Importance sampling for a multiscale diffusion

Paul Dupuis

Division of Applied Mathematics
Brown University

(Kostas Spiliopoulos, Hui Wang)

SAMSI

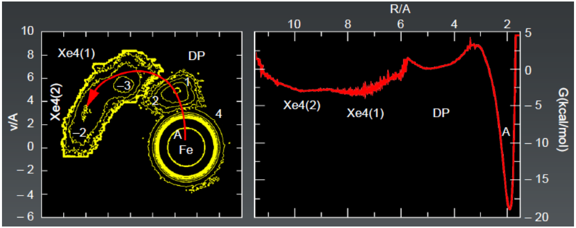
14 February 2012

Problem of interest

Where we want to end up.

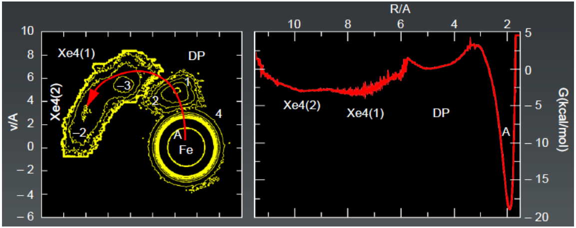
Problem of interest

Where we want to end up. Good Monte Carlo estimators to calculate transition probabilities for complex dynamics, e.g., diffusion with drift given by the potential



Problem of interest

Where we want to end up. Good Monte Carlo estimators to calculate transition probabilities for complex dynamics, e.g., diffusion with drift given by the potential



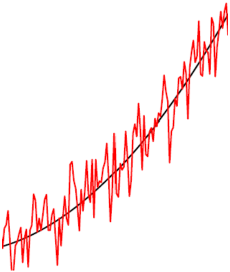
This talk concerns estimators based on change-of-measure, i.e., *importance sampling*.

Problem of interest

What is a good (local) model?

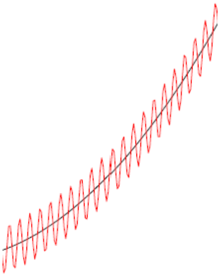
Problem of interest

What is a good (local) model? *Local model 1*: Fast oscillating ergodic random field, superimposed on slowly varying surface



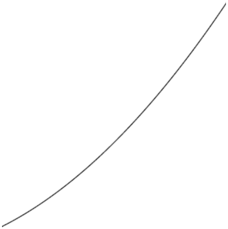
Problem of interest

Local model 2: Fast oscillating periodic function, superimposed on slowly varying surface



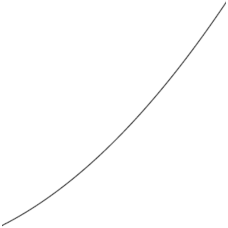
Problem of interest

Local model 3: Slowly varying surface



Problem of interest

Local model 3: Slowly varying surface



Methodology–there is a PDE/subsolutions approach that can be applied to this model. Can it be adapted to the previous models?

Problem of interest

Answers:

- Yes, numerically and theoretically, to fast oscillating periodic function.

Problem of interest

Answers:

- Yes, numerically and theoretically, to fast oscillating periodic function.
- Yes numerically, but not yet theoretically, to fast oscillating ergodic random field.

Problem of interest

Answers:

- Yes, numerically and theoretically, to fast oscillating periodic function.
- Yes numerically, but not yet theoretically, to fast oscillating ergodic random field.
- Nothing attempted yet with original problem

Outline

1. Process model and main assumptions
2. Large deviation theory for periodic model
3. Importance sampling and subsolutions
4. Formal extension to ergodic coefficients

Process model and assumptions

We consider a diffusion with multiple spatial scales:

$$dX^{\varepsilon,\delta} = \left[\frac{\varepsilon}{\delta} b\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right) + c\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right)\right] dt + \sqrt{\varepsilon}\sigma\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right) dW.$$

Process model and assumptions

We consider a diffusion with multiple spatial scales:

$$dX^{\varepsilon,\delta} = \left[\frac{\varepsilon}{\delta} b\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right) + c\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right)\right] dt + \sqrt{\varepsilon}\sigma\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right) dW.$$

Three different limiting behaviors are possible, depending on

$$\frac{\varepsilon}{\delta} \to \begin{cases} 0 \\ c \in (0,\infty) \\ \infty \end{cases} \quad \text{as } \varepsilon \to 0 \text{ and } \delta \to 0.$$

Process model and assumptions

We consider a diffusion with multiple spatial scales:

$$dX^{\varepsilon,\delta} = \left[\frac{\varepsilon}{\delta} b\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right) + c\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right)\right] dt + \sqrt{\varepsilon}\sigma\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right) dW.$$

Three different limiting behaviors are possible, depending on

$$\frac{\varepsilon}{\delta} \to \begin{cases} 0 \\ c \in (0,\infty) \\ \infty \end{cases} \quad \text{as } \varepsilon \to 0 \text{ and } \delta \to 0.$$

This talk focuses on $\varepsilon/\delta \to \infty$: averaging or homogenization first, then small noise (see reference for other cases).

Process model and assumptions

An important special case:

$$dX^{\varepsilon,\delta} = \left[-\frac{\varepsilon}{\delta}\nabla Q\left(\frac{X^{\varepsilon,\delta}}{\delta}\right) - \nabla S\left(X^{\varepsilon,\delta}\right)\right] dt + \sqrt{2\varepsilon}\, dW,$$

where S is a slowly varying potential and Q is periodic or ergodic random field.

Process model and assumptions

An important special case:

$$dX^{\varepsilon,\delta} = \left[-\frac{\varepsilon}{\delta} \nabla Q \left(\frac{X^{\varepsilon,\delta}}{\delta} \right) - \nabla S \left(X^{\varepsilon,\delta} \right) \right] dt + \sqrt{2\varepsilon} dW,$$

where S is a slowly varying potential and Q is periodic or ergodic random field. What is a simpler approximating model?

Process model and assumptions

An important special case:

$$dX^{\varepsilon,\delta} = \left[-\frac{\varepsilon}{\delta} \nabla Q \left(\frac{X^{\varepsilon,\delta}}{\delta} \right) - \nabla S \left(X^{\varepsilon,\delta} \right) \right] dt + \sqrt{2\varepsilon} dW,$$

where S is a slowly varying potential and Q is periodic or ergodic random field. What is a simpler approximating model? Consider 1-D. Since $\varepsilon/\delta \to \infty$ and $\delta \to 0$, by rescaling time the occupation measure

$$\theta^{\varepsilon,\delta}(A) = \frac{1}{\Delta} \int_0^{\Delta} \delta_{X^{\varepsilon,\delta}(s)/\delta}(A) ds$$

is approximately the invariant distribution associated with generator

$$\mathcal{L}f(y) = -\nabla Q(y) \cdot \nabla f(y) + \nabla^2 f(y), \quad y \in [0, \lambda]$$

and periodic b.c.'s, which is

$$\mu(dy) = \frac{1}{L} e^{-Q(y)} dy, \quad L = \int_0^{\lambda} e^{-Q(y)} dy.$$

Process model and assumptions

Let $\xi(y)$ solve

$$\mathcal{L}\xi(y) = -\nabla Q(y), \quad \int_0^\lambda \xi(y)\mu(dy) = 0,$$

which turns out here to be

$$\xi(y) = \frac{1}{\hat{L}} \int_0^y e^{Q(r)} dr \quad \hat{L} = \int_0^\lambda e^{Q(y)} dy.$$

Process model and assumptions

Let $\xi(y)$ solve

$$\mathcal{L}\xi(y) = -\nabla Q(y), \quad \int_0^{\lambda} \xi(y)\mu(dy) = 0,$$

which turns out here to be

$$\xi(y) = \frac{1}{\hat{L}} \int_0^{y} e^{Q(r)} dr \quad \hat{L} = \int_0^{\lambda} e^{Q(y)} dy.$$

Recall

$$dX^{\varepsilon,\delta} = \left[-\frac{\varepsilon}{\delta} \nabla Q \left(\frac{X^{\varepsilon,\delta}}{\delta} \right) - \nabla S \left(X^{\varepsilon,\delta} \right) \right] dt + \sqrt{2\varepsilon} dW.$$

Process model and assumptions

Let $\xi(y)$ solve

$$\mathcal{L}\xi(y) = -\nabla Q(y), \quad \int_0^\lambda \xi(y)\mu(dy) = 0,$$

which turns out here to be

$$\xi(y) = \frac{1}{\hat{L}} \int_0^y e^{Q(r)} dr \quad \hat{L} = \int_0^\lambda e^{Q(y)} dy.$$

Recall

$$dX^{\varepsilon,\delta} = \left[-\frac{\varepsilon}{\delta} \nabla Q \left(\frac{X^{\varepsilon,\delta}}{\delta} \right) - \nabla S \left(X^{\varepsilon,\delta} \right) \right] dt + \sqrt{2\varepsilon} dW.$$

Then by Itô applied to $G(y) = y + \delta\xi(y/\delta)$, we get

$$\begin{aligned} X^{\varepsilon,\delta}(t) - x &= -\int_0^t \left[1 + \xi \left(\frac{X^{\varepsilon,\delta}}{\delta} \right) \right] \nabla S \left(X^{\varepsilon,\delta} \right) dt \\ &\quad + \sqrt{2\varepsilon} \int_0^t \left[1 + \xi \left(\frac{X^{\varepsilon,\delta}}{\delta} \right) \right] dW + \text{small}. \end{aligned}$$

Process model and assumptions

Since

$$\begin{aligned} X^{\varepsilon,\delta}(t) - x &= -\int_0^t \left[1 + \xi\left(\frac{X^{\varepsilon,\delta}}{\delta}\right)\right] \nabla S\left(X^{\varepsilon,\delta}\right) dt \\ &\quad + \sqrt{2\varepsilon}\int_0^t \left[1 + \xi\left(\frac{X^{\varepsilon,\delta}}{\delta}\right)\right] dW + \text{small}, \end{aligned}$$

suggests the approximating process

$$d\bar{X}^{\varepsilon} = r\left(\bar{X}^{\varepsilon}\right) dt + \sqrt{\varepsilon} q\left(\bar{X}^{\varepsilon}\right)^{1/2} dW,$$

with

$$r(x) = -\nabla S(x) \int_0^{\lambda} [1 + \xi(y)]\, \mu(dy), \quad q(x) = \int_0^{\lambda} 2\,[1 + \xi(y)]^2\, \mu(dy).$$

Process model and assumptions

More precisely

$$d\bar{X}^{\varepsilon} = -\frac{\lambda^2}{L\hat{L}}\nabla S\left(\bar{X}^{\varepsilon}\right)dt + \sqrt{\varepsilon}\sqrt{\frac{\lambda^2}{L\hat{L}}2}dW, \quad \hat{L} = \int_0^{\lambda} e^{Q(y)}dy.$$

Process model and assumptions

More precisely

$$d\bar{X}^{\varepsilon} = -\frac{\lambda^2}{L\hat{L}}\nabla S\left(\bar{X}^{\varepsilon}\right)dt + \sqrt{\varepsilon}\sqrt{\frac{\lambda^2}{L\hat{L}}2dW}, \quad \hat{L} = \int_0^{\lambda} e^{Q(y)}dy.$$

By Hölder

$$\lambda^2 = \left(\int_0^{\lambda} dy\right)^2 < L\hat{L},$$

i.e., slower drift and less diffusive.

Process model and assumptions

1. The functions $b(x, y), c(x, y)$ and $\sigma(x, y)$ are bounded, possess appropriate smoothness, and are periodic with period λ in y in each direction,

Process model and assumptions

1. The functions $b(x,y), c(x,y)$ and $\sigma(x,y)$ are bounded, possess appropriate smoothness, and are periodic with period λ in y in each direction,
2. $\sigma(x,y)\sigma^T(x,y)$ is uniformly positive definite,

Process model and assumptions

1. The functions $b(x, y), c(x, y)$ and $\sigma(x, y)$ are bounded, possess appropriate smoothness, and are periodic with period λ in y in each direction,
2. $\sigma(x, y)\sigma^T(x, y)$ is uniformly positive definite,
3. Let $\mathbb{T}^d$ be the d-dimensional torus and for smooth $f : \mathbb{T}^d \to \mathbb{R}$

$$\mathcal{L}_x f(y) = \langle b(x, y), Df(y)\rangle + \frac{1}{2}\text{tr}\left[\sigma(x, y)\sigma^T(x, y)D^2 f(y)\right]$$

with periodic boundary conditions.

Process model and assumptions

1. The functions $b(x, y), c(x, y)$ and $\sigma(x, y)$ are bounded, possess appropriate smoothness, and are periodic with period λ in y in each direction,
2. $\sigma(x, y)\sigma^T(x, y)$ is uniformly positive definite,
3. Let $\mathbb{T}^d$ be the d-dimensional torus and for smooth $f : \mathbb{T}^d \to \mathbb{R}$

$$\mathcal{L}_x f(y) = \langle b(x, y), Df(y) \rangle + \frac{1}{2}\mathrm{tr}\left[\sigma(x, y)\sigma^T(x, y) D^2 f(y)\right]$$

with periodic boundary conditions.

Process model and assumptions

1. The functions $b(x,y), c(x,y)$ and $\sigma(x,y)$ are bounded, possess appropriate smoothness, and are periodic with period λ in y in each direction,
2. $\sigma(x,y)\sigma^T(x,y)$ is uniformly positive definite,
3. Let $\mathbb{T}^d$ be the d-dimensional torus and for smooth $f:\mathbb{T}^d\to\mathbb{R}$

$$\mathcal{L}_x f(y) = \langle b(x,y), Df(y)\rangle + \frac{1}{2}\mathrm{tr}\left[\sigma(x,y)\sigma^T(x,y)D^2 f(y)\right]$$

with periodic boundary conditions. Let $\mu(dy|x)$ be the associated invariant probability. Then the "centering"

$$\int_{\mathbb{T}^d} b(x,y)\mu(dy|x) = 0$$

holds, and a C^2 solution to the "cell problem" exists:

$$\mathcal{L}_x \xi_\ell(x,y) = b(x,y), \quad \int_{\mathbb{T}^d} \xi_\ell(x,y)\mu(dy|x) = 0, \quad \ell = 1,\dots,d.$$

Large deviation theory

The large deviation theory for

$$dX^{\varepsilon,\delta} = \left[\frac{\varepsilon}{\delta} b\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right) + c\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right)\right] dt + \sqrt{\varepsilon}\sigma\left(X^{\varepsilon,\delta}, \frac{X^{\varepsilon,\delta}}{\delta}\right) dW.$$

for all limits

$$\frac{\varepsilon}{\delta} \to \begin{cases} 0 \\ c \in (0,\infty) \\ \infty \end{cases} \quad \text{as } \varepsilon \to 0 \text{ and } \delta \to 0$$

is developed in

- Large deviations for multiscale diffusions via weak convergence methods (D. and K. Spiliopoulos), to appear in *SPA*.

Large deviation theory

While proofs technical, result for $\varepsilon/\delta \to \infty$ motivated by letting $\delta \to 0$ first. Then $X^{\varepsilon,\delta} \to \bar{X}^{\varepsilon}$ in distribution,

$$d\bar{X}^{\varepsilon} = r\left(\bar{X}^{\varepsilon}\right)dt + \sqrt{\varepsilon}q\left(\bar{X}^{\varepsilon}\right)^{1/2}dW,$$

Large deviation theory

While proofs technical, result for $\varepsilon/\delta \to \infty$ motivated by letting $\delta \to 0$ first. Then $X^{\varepsilon,\delta} \to \bar{X}^{\varepsilon}$ in distribution,

$$d\bar{X}^{\varepsilon} = r\left(\bar{X}^{\varepsilon}\right) dt + \sqrt{\varepsilon} q\left(\bar{X}^{\varepsilon}\right)^{1/2} dW,$$

with

$$\begin{aligned} r(x) &= \int_{\mathbb{T}^d} \left(I + \frac{\partial \xi}{\partial y}(x,y) \right) c(x,y)\mu(dy|x) \\ q(x) &= \int_{\mathbb{T}^d} \left(I + \frac{\partial \xi}{\partial y}(x,y) \right) \sigma(x,y)\sigma^T(x,y) \left(I + \frac{\partial \xi}{\partial y}(x,y) \right)^T \mu(dy|x) \end{aligned}$$

Large deviation theory

While proofs technical, result for $\varepsilon/\delta \to \infty$ motivated by letting $\delta \to 0$ first. Then $X^{\varepsilon,\delta} \to \bar{X}^{\varepsilon}$ in distribution,

$$d\bar{X}^{\varepsilon} = r\left(\bar{X}^{\varepsilon}\right) dt + \sqrt{\varepsilon} q\left(\bar{X}^{\varepsilon}\right)^{1/2} dW,$$

with

$$\begin{aligned} r(x) &= \int_{\mathbb{T}^d} \left(I + \frac{\partial \xi}{\partial y}(x,y)\right) c(x,y)\mu(dy|x) \\ q(x) &= \int_{\mathbb{T}^d} \left(I + \frac{\partial \xi}{\partial y}(x,y)\right) \sigma(x,y)\sigma^T(x,y) \left(I + \frac{\partial \xi}{\partial y}(x,y)\right)^T \mu(dy|x) \end{aligned}$$

Theorem. The large deviation behaviors of $X^{\varepsilon,\delta}$ and $\bar{X}^{\varepsilon}$ are the same.

Large deviation theory

A key point in the analysis. The proof is *not* based on super-exponential closeness.

A key point in the analysis. The proof is *not* based on super-exponential closeness. Representation for local rate:

$$\mathbb{P}\left[\frac{X^{\varepsilon,\delta}(\Delta) - X^{\varepsilon,\delta}(0)}{\Delta} \approx \beta \,\middle|\, X^{\varepsilon,\delta}(0) = x \right] \approx \exp -\frac{\Delta}{\varepsilon} L(x,\beta).$$

A key point in the analysis. The proof is *not* based on super-exponential closeness. Representation for local rate:

$$\mathbb{P}\left[\frac{X^{\varepsilon,\delta}(\Delta)-X^{\varepsilon,\delta}(0)}{\Delta}\approx\beta\middle|X^{\varepsilon,\delta}(0)=x\right]\approx\exp-\frac{\Delta}{\varepsilon}L(x,\beta).$$

Then (Boué-D)

$$L(x,\beta)\approx\inf\left\{E\frac{1}{\Delta}\int_0^\Delta\|v\|^2:\hat{X}^{\varepsilon,\delta}(\Delta)\approx x+\Delta\beta\right\},$$

where $\hat{X}^{\varepsilon,\delta}(0)=x$ and

$$\begin{aligned}d\hat{X}^{\varepsilon,\delta} &= \left[\frac{\varepsilon}{\delta}b\left(\hat{X}^{\varepsilon,\delta},\frac{\hat{X}^{\varepsilon,\delta}}{\delta}\right)+c\left(\hat{X}^{\varepsilon,\delta},\frac{\hat{X}^{\varepsilon,\delta}}{\delta}\right)\right]dt+\sigma\left(\hat{X}^{\varepsilon,\delta},\frac{\hat{X}^{\varepsilon,\delta}}{\delta}\right)vdt\\ &\quad+\sqrt{\varepsilon}\sigma\left(\hat{X}^{\varepsilon,\delta},\frac{\hat{X}^{\varepsilon,\delta}}{\delta}\right)dW.\end{aligned}$$

Large deviation theory

$$L(x,\beta) \approx \inf \left\{ E\frac{1}{\Delta}\int_0^\Delta \|v\|^2 : \hat{X}^{\varepsilon,\delta}(\Delta) \approx x + \Delta\beta \right\},$$

where $\hat{X}^{\varepsilon,\delta}(0) = x$ and

$$\begin{aligned} d\hat{X}^{\varepsilon,\delta} &= \left[\frac{\varepsilon}{\delta} b\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta}\right) + c\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta}\right)\right] dt + \sigma\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta}\right) v dt \\ &\quad + \sqrt{\varepsilon}\sigma\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta}\right) dW. \end{aligned}$$

Large deviation theory

$$L(x,\beta) \approx \inf \left\{ E \frac{1}{\Delta} \int_0^\Delta \|v\|^2 : \hat{X}^{\varepsilon,\delta}(\Delta) \approx x + \Delta\beta \right\},$$

where $\hat{X}^{\varepsilon,\delta}(0) = x$ and

$$\begin{aligned} d\hat{X}^{\varepsilon,\delta} &= \left[\frac{\varepsilon}{\delta} b\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right) + c\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right) \right] dt + \sigma\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right) v dt \\ &\quad + \sqrt{\varepsilon}\sigma\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right) dW. \end{aligned}$$

With standard small noise analysis can take v to be constant (depends on x, β).

Large deviation theory

$$L(x,\beta) \approx \inf \left\{ E \frac{1}{\Delta} \int_0^\Delta \|v\|^2 : \hat{X}^{\varepsilon,\delta}(\Delta) \approx x + \Delta\beta \right\},$$

where $\hat{X}^{\varepsilon,\delta}(0) = x$ and

$$\begin{aligned} d\hat{X}^{\varepsilon,\delta} &= \left[\frac{\varepsilon}{\delta} b\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right) + c\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right) \right] dt + \sigma\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right) v dt \\ &\quad + \sqrt{\varepsilon}\sigma\left(\hat{X}^{\varepsilon,\delta}, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right) dW. \end{aligned}$$

With standard small noise analysis can take v to be constant (depends on x,β). Here need form

$$v(t) = -\sigma\left(x, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right)^T \left(I + \frac{\partial \xi}{\partial y}\left(x, \frac{\hat{X}^{\varepsilon,\delta}}{\delta} \right) \right)^T u,$$

with u constant. *Need feedback information on fast variable.*

Importance sampling and subsolutions

How should homogenization (i.e., the limit $\delta \to 0$) be accounted for in the design of importance sampling?

How should homogenization (i.e., the limit $\delta \to 0$) be accounted for in the design of importance sampling?

1. obtain the limit problem sending first $\delta \to 0$ (the homogenized diffusion), with "effective" drift and diffusion coefficients;

Importance sampling and subsolutions

How should homogenization (i.e., the limit $\delta \to 0$) be accounted for in the design of importance sampling?

1. obtain the limit problem sending first $\delta \to 0$ (the homogenized diffusion), with "effective" drift and diffusion coefficients;
2. as noted the large deviation rate as $\varepsilon \to 0$ and $\delta \to 0$ with $\varepsilon/\delta \to \infty$ is the same as for the homogenized SDE;

Importance sampling and subsolutions

How should homogenization (i.e., the limit $\delta \to 0$) be accounted for in the design of importance sampling?

1. obtain the limit problem sending first $\delta \to 0$ (the homogenized diffusion), with "effective" drift and diffusion coefficients;
2. as noted the large deviation rate as $\varepsilon \to 0$ and $\delta \to 0$ with $\varepsilon/\delta \to \infty$ is the same as for the homogenized SDE;
3. not surprisingly, importance sampling based in the homogenized SDE is inadequate for any kind of asymptotic optimality;

Importance sampling and subsolutions

How should homogenization (i.e., the limit $\delta \to 0$) be accounted for in the design of importance sampling?

1. obtain the limit problem sending first $\delta \to 0$ (the homogenized diffusion), with "effective" drift and diffusion coefficients;
2. as noted the large deviation rate as $\varepsilon \to 0$ and $\delta \to 0$ with $\varepsilon/\delta \to \infty$ is the same as for the homogenized SDE;
3. not surprisingly, importance sampling based in the homogenized SDE is inadequate for any kind of asymptotic optimality;
4. bring in the solution to the eigenvector/eigenvalue problem used to identify homogenized problem and for large deviation analysis.

Importance sampling and subsolutions

Consider functionals such as

$$J^{\varepsilon,\delta} = \mathbb{E}\left[e^{-\frac{1}{\varepsilon}h(X^{\varepsilon,\delta}(T))}\right], \quad \bar{J}^{\varepsilon} = \mathbb{E}\left[e^{-\frac{1}{\varepsilon}h(\bar{X}^{\varepsilon}(T))}\right]$$

where $X^{\varepsilon,\delta}(0) = \bar{X}^{\varepsilon}(0) = x_0$ and h is continuous or indicator of a "nice" set. How to approximate via importance sampling?

Importance sampling and subsolutions

Consider functionals such as

$$J^{\varepsilon,\delta} = \mathbb{E}\left[e^{-\frac{1}{\varepsilon}h(X^{\varepsilon,\delta}(T))}\right], \quad \bar{J}^{\varepsilon} = \mathbb{E}\left[e^{-\frac{1}{\varepsilon}h(\bar{X}^{\varepsilon}(T))}\right]$$

where $X^{\varepsilon,\delta}(0) = \bar{X}^{\varepsilon}(0) = x_0$ and h is continuous or indicator of a "nice" set. How to approximate via importance sampling?

Subsolutions for the homogenized problem

Associated with $\bar{X}^{\varepsilon}$ is the Hamilton-Jacobi-Bellman equation

$$U_t(x,t) + \langle \nabla U(x,t), r(x)\rangle - \frac{1}{2}\left\|q^{1/2}(x)\nabla U(x,t)\right\|^2 = 0, \quad U(x,T) = h(x),$$

and $-\varepsilon \log \bar{J}^{\varepsilon} \to U(x_0, 0)$.

Importance sampling and subsolutions

Consider functionals such as

$$J^{\varepsilon,\delta} = \mathbb{E}\left[e^{-\frac{1}{\varepsilon}h(X^{\varepsilon,\delta}(T))}\right], \quad \bar{J}^{\varepsilon} = \mathbb{E}\left[e^{-\frac{1}{\varepsilon}h(\bar{X}^{\varepsilon}(T))}\right]$$

where $X^{\varepsilon,\delta}(0) = \bar{X}^{\varepsilon}(0) = x_0$ and h is continuous or indicator of a "nice" set. How to approximate via importance sampling?

Subsolutions for the homogenized problem

Associated with $\bar{X}^{\varepsilon}$ is the Hamilton-Jacobi-Bellman equation

$$U_t(x,t) + \langle \nabla U(x,t), r(x)\rangle - \frac{1}{2}\left\|q^{1/2}(x)\nabla U(x,t)\right\|^2 = 0, \quad U(x,T) = h(x),$$

and $-\varepsilon \log \bar{J}^{\varepsilon} \to U(x_0,0)$. Let V be a subsolution of the equation, i.e.,

$$V_t(x,t) + \langle \nabla V(x,t), r(x)\rangle - \frac{1}{2}\left\|q^{1/2}(x)\nabla V(x,t)\right\|^2 \geq 0, \quad V(x,T) \leq h(x),$$

Importance sampling and subsolutions

Let $u_2(x,t) = -q^{1/2}(x)\nabla V(x,t)$, simulate

$$d\bar{Y}^{\varepsilon} = r\left(\bar{Y}^{\varepsilon}\right)dt + q\left(\bar{Y}^{\varepsilon}\right)u_2(\bar{Y}^{\varepsilon},t) + \sqrt{\varepsilon}q\left(\bar{Y}^{\varepsilon}\right)dW,$$

and use samples

$$e^{-\frac{1}{\varepsilon}h(\bar{Y}^{\varepsilon}(T))}\frac{dP^x}{dP^y}(\bar{Y}^{\varepsilon}) \quad \text{in place of} \quad e^{-\frac{1}{\varepsilon}h(\bar{X}^{\varepsilon}(T))}.$$

Importance sampling and subsolutions

Let $u_2(x,t) = -q^{1/2}(x)\nabla V(x,t)$, simulate

$$d\bar{Y}^{\varepsilon} = r\left(\bar{Y}^{\varepsilon}\right)dt + q\left(\bar{Y}^{\varepsilon}\right)u_2(\bar{Y}^{\varepsilon},t) + \sqrt{\varepsilon}q\left(\bar{Y}^{\varepsilon}\right)dW,$$

and use samples

$$e^{-\frac{1}{\varepsilon}h(\bar{Y}^{\varepsilon}(T))}\frac{dP^x}{dP^y}(\bar{Y}^{\varepsilon}) \quad \text{in place of} \quad e^{-\frac{1}{\varepsilon}h(\bar{X}^{\varepsilon}(T))}.$$

One has the decay rate

$$\liminf_{\varepsilon\to 0}\varepsilon\log\mathbb{E}^y\left[e^{-\frac{1}{\varepsilon}h(\bar{Y}^{\varepsilon}(T))}\frac{dP^x}{dP^y}(\bar{Y}^{\varepsilon})\right]^2 \geq -U(x_0,0) - V(x_0,0).$$

Importance sampling and subsolutions

Let $u_2(x,t) = -q^{1/2}(x)\nabla V(x,t)$, simulate

$$d\bar{Y}^{\varepsilon} = r\left(\bar{Y}^{\varepsilon}\right)dt + q\left(\bar{Y}^{\varepsilon}\right)u_2(\bar{Y}^{\varepsilon},t) + \sqrt{\varepsilon}q\left(\bar{Y}^{\varepsilon}\right)dW,$$

and use samples

$$e^{-\frac{1}{\varepsilon}h(\bar{Y}^{\varepsilon}(T))}\frac{dP^x}{dP^y}(\bar{Y}^{\varepsilon}) \quad \text{in place of} \quad e^{-\frac{1}{\varepsilon}h(\bar{X}^{\varepsilon}(T))}.$$

One has the decay rate

$$\liminf_{\varepsilon\to 0}\varepsilon\log\mathbb{E}^y\left[e^{-\frac{1}{\varepsilon}h(\bar{Y}^{\varepsilon}(T))}\frac{dP^x}{dP^y}(\bar{Y}^{\varepsilon})\right]^2 \geq -U(x_0,0) - V(x_0,0).$$

If $V(x_0,0) = U(x_0,0)$ then one has *logarithmic optimality*. However, analogous change of measure does **not** work for the $X^{\varepsilon,\delta}$ process.

Importance sampling and subsolutions

For $X^{\varepsilon,\delta}$ we should use

$$u_1(x,y,t) = -\sigma^T(x,y)\left(I + \frac{\partial\xi}{\partial y}(x,y)\right)^T \nabla V(x,t),$$

and simulate

$$\begin{aligned} dY^{\varepsilon,\delta} \quad = \quad & \left[\frac{\varepsilon}{\delta} b\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right) + c\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right)\right] dt + \sigma\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right) \\ & \times \left[\sqrt{\varepsilon} dW - \sigma^T\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right)\left(I + \frac{\partial\xi}{\partial y}\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right)\right)^T DV(Y^{\varepsilon,\delta}, t)\right]. \end{aligned}$$

Importance sampling and subsolutions

For $X^{\varepsilon,\delta}$ we should use

$$u_1(x,y,t) = -\sigma^T(x,y)\left(I + \frac{\partial\xi}{\partial y}(x,y)\right)^T \nabla V(x,t),$$

and simulate

$$\begin{aligned} dY^{\varepsilon,\delta} &= \left[\frac{\varepsilon}{\delta} b\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right) + c\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right)\right] dt + \sigma\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right) \\ &\times \left[\sqrt{\varepsilon}dW - \sigma^T\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right)\left(I + \frac{\partial\xi}{\partial y}\left(Y^{\varepsilon,\delta}, \frac{Y^{\varepsilon,\delta}}{\delta}\right)\right)^T DV(Y^{\varepsilon,\delta}, t)\right]. \end{aligned}$$

Analogous to the *eigenfunction* appearing in importance sampling for *Markov modulated* processes. With $u_1(x,y,t)$ obtain

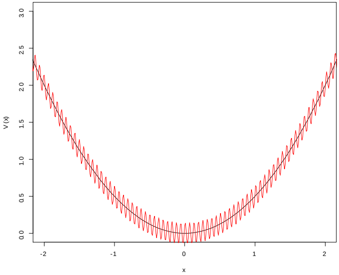
$$\liminf_{\varepsilon\to 0} \varepsilon \log \mathbb{E}^y \left[e^{-\frac{1}{\varepsilon}h(Y^{\epsilon,\delta}(T))}\frac{dP^x}{dP^y}(Y^{\epsilon,\delta}(T))\right]^2 \geq -U(x_0,0) - V(x_0,0).$$

Importance sampling and subsolutions

Numerical example

$$dX^{\varepsilon,\delta} = \left[-\frac{\varepsilon}{\delta}\nabla Q\left(\frac{X^{\varepsilon,\delta}}{\delta}\right) - \nabla S\left(X^{\varepsilon,\delta}\right)\right]dt + \sqrt{\varepsilon}\sqrt{C}dW,$$

$$S(x) = \frac{x^2}{2}, \quad Q(y) = \cos y + \sin y, \quad h(x) = (|x|-1)^2.$$



Importance sampling and subsolutions

No.	ε	δ	ε/δ	$\hat{\theta}_1(\varepsilon,\delta)$	$\hat{\rho}_0(\varepsilon,\delta)$	$\hat{\rho}_1(\varepsilon,\delta)$	$\hat{\rho}_2(\varepsilon,\delta)$
1	0.25	0.1	2.5	$2.25e-01$	1	6	20
2	0.125	0.04	3.125	$3.65e-02$	3	6	5
3	0.0625	0.015625	4	$8.75e-04$	34	4	13
4	0.03125	0.007	4.46	$6.87e-07$	141	3	105
5	0.025	0.004	6.25	$1.61e-08$	217	2	97
6	0.02	0.002	10	$1.99e-10$	1294	1	157
7	0.015	0.0013	11.54	$1.37e-13$	800	1	588

where

$$\hat{\rho}_i(\varepsilon,\delta) \approx \frac{\text{Estimated standard deviation of one sample}}{\theta(\varepsilon,\delta)},$$

and $i=0$ means ordinary Monte Carlo, $i=1$ subsolution with correction, $i=2$ subsolution and no correction.

Formal extension to ergodic coefficients

Consider model

$$dX^{\varepsilon,\delta} = \left[-\frac{\varepsilon}{\delta} \nabla Q \left(\frac{X^{\varepsilon,\delta}}{\delta} \right) - \nabla S \left(X^{\varepsilon,\delta} \right) \right] dt + \sqrt{\varepsilon} \sqrt{C} dW,$$

assume Q is stationary, ergodic, and a.s. C^2 and bounded.

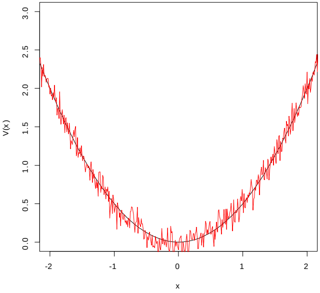
Formal extension to ergodic coefficients

Consider model

$$dX^{\varepsilon,\delta} = \left[-\frac{\varepsilon}{\delta} \nabla Q \left(\frac{X^{\varepsilon,\delta}}{\delta} \right) - \nabla S \left(X^{\varepsilon,\delta} \right) \right] dt + \sqrt{\varepsilon}\sqrt{C} dW,$$

assume Q is stationary, ergodic, and a.s. C^2 and bounded. E.g., $S(x) = x^2/2$ and $Q(y)$ zero mean Gaussian with

$$\mathbb{E}\left[Q(y_1)Q(y_2)\right] = e^{-|y_1 - y_2|^2}.$$



Formal extension to ergodic coefficients

Then save that

$$\int_{\mathbb{T}} e^{-\frac{Q(y)}{C}} dy, \int_{\mathbb{T}} e^{\frac{Q(y)}{C}} dy \quad \text{replaced by} \quad \mathbb{E}\left[e^{-\frac{Q(y)}{C}}\right], \mathbb{E}\left[e^{\frac{Q(y)}{C}}\right],$$

one expects it to go through as before, with

$$u_1(y) = \sqrt{2} e^{\frac{Q(y)}{C}} \Big/ \sqrt{C} \mathbb{E}\left[e^{\frac{Q(y)}{C}}\right], \quad u_2 = \sqrt{2} \Big/ \sqrt{C \mathbb{E}\left[e^{-\frac{Q(y)}{C}}\right] \mathbb{E}\left[e^{\frac{Q(y)}{C}}\right]}.$$

Formal extension to ergodic coefficients

Then save that

$$\int_{\mathbb{T}} e^{-\frac{Q(y)}{C}} dy, \int_{\mathbb{T}} e^{\frac{Q(y)}{C}} dy \quad \text{replaced by} \quad \mathbb{E}\left[e^{-\frac{Q(y)}{C}}\right], \mathbb{E}\left[e^{\frac{Q(y)}{C}}\right],$$

one expects it to go through as before, with

$$u_1(y) = \sqrt{2} e^{\frac{Q(y)}{C}} \Big/ \sqrt{C}\mathbb{E}\left[e^{\frac{Q(y)}{C}}\right], \quad u_2 = \sqrt{2} \Big/ \sqrt{C\mathbb{E}\left[e^{-\frac{Q(y)}{C}}\right]\mathbb{E}\left[e^{\frac{Q(y)}{C}}\right]}.$$

We look at estimating $\mathbb{P}\left\{X^{\varepsilon,\delta} \text{ hits } 0.8 \text{ before } 0.0 | X^{\varepsilon,\delta}(0) = 0.1\right\}$.

No.	ε	δ	ε/δ	$\hat{\theta}_1(\varepsilon,\delta)$	$\hat{\rho}_0(\varepsilon,\delta)$	$\hat{\rho}_1(\varepsilon,\delta)$	$\hat{\rho}_2(\varepsilon,\delta)$
1	0.25	0.1	2.5	$1.38e-1$	3	0.5	3
2	0.125	0.04	3.125	$1.31e-2$	7	16	8
3	0.0625	0.018	3.472	$6.13e-4$	36	18	42
4	0.05	0.01	5	$2.30e-5$	212	28	316
5	0.04	0.007	5.72	$5.93e-6$	396	75	332
6	0.025	0.004	6.25	$7.82e-10$	–	22	1856

References

- A comparison of homogenization and large deviations, with applications to wavefront propagation (M. Freidlin and R. Sowers), *SPA*, Vol. 82, Issue 1, (1999), pp. 23–52.
- *Asymptotic Analysis for Periodic Structures*, (A. Bensoussan, J.L. Lions, and G. Papanicolaou), Vol 5, Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam, 1978.
- Large deviations for multiscale diffusions via weak convergence methods (D. and K. Spiliopoulos), to appear in *SPA*.
- Importance sampling for multiscale diffusions (D., K. Spiliopoulos and H. Wang), to appear in *SIAM J. MMS*.