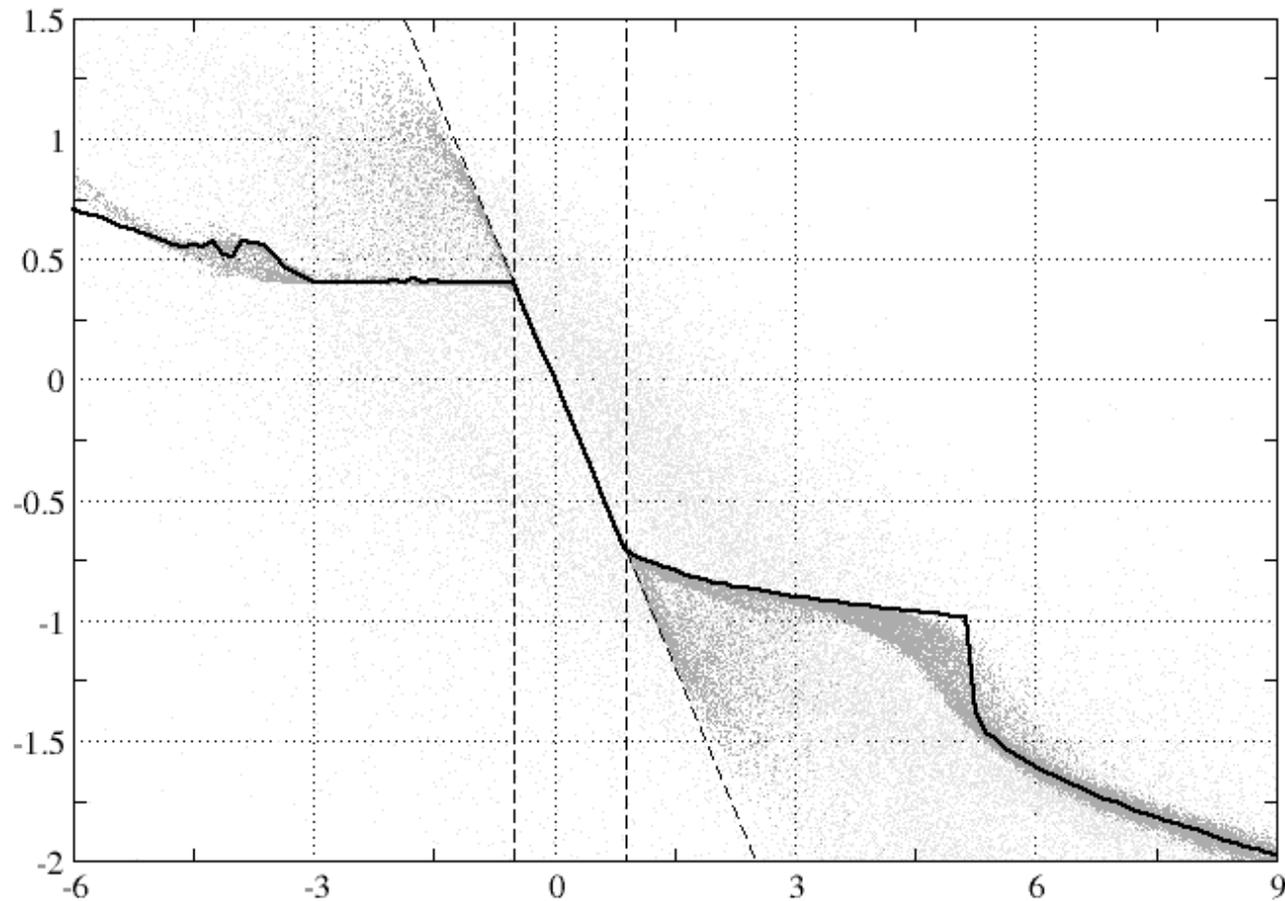
This assumption is corroborated by numerical experiments.



Scatterplots of the bare forcing $\varepsilon h_x \sum_{j=1}^{\lfloor 1/\varepsilon \rfloor} Y_{j,k}$ (no averaging here); light grey: $J = 1/\varepsilon = 128$; dark grey: $J = 1/\varepsilon = 4096$ ($K = 9$). The sharpness of these graphs confirm that bare forcing self-averages consistent with (??). The subplot: $J = 1/\varepsilon = 1024$, $h_x = 1.2$ (instead of $h_x = -0.8$ taken otherwise); it can be obtained by rescaling the forcing in the main plot respectively by $1.2/(-0.8)$.



Comparison of ACFs and PDFs (subplot) of the slow mode for various simulations in $J = 1/\varepsilon$ regime. Light grey: $J = 128$; dark grey: $J = 256$ (practically indistinguishable from the previous one); solid black: result from the multiscale scheme with tabulated forcing ($J = 16$; gain = ∞); dashed black: result from the multiscale scheme with $J = 8$, $\Delta t = 2^{-7}$, $M_T = R = 1$ (gain = 256).



Dark grey: scatterplots of the bare forcing $\varepsilon h_x \sum_{j=1}^{\lfloor 1/\varepsilon \rfloor} Y_{j,k}$ with $J = 1/\varepsilon = 4096$ ($K = 9$). Light grey: effective forcing $F_k(X)$ produced on-the-fly by the multiscale scheme with $K = 9$, $J = 8$, $M_T = 1$, $R = 64$, $\Delta t = 2^{-7}$. Solid black line: tabulated effective forcing computed via the multiscale scheme with $K = 9$, $J = 16$.

Warning: The problem of hidden slow variables

$$\begin{cases} \dot{X}_k = -X_{k-1}(X_{k-2} - X_{k+1}) - \frac{1}{\sqrt{J}} \sum_{j=1}^J (Y_{j,k+1}^2 - Y_{j,k-1}^2) \\ \dot{Y}_{j,k} = -\frac{1}{\varepsilon} Y_{j+1,k} (Y_{j+2,k} - Y_{j-1,k}) - \frac{1}{\sqrt{J}} Y_{j,k} (X_{k+1} - X_{k-1}). \end{cases} \quad (\text{hL96})$$

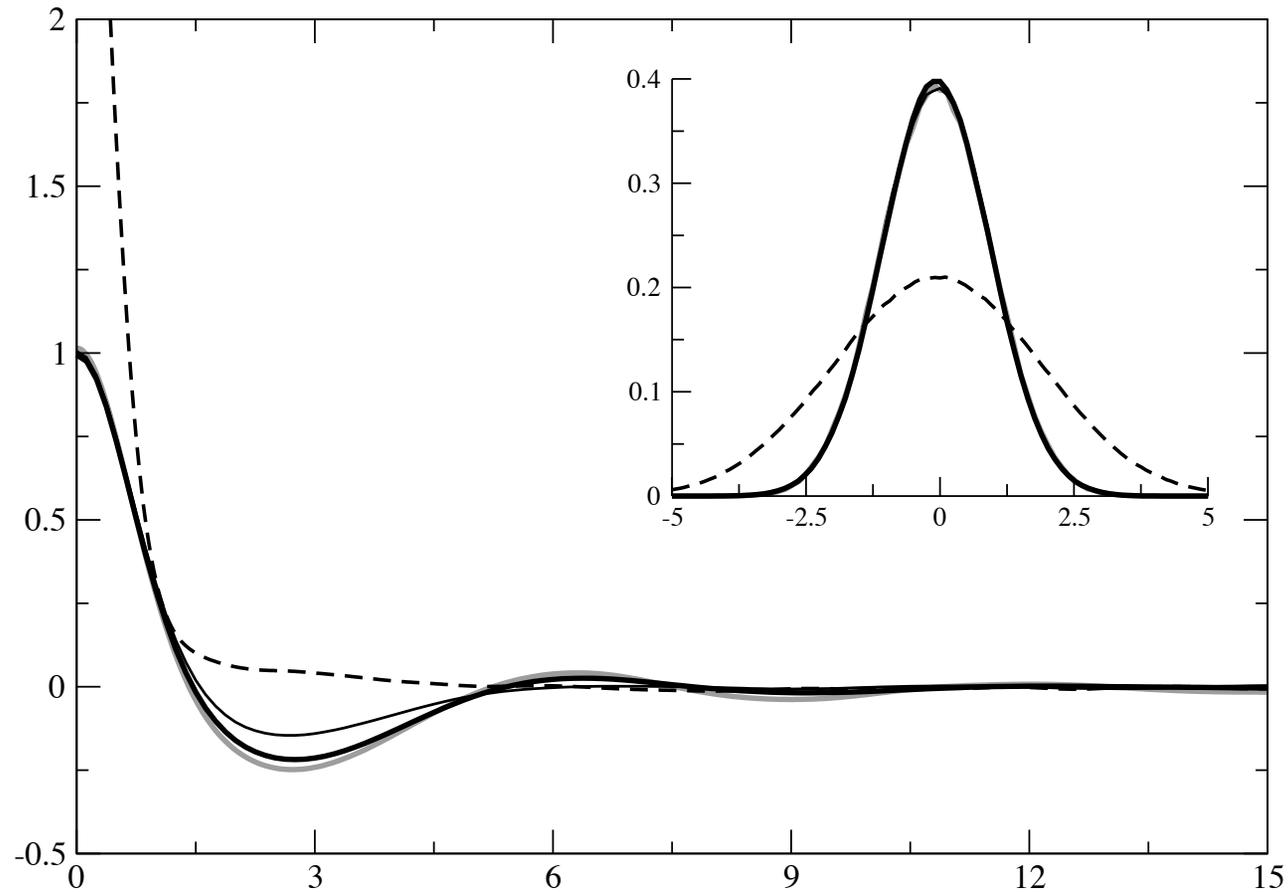
with $J = 1/\varepsilon$.

The following is an additional slow variables (hidden in (hL96)):

$$\bar{B}_k = \frac{1}{\sqrt{J}} \sum_{j=1}^J Y_{j,k}^2.$$

Due to the absence of forcing and damping in (hL96), this equation conserves the energy

$$E = \sum_{k=1}^K \left(X_k^2 + \sum_{j=1}^J Y_{j,k}^2 \right).$$



ACFs and PDFs of X_k . Grey line: direct simulation with $\varepsilon = 1/64$, $J = 512$ ($\delta t = 2^{-10}$); thick solid black line: multiscale scheme with $\Delta t = 2^{-8}$, $J = 32$ (efficiency gain: 64); thin black line: multiscale with $\Delta t = 2^{-7}$, $J = 8$ (efficiency gain: 512). Dashed black line: multiscale scheme with $\Delta t = 2^{-8}$, $J = 32$ (efficiency gain: 64) where the hidden slow variables B_k are not accounted for. The discrepancy clearly indicates that accounting for the B_k 's is necessary. In all the multiscale computations $M_T = R = 1$.

Toward seamless multiscale integrators?

$$\begin{cases} \dot{X} = -Y \\ dY = -\frac{1}{\varepsilon}(Y - X)dt + \frac{1}{\sqrt{\varepsilon}}dW(t) \end{cases} \quad (\star)$$

Recall the scheme:

1. Use:

$$Y^{n,m+1} = Y^{n,m} - \frac{\delta t}{\varepsilon}(Y^{n,m} - X^n) + \sqrt{\frac{\delta t}{\varepsilon}}\xi^{n,m}, \quad Y^{n,0} = Y^{n-1,M}$$

2. Use

$$X^{n+1} = X^n - \Delta t Y^{n,M}$$

3. Repeat.

Denoting $\lambda = \Delta t/M\delta t$, this integrator approximates

$$\begin{cases} \dot{X} = -Y \\ dY = -\frac{1}{\lambda\varepsilon}(Y - X)dt + \frac{1}{\sqrt{\lambda\varepsilon}}dW(t) \end{cases} \quad (\star\star)$$

Hence, by the limit theorem, it approximates (\star) and boost the efficiency by λ if

$$\varepsilon \ll \varepsilon\lambda \ll 1 \quad \text{or equivalently} \quad M\delta t \ll \Delta t \ll \varepsilon^{-1}M\delta t$$

Seamless multiscale algorithm – Boosting method

Consider

$$dZ = \frac{1}{\varepsilon} h_1(Z) dt + \frac{1}{\sqrt{\varepsilon}} h_2(Z) dW(t) + h_3(Z) dt + h_4(Z) d\bar{W}(t)$$

and assume that there exist slow variables $X = \phi(Z)$ which satisfies

$$dX = F(X) dt + G(X) dW(t) \quad \text{as } \varepsilon \rightarrow 0$$

1. Use

$$Z^{n,m+1} = Z^{n,m} + \frac{\delta t}{\varepsilon} h_1(Z^{n,m}) + \sqrt{\frac{\delta t}{\varepsilon}} h_2(Z^{n,m}) \xi^{n,m},$$

2. Use

$$Z^{n+1,0} = Z^{n,M} + \Delta t h_3(Z^{n,M}) + \sqrt{\Delta t} h_4(\tilde{Z}^{n,M}) \eta^n$$

3. Repeat.

Seamless in that one does not need to know the slow variables X are. Works because it approximates

$$dZ = \frac{1}{\lambda\varepsilon} h_1(Z) dt + \frac{1}{\sqrt{\lambda\varepsilon}} h_2(Z) dW(t) + h_3(Z) dt + h_4(Z) d\bar{W}(t)$$

with $\lambda = \Delta t / M \delta t$.

Similar in spirit to Chorin's artificial compressibility method and the Car-Parrinello method in molecular dynamics.

More sophisticated seamless integrators?

Reformulating the main theorem:

Consider the system

$$\dot{Z}_t = \frac{1}{\varepsilon} f(Z_t) + g(Z_t)$$

for some variable $Z_t \in \mathbb{R}^n$. Assume that there exists a vector valued function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ($m < n$) such that:

1. We have

$$f(z) \cdot \nabla \varphi(z) = 0;$$

2. The dynamics

$$\dot{Z}_t^x = f(Z_t^x)$$

is ergodic on every component indexed by $\varphi(z) = x \in \mathbb{R}^m$ with respect to the equilibrium distribution $d\mu_x(z)$.

Then $X_t = \varphi(Z_t)$ are slow variables satisfying the following equation

$$\dot{X}_t = H(X_t)$$

where

$$H(x) = \int_{\mathbb{R}^n} g(z) \cdot \nabla \varphi(z) d\mu_x(z).$$

The previous result is more general, but unfortunately it does not lead to a seamless multiscale algorithm because the mapping φ defining the slow variable is usually nonlinear.

In particular, it is easy to see that averaging the original equation, i.e. using

$$\dot{\bar{Z}}_t = G(\varphi(\bar{Z}_t))$$

where

$$G(z) = \int_{\mathbb{R}^n} g(z) d\mu_x(z),$$

will, in general not be correct, in the sense that

$$\varphi(\bar{Z}_t) \neq X_t$$

unless $\nabla\varphi(z)$ is a function of x alone, i.e.

$$\nabla\varphi(z) = J(\varphi(z))$$

for some $J : \mathbb{R}^m \rightarrow \mathbb{R}^n \times \mathbb{R}^m$.

Only if (??) is satisfied do we have

$$F(z) = G(z)J(z) \quad \text{and} \quad \varphi(\bar{Z}_t) = X_t$$

When this is the case, we can build a seamless multiscale algorithm based only on our ability to decompose the velocity field in its fast $O(\varepsilon^{-1})$ component and its slow $O(1)$ component.

Seamless multiscale algorithm - bis

If

$$\nabla\varphi(z) = J(\varphi(z))$$

for some $J : \mathbb{R}^m \rightarrow \mathbb{R}^n \times \mathbb{R}^m$, then we can do the following to integrate

$$\dot{Z}_t = \frac{1}{\varepsilon} f(Z_t) + g(Z_t)$$

1. Use

$$\tilde{Z}_{m+1,n} = \tilde{Z}_{m,n} + \frac{\delta t}{\varepsilon} f(\tilde{Z}_{m,n}), \quad \tilde{Z}_{0,0} = Z_{t=0}, \quad \tilde{Z}_{0,n+1} = \tilde{Z}_{M+M_T,n}$$

and compute

$$\tilde{G}_n = \frac{1}{M} \sum_{m=M_T}^{M+M_T-1} g(\tilde{Z}_{n,m})$$

2. Use

$$Z_{n+1} = Z_n + \Delta t \tilde{G}_n, \quad Z_0 = Z_{t=0}$$

3. Repeat.

Here $\varphi(Z_n) \approx X(n\Delta t)$.

Notice that the algorithm is totally seamless, i.e. one does not need to know φ nor $\nabla\varphi$.

Lecture 3:

Application to Markov jump processes (aka Kinetic Monte-Carlo - KMC)

Evolution of an isothermal, spatially homogeneous mixture of chemically reacting molecules contained in a fixed volume V .

N_S species of molecules S_i , $i = 1, \dots, N_S$ involved in M_R reactions R_j , $j = 1, \dots, M_R$.

Each reaction R_j is characterized by a rate function $a_j(x)$ and a state change (or stoichiometric) vector ν_j :

$$R_j = (a_j, \nu_j), \quad R = \{R_1, \dots, R_{M_R}\}.$$

Let x_i be the number of molecules of species S_i . Given the state $x = (x_1, \dots, x_{N_S})$, the occurrences of the reactions on an infinitesimal time interval dt are independent of each other and the probability for reaction R_j to happen during this time interval is given by $a_j(x)dt$. The state of the system after reaction R_j is $x + \nu_j$.

Equivalently: Given that the state of the system is $X_t = x$ at time t ;

1. The probability that the next reaction happens after time $t + s$ is $e^{-a(x)s}$ where $a(x) = \sum_{j=1}^{M_R} a_j(x)$.
2. Given that a reaction happens at time $t + s$, the probability that it be reaction j is $a_j(x)/a(x)$.

Gillespie's Stochastic Simulation Algorithm (SSA)

Randomly choose when the next reaction occurs according to 1. above; then:

Randomly choose which one occurs according to 2. above.

Gillespie's Stochastic Simulation Algorithm (SSA)

Let

$$a(x) = \sum_{j=1}^{M_R} a_j(x).$$

Assume that the current time is t_n , and the system is at state X_n . We perform the following steps:

1. Generate independent random numbers r_1 and r_2 with uniform distribution on the unit interval $(0, 1]$. Let

$$\delta t_{n+1} = -\frac{\ln r_1}{a(X_n)},$$

and k_{n+1} be the natural number such that

$$\frac{1}{a(X_n)} \sum_{j=0}^{k_{n+1}-1} a_j(X_n) < r_2 \leq \frac{1}{a(X_n)} \sum_{j=0}^{k_{n+1}} a_j(X_n),$$

where $a_0 = 0$ by convention.

2. Update the time and the state of the system by

$$t_{n+1} = t_n + \delta t_{n+1}, \quad X_{n+1} = X_n + \nu_{k_{n+1}}.$$

3. Repeat

Suppose that are fast and slow reactions:

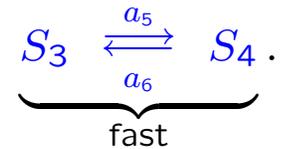
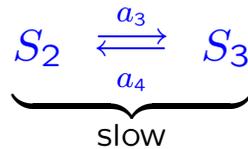
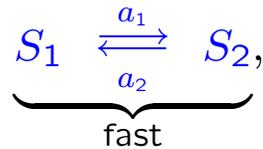
$$R_j^s = (a_j^s(x), \nu_j^s), \quad R_j^f = (\varepsilon^{-1} a_j^f(x), \nu_j^f).$$

where $\varepsilon \ll 1$ represents the ratio of time scales of the system.

Then: *The time-step between reactions is $O(\varepsilon)$ and with probability $1 - O(\varepsilon)$ a fast reaction happens.*

Difficult to simulate the evolution up to the $O(1)$ time-scale of the slow reactions!

Simple example:



with

$$a_1 = 10^5 x_1,$$

$$\nu_1 = (-1, +1, 0, 0),$$

$$a_2 = 10^5 x_2,$$

$$\nu_2 = (+1, -1, 0, 0),$$

$$a_3 = x_2,$$

$$\nu_3 = (0, -1, +1, 0),$$

$$a_4 = x_3,$$

$$\nu_4 = (0, +1, -1, 0),$$

$$a_5 = 10^5 x_3,$$

$$\nu_5 = (0, 0, -1, +1),$$

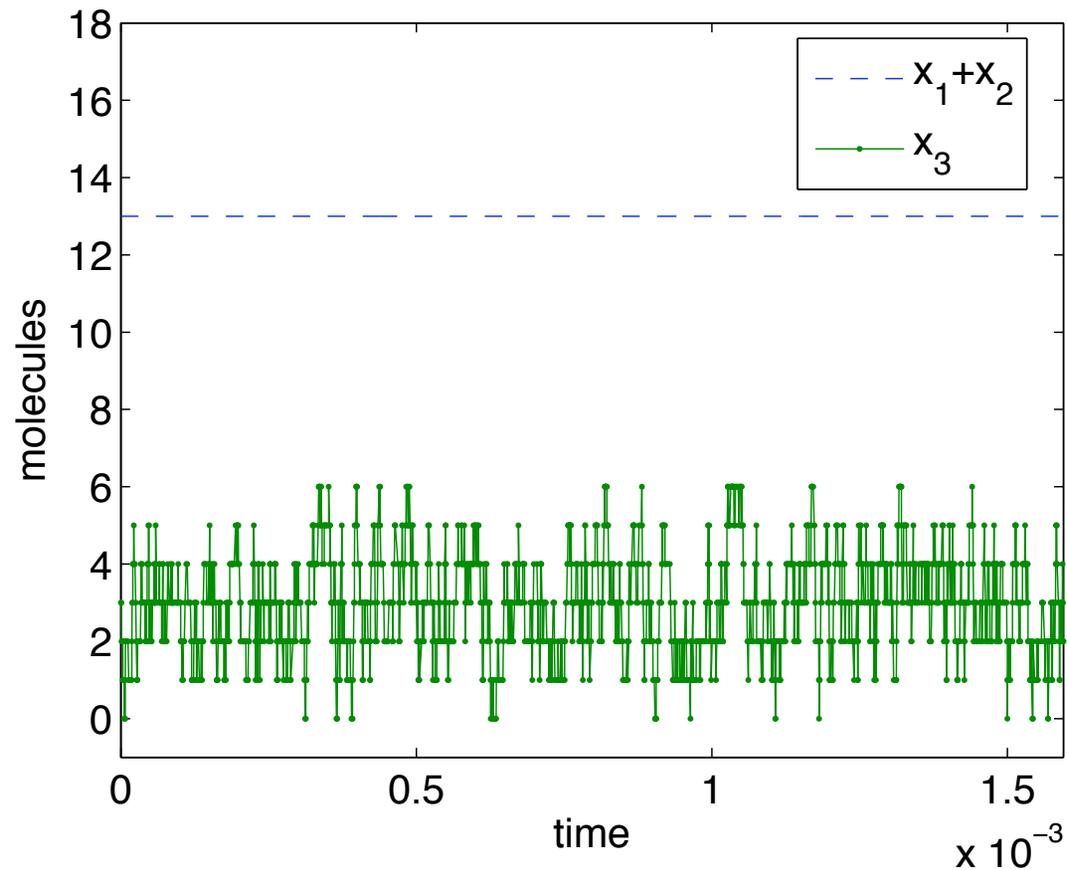
$$a_6 = 10^5 x_4,$$

$$\nu_6 = (0, 0, +1, -1).$$

i.e. the first and third reactions are faster than the second one.

Every species is involved in at least one fast reaction so there is no slow species.

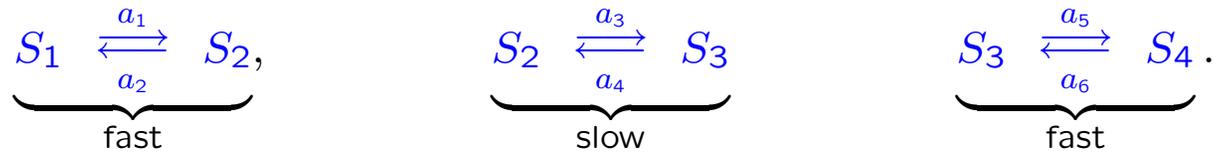
But the variables $y_1 = x_1 + x_2$ and $y_2 = x_3 + x_4$ are conserved during the fast reactions and only evolve during the slow reaction.



Evolution of one of the slow variable, $y_1 = x_1 + x_2$ (the other, $y_2 = x_3 + x_4$ behaves similarly), and one of the fast variables, x_3 , on an intermediate time scale.

Note that x_3 is also the “instantaneous” rate of reaction for the slow variable $y_2 = x_3 + x_4$.

Averaging thm:



The slow variables are

$$y_1 = x_1 + x_2, \quad y_2 = x_3 + x_4$$

The equilibrium distribution of the virtual fast process is

$$\mu_{y_1, y_2}(x_1, x_2, x_3, x_4) = \frac{y_1! y_2!}{x_1! x_2! x_3! x_4!} (1/2)^{y_1} (1/2)^{y_2} \delta_{x_1+x_2=y_1} \delta_{x_3+x_4=y_2}.$$

Effective dynamics:

$$\begin{aligned}
 \bar{a}_3^s = P x_2 &= \frac{x_1 + x_2}{2} = \frac{y_1}{2}, & \bar{v}_3^s &= (-1, +1), \\
 \bar{a}_4^s = P x_3 &= \frac{x_3 + x_4}{2} = \frac{y_2}{2}, & \bar{v}_4^s &= (+1, -1).
 \end{aligned}$$

Key observation: $v(x) = b \cdot x$ is a slow variable if

$$b \cdot \nu_j^f = 0$$

for all ν_j^f . The set of such vectors form a linear subspace in R^{N_s} .

Let b_1, b_2, \dots, b_J be a set of basis vectors of this subspace, and define

$$y_j = b_j \cdot x \quad \text{for } j = 1, \dots, J,$$

then y_1, y_2, \dots, y_J defines a complete set of slow variables, i.e. all slow observables can be expressed as functions of y_1, y_2, \dots, y_J .

In other words, in the present case, the slow variables are linear functions of the original variables! This allows for a seamless formulation of the limiting theorem (see next).

Notice that the slow variables are *not* slow species (which do not exist in general)!

Seamless limit theorem:

Consider a Markov jump process with generator $L = L_0 + \varepsilon^{-1}L_1$ where

$$\begin{cases} (L_0 f)(x) = \sum_{j=1}^{N_s} a_j^s(x)(f(x + \nu_j^s) - f(x)) \\ (L_1 f)(x) = \sum_{j=1}^{N_f} a_j^f(x)(f(x + \nu_j^f) - f(x)) \end{cases}$$

Assume that the fast process is ergodic with respect to the foliated measure $\mu_y(x')$ indexed by $y = b \cdot x$, i.e.

$$\frac{1}{T} \int_0^T F(X_t^f) dt \rightarrow \sum_{x'} f(x') \mu_{b \cdot x}(x')$$

where X_t^f is a sample path with $X_{t=0}^f = x$ of the process with generator L_1 .

Then, for any $T > 0$, there exists a constant $C > 0$ such that

$$\sup_{0 \leq t \leq T} |\mathbb{E} f(X_t) - \mathbb{E} \sum_{x'} \mu_{b \cdot \bar{X}_t}(x') f(x')| \leq C\varepsilon$$

where \bar{X}_t is a sample path of the process with generator

$$(\bar{L}f)(x) = \sum_{j=1}^{N_s} \bar{a}_j(b \cdot x)(f(x + \nu_j^s) - f(x)) \quad \bar{a}_j(y) = \sum_{x'} \mu_y(x') a_j^s(x')$$

Nested Stochastic Simulation Algorithm (nSSA)

1. **Inner SSA:** Run N independent replicas of SSA with the fast reactions $R^f = \{(\epsilon^{-1}a^f, \nu^f)\}$ only, for a time interval of $T_0 + T_f$.

During this calculation, compute the modified slow rates for $j = 1, \dots, M_s$

$$\tilde{a}_j^s = \frac{1}{N} \sum_{k=1}^N \frac{1}{T_f} \int_{T_0}^{T_f+T_0} a_j^s(X_\tau^k) d\tau,$$

where X_τ^k is the result of the k -th replica of this auxiliary virtual fast process at virtual time τ whose initial value is $X_{t=0}^k = X_n$, and T_0 is a parameter we choose in order to minimize the effect of the transients to the equilibrium in the virtual fast process.

2. **Outer SSA:** Run one step of SSA for the modified slow reactions $\tilde{R}^s = (\tilde{a}^s, \nu^s)$ to generate (t_{n+1}, X_{n+1}) from (t_n, X_n) .

Then repeat.

Totally seamless!

Error estimate:

For any $T > 0$, there exist constants C and α independent of (N, T_0, T_f) such that,

$$\sup_{0 \leq t \leq T} \mathbb{E} |v(x, t) - u(x, t)| \leq C \left(\epsilon + \frac{e^{-\alpha T_0/\epsilon}}{1 + T_f/\epsilon} + \frac{1}{\sqrt{N(1 + T_f/\epsilon)}} \right).$$

Here:

$v(x, t) = \mathbb{E}f(X_t^\epsilon)$ where X_t^ϵ is an exact path, and

$v(x, t) = \mathbb{E}f(X_t)$ where X_t is a pathway from the nested SSA.

Efficiency:

Given an error tolerance λ :

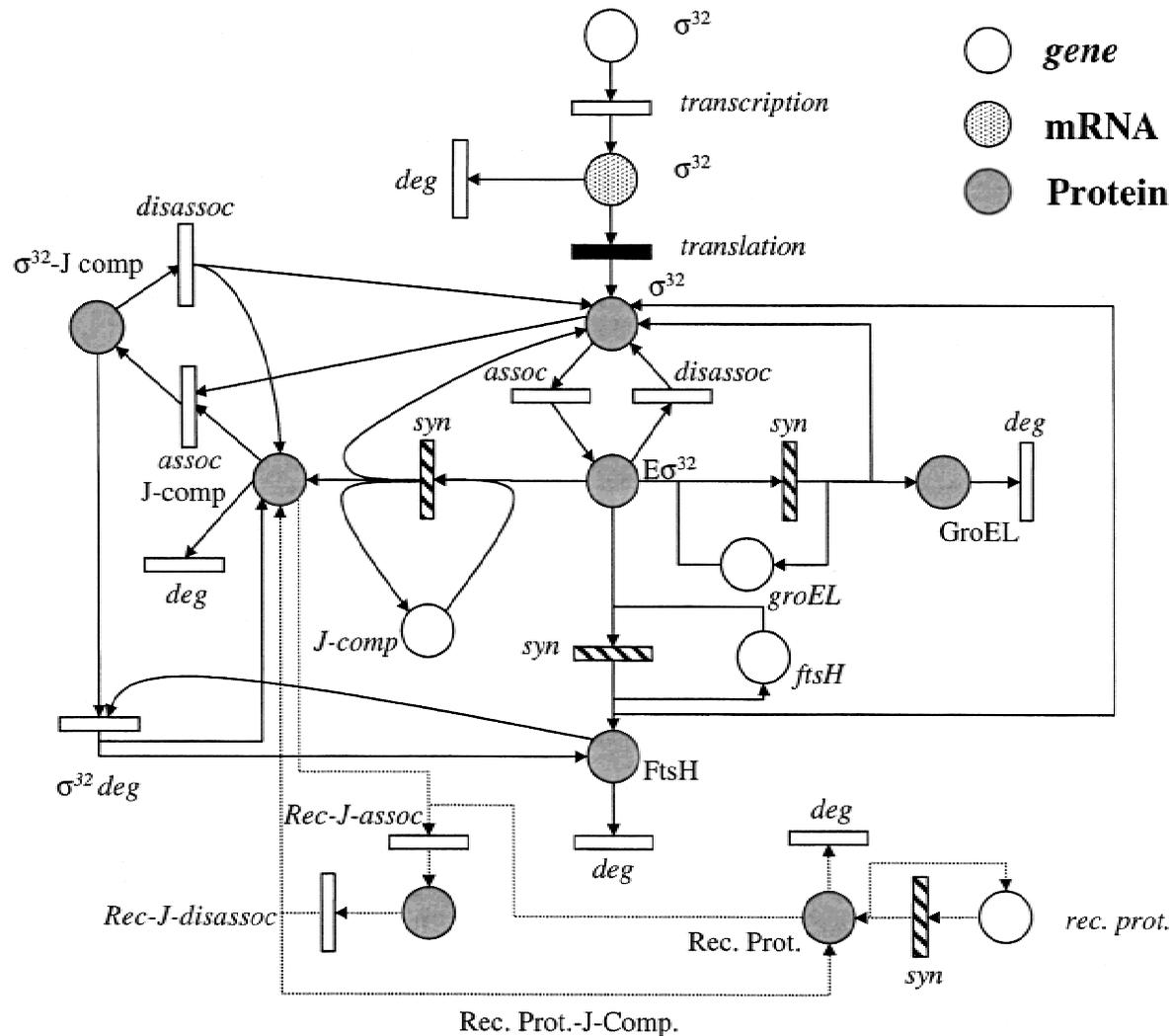
$$\text{cost} = O(N(1 + T_0/\epsilon + T_f/\epsilon)) = O\left(\frac{1}{\lambda^2}\right) \quad (\text{nested SSA})$$

$$\text{cost} = O\left(\frac{1}{\epsilon}\right) \quad (\text{direct SSA})$$

Example: heat shock response of Escherichia Coli

Reaction	Rates magnitude
DNA. $\sigma^{32} \rightarrow$ mRNA. σ^{32}	1.4×10^{-3}
mRNA. $\sigma^{32} \rightarrow \sigma^{32} +$ mRNA. σ^{32}	1.19
mRNA. $\sigma_{32} \rightarrow$ degradation	2.38×10^{-5}
$\sigma_{32} \rightarrow$ RNAP σ^{32}	10.5
RNAP $\sigma^{32} \rightarrow \sigma^{32}$	9.88
$\sigma^{32} +$ DnaJ $\rightarrow \sigma^{32}$.DnaJ (**)	25.2
DnaJ \rightarrow degradation (**)	2.97×10^{-6}
σ^{32} .DnaJ $\rightarrow \sigma^{32} +$ DnaJ	1.30
DNA.DnaJ + RNAP $\sigma^{32} \rightarrow$ DnaJ + DNA.DnaJ + σ^{32}	3.71
DNA.FtsH + RNAP. $\sigma^{32} \rightarrow$ FtsH + DNA.FtsH + σ^{32}	0
FtsH \rightarrow degradation	1.48×10^{-8}
GroEL \rightarrow degradation	7.76×10^{-5}
σ^{32} .DnaJ + FtsH \rightarrow DnaJ + FtsH	8.4
DNA.GroEL + RNAP $\sigma^{32} \rightarrow$ GroEL + DNA.GroEL + σ^{32}	4.78
Protein \rightarrow UnfoldedProtein (*)	10^6
DnaJ+ UnfoldedProtein \rightarrow DnaJ.UnfoldedProtein (*)	10^7
DnaJ.UnfoldedProtein \rightarrow DnaJ+ UnfoldedProtein (*)	10^6

Reaction list for the heat shock model of E. Coli proposed in: R. Srivastava, M. S. Peterson and W. E. Bently, *Biotech. Bioeng.* **75**, 120–129 (2001). The rate magnitude is the value of $a_i(x)$ evaluated at initial time or equilibrium. The last three reactions marked with a (*) in the table are fast: they are used in the Inner SSA. All the other reactions are used in the Outer SSA, and the rates of the reactions marked with a (**) are averaged.



Stochastic petri net diagram for the model of heat shock of E. Coli [from R. Srivastava, M. S. Peterson and W. E. Bently, *Biotech. Bioeng.* **75**, 120–129 (2001)]

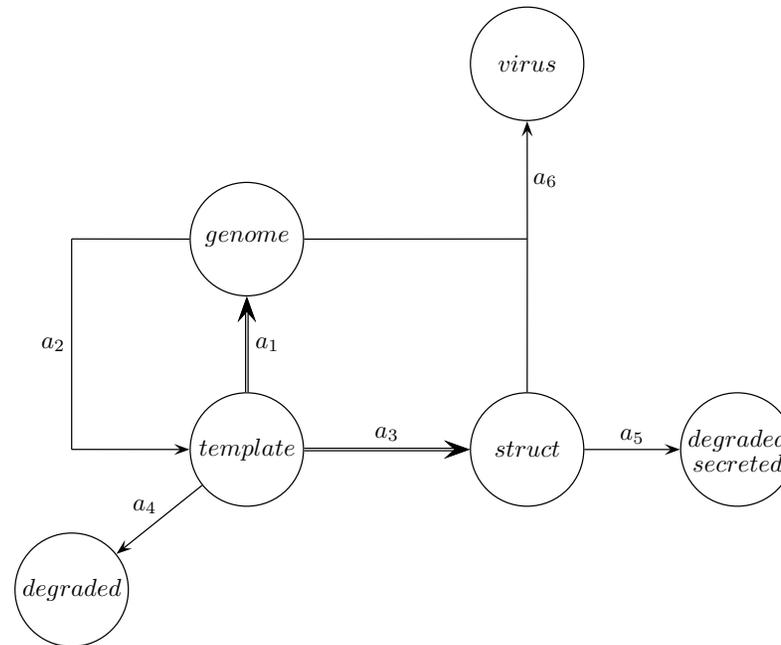
$(N, T_f/10^{-6})$	(1, 1)	(1, 4)	(1, 16)	(1, 64)	(1, 256)	(1, 1024)
CPU	0.62	1.32	2.98	9.56	35.81	142.08
$\overline{\sigma^{32}}$	4.60	8.66	13.60	14.52	14.98	15.00
$\text{var}(\sigma^{32})$	4.41	8.11	12.22	13.13	13.73	14.66

Efficiency of nested SSA when $N = 1$. Since we used $N_0 = 1000$ realizations of the Outer SSA to compute $\overline{\sigma^{32}}$ and $\text{var}(\sigma^{32})$, the statistical errors on these quantities is about 0.2. For comparison, the actual values of these quantities are

$$\overline{\sigma^{32}} = 14.8 \pm 0.2, \quad \text{var}(\sigma^{32}) = 14.2 \pm 0.2.$$

and the direct SSA took 19719.2 seconds of CPU time to compute them.

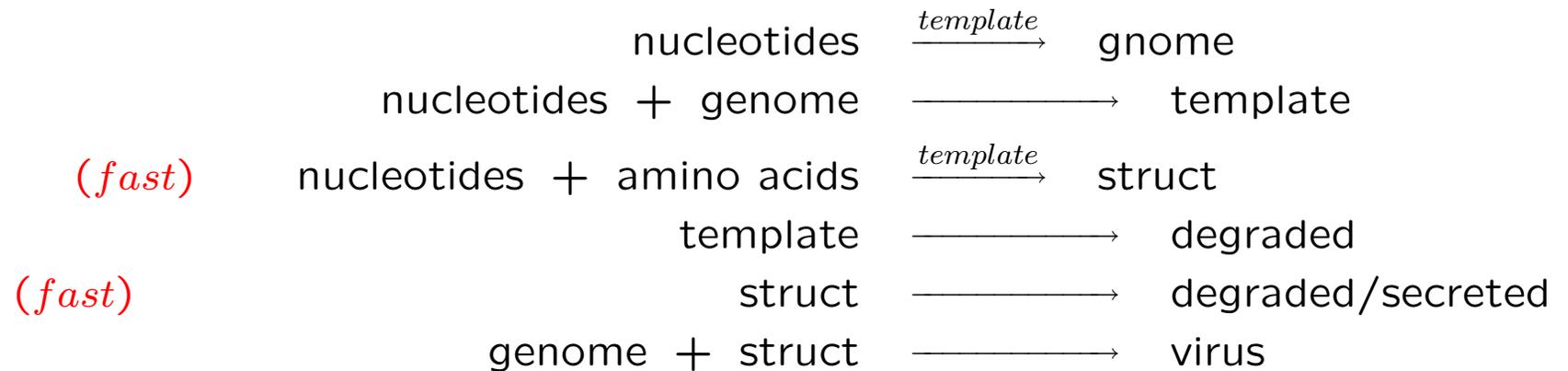
Example: virus infection model



General virus infection model proposed in: R. Srivastava, L. You, J. Summers and J. Yin, *J. Theor. Biol.* **218**, 309-321 (2002).

The nucleotides and amino acids are assumed to be available at constant concentrations. The reacting species that need to be modeled are genome, struct, template and virus ($N_s = 4$). The quantity genome represents the vehicle of the viral genetic information which can take the form of DNA, positive-strand RNA, negative-strand RNA, or some other variants. The structural proteins making up the virus are denoted by struct. Template refers to the form of the nucleic acid that is transcribed and involved in catalytically synthesizing each viral component.

Reaction list in the virus model:



template ——— nucleic acids catalyzing the synthesis of the virus components

genome ——— DNA (RNA) transporting the viral genetic information

struct ——— structural protein

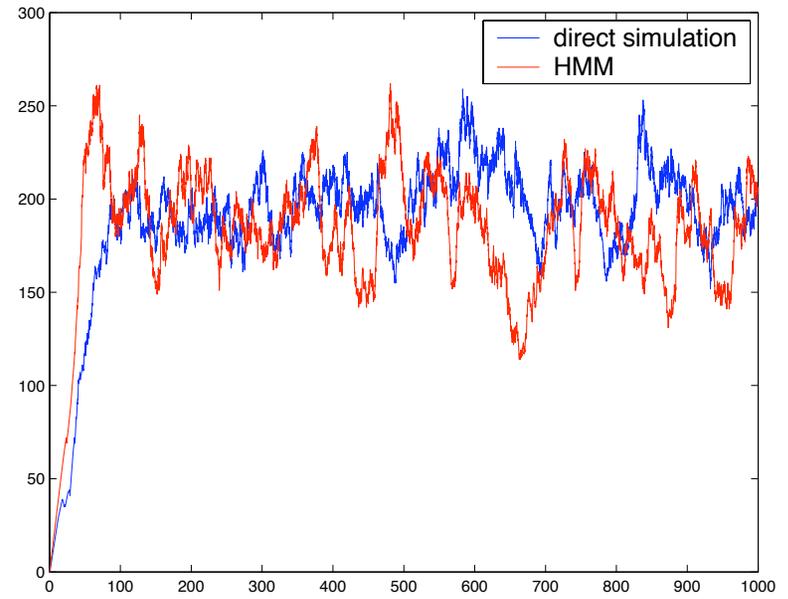
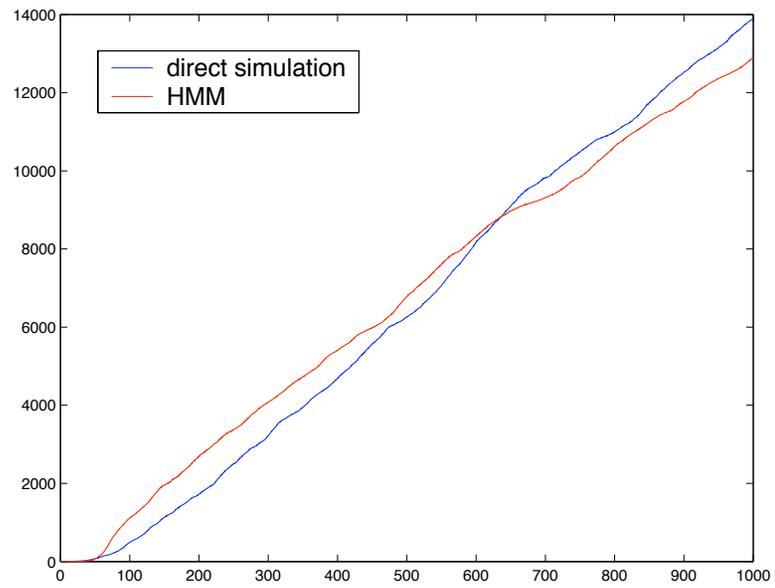
Reactions		Rates
<i>nucleotides</i>	$\xrightarrow{\text{template}}$ <i>genome</i>	$a_1 = 1. \times \text{template}$
<i>nucleotides + genome</i>	\longrightarrow <i>template</i>	$a_2 = .025 \times \text{genome}$
<i>template</i>	\longrightarrow <i>degraded</i>	$a_4 = .25 \times \text{template}$
<i>genome + struct</i>	\longrightarrow <i>virus</i>	$a_6 = 3.75d - 3 \times \text{genome}^2 \times \text{struct}$

T_f/ϵ	1	4	16	64
<i>CPU</i>	154.8	461.3	2068.2	9190.9
$\overline{\text{template}}$	4.027	3.947	3.796	3.757
$\text{var}(\text{template})$	5.401	5.254	5.007	4.882

Efficiency of the nested SSA for the virus infection model: the direct SSA cost was 34806.39 seconds of CPU time

The exact values are :

$$\overline{\text{template}} = 3.7170 \pm 0.005, \quad \text{var}(\text{template}) = 4.9777 \pm 0.005.$$



Left: Growth of the virus; Right: evolution of template

Asymptotic techniques for singularly perturbed Markov processes:

R. Z. Khasminsky, On Stochastic Processes Defined by Differential Equations with a Small Parameter, *Theory Prob. Applications*, **11**:211–228, 1966.

R. Z. Khasminsky, A Limit Theorem for the Solutions of Differential Equations with Random Right-Hand Sides, *Theory Prob. Applications*, **11**:390–406, 1966.

T. G. Kurtz, Semigroups of conditioned shifts and approximations of Markov processes, *Annals of Probability*, **3**:618–642, 1975.

G. Papanicolaou, Some probabilistic problems and methods in singular perturbations, *Rocky Mountain J. Math*, **6**:653–673, 1976.

M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, 2nd edition, Springer-Verlag, 1998.

G. Pavliotis and A. Stuart. *Multiscale Methods: Averaging and Homogenization*. (to appear)

Explicit integrators for stiff ODEs:

C.W. Gear and I.G. Kevrekidis, "Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum," *SIAM J. Sci. Comp.* **24**(4):109-110 (2003).

K. Eriksson, C. Johnson, and A. Logg, "Explicit Time-stepping for stiff ODEs," *SIAM J. Sci. Comp.* **25**(4):1142-1157 (2003).

V. I. Lebedev and S. I. Finogenov, "Explicit methods of second order for the solution of stiff systems of ordinary differential equations", *Zh. Vychisl. Mat. Mat Fiziki*, **16**:895–910 (1976).

A. Abdulle, "Fourth order Chebychev methods with recurrence relations", *SIAM J. Sci. Comput.* **23**:2041–2054 (2002).

HMM-like multiscale integrators for stochastic systems

E. Vanden-Eijnden, "Numerical techniques for multiscale dynamical systems with stochastic effects", *Comm. Math. Sci.*, **1**: 385–391 (2003).

W. E, D. Liu and E. Vanden-Eijnden, "Analysis of Multiscale Methods for Stochastic Differential Equations," *Comm. Pure App. Math.* **58**: 1544–1585 (2005).

E. Vanden-Eijnden, On HMM-like integrators and projective integration methods for systems with multiple time scales, *Comm. Math. Sci.* (2007).

HMM and Equation Free:

W. E, B. Engquist, The heterogeneous multi-scale methods, *Comm. Math. Sci.* **1** (2003), 87–133.

Kevrekidis, I. G.; Gear, C. W.; Hyman, J. M.; Panagiotis, G. K.; Runborg, O.; Theodoropoulos, C. Equation-Free, Coarse-Grained Multiscale Computation: Enabling Microscopic Simulators to Perform System-Level Analysis. *Comm. Math. Sci.* **1**: 87–133 (2003).

Effective dynamics in L96:

E. N. Lorenz, Predictability – A problem partly solved, pp. 1–18 in: ECMWF Seminar Proceedings on Predictability, Reading, United Kingdom, ECMWF, 1995.

C. Rödenbeck, C. Beck, H. Kantz, Dynamical systems with time scale-separation: averaging, stochastic modeling, and central limit theorems, pp. 187–209 in: Stochastic Climate Models, P. Imkeller, J.-S. von Storch eds., Progress in Probability **49**, Birkhäuser Verlag, Basel, 2001.

I. Fatkullin and E. Vanden-Eijnden, A computational strategy for multiscale systems with applications to Lorenz 96 model, *J. Comp. Phys.* **200**:605–638, 2004.

(Coupled) TBH:

A. J. Majda and I. Timofeyev, Remarkable statistical behavior for truncated Burgers-Hopf dynamics, *Proc. Nat. Acad. Sci. USA*, **97**:12413–12417, 2000.

A. J. Majda and I. Timofeyev, Statistical mechanics for truncations of the Burgers-Hopf equation: a model for intrinsic stochastic behavior with scaling, *Milan Journal of Mathematics*, **70**(1):39–96, 2002.

A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden, Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.*, **60**(14):1705–1722, 2003.

A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden, Stochastic models for selected slow variables in large deterministic systems, *Nonlinearity*, **19**:769–794, 2006.

Stochastic Simulation Algorithm:

D. T. Gillespie, *J. Comp. Phys.* **22**, 403 (1976)

A. B. Bortz, M. H. Kalos and J. L. Lebowitz, *J. Comp. Phys.* **17**, 10 (1975)

Seamless scheme for Kinetic Monte-Carlo:

W. E, D. Liu, E. Vanden-Eijnden, “Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates,” *J. Chem. Phys.* **123**: 194107 (2005)

W. E, D. Liu, E. Vanden-Eijnden, “Nested stochastic simulation algorithm for chemical kinetic systems with multiple time-scales,” *J. Comp. Phys.* **221**, 158–180 (2007).

Related works:

Y. Cao, D. Gillespie, L. Petzold, “Multiscale Stochastic Simulation Algorithm with Stochastic Partial Equilibrium Assumption for Chemically Reacting Systems,” *J. Comp. Phys.* **206**(2):395–411 (2005).

A. Samant and D. G. Vlachos, “Overcoming stiffness in stochastic simulation stemming from partial equilibrium: A multiscale Monte Carlo algorithm.,” *J. Chem. Phys.* **123**, 144114 (2005)