Toward practical rare event simulation for small noise diffusions

Jonathan Weare

Courant Institute, NYU

June 23, 2010

Goals

1. Generate samples of the rare event
2. Accurately estimate their probability

Applications

- Device reliability
- Options pricing
- Chemical reactions
- Data assimilation?

Goals

1. Generate samples of the rare event
2. Accurately estimate their probability

Applications

- Device reliability
- Options pricing
- Chemical reactions
- Data assimilation?

Example: the Kuroshio current

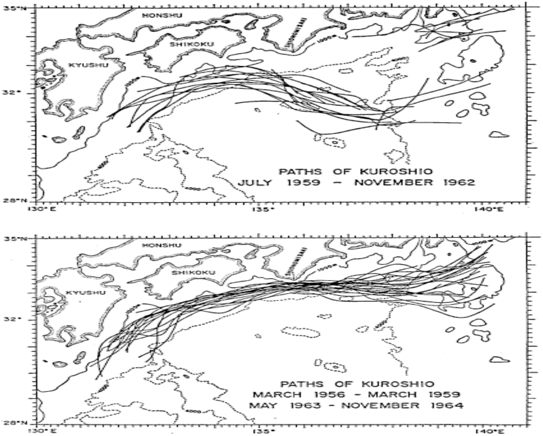


Figure: **Top:** Mean flow paths in the large meander state. **Bottom:** Mean flow paths in the small meander state

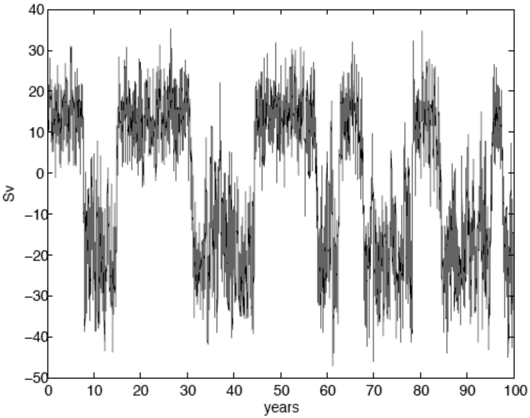


Figure: Projected view of transitions.

Rare event simulation for diffusions

X^ϵ is the solution of the stochastic differential equation

$$dX^\epsilon(t) = b(X^\epsilon(t))\,dt + \sqrt{\epsilon}\,\sigma(X^\epsilon(t))\,dW(t), \quad X^\epsilon(0) = x_0$$

How should we approximate

$$\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right]$$

where g is a functional of the path of X^ϵ?

Or the special case

$$P\left(X^\epsilon \in A\right)$$

If we use the standard Monte Carlo estimator,

$$\delta^\epsilon = \frac{1}{M}\sum_{j=1}^{M} e^{-\frac{1}{\epsilon}g(X_j^\epsilon)} \qquad X_j^\epsilon \quad \text{i.i.d.},$$

the variance is

$$\mathrm{Var}(\delta^\epsilon) = \frac{1}{\sqrt{M}}\left(\mathbf{E}\left[e^{-\frac{1}{\epsilon}2g(X^\epsilon)}\right] - \mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right]^2\right)$$

and the relative error is

$$\frac{\sqrt{\mathrm{Var}(\delta^\epsilon)}}{\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right]} = \frac{1}{\sqrt{M}}\sqrt{R^\epsilon - 1}$$

where

$$R^\epsilon = \frac{\mathbf{E}\left[e^{-\frac{2}{\epsilon}g(X^\epsilon)}\right]}{\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right]^2}$$

We'll focus on

$$R^\epsilon = \frac{\mathbf{E}\left[e^{-\frac{2}{\epsilon}g(X^\epsilon)}\right]}{\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right]^2}$$

First Notice that $R^\epsilon \geq 1$ from Jensen's inequality.

The **Laplace Principle** for X^ϵ gives us constants γ_1 and γ_2 such that

$$\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right] = e^{\frac{-\gamma_1+o(1)}{\epsilon}} \quad \text{and} \quad \mathbf{E}\left[e^{-\frac{1}{\epsilon}2g(X^\epsilon)}\right] = e^{\frac{-\gamma_2+o(1)}{\epsilon}}$$

Therefore

$$R^\epsilon = \exp\left(\frac{2\gamma_1 - \gamma_2 + o(1)}{\epsilon}\right)$$

Since $\gamma_2 \leq 2\gamma_1$ this is very bad news. We'll need exponentially many samples.

We'll focus on

$$R^\epsilon = \frac{\mathbf{E}\left[e^{-\frac{2}{\epsilon}g(X^\epsilon)}\right]}{\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right]^2}$$

First Notice that $R^\epsilon \geq 1$ from Jensen's inequality.

The **Laplace Principle** for X^ϵ gives us constants γ_1 and γ_2 such that

$$\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right] = e^{\frac{-\gamma_1+o(1)}{\epsilon}} \quad \text{and} \quad \mathbf{E}\left[e^{-\frac{1}{\epsilon}2g(X^\epsilon)}\right] = e^{\frac{-\gamma_2+o(1)}{\epsilon}}$$

Therefore

$$R^\epsilon = \exp\left(\frac{2\gamma_1 - \gamma_2 + o(1)}{\epsilon}\right)$$

Since $\gamma_2 \leq 2\gamma_1$ this is very bad news. We'll need exponentially many samples.

More precisely

$$R^\epsilon = \exp\left(\frac{2\gamma_1 - \gamma_2 + o(1)}{\epsilon}\right)$$

where

$$\gamma_1 = \inf_{\substack{\varphi \in \mathcal{AC}([0,T]),\\ \varphi(0)=x_0}} \left\{ \int_0^T \frac{1}{2}\|\sigma^{-1}(\dot{\varphi} - b)\|^2\, ds + g(\varphi) \right\}, \tag{1}$$

and

$$\gamma_2 = \inf_{\substack{\varphi \in \mathcal{AC}([0,T]),\\ \varphi(0)=x_0}} \left\{ \int_0^T \frac{1}{2}\|\sigma^{-1}(\dot{\varphi} - b)\|^2\, ds + 2g(\varphi) \right\}, \tag{2}$$

Data assimilation picture:

Suppose for example that the observation model is of the form

$$Y_n = r\left(X_{t_n}\right) + \sqrt{\epsilon}\,\xi_n,$$

where the ξ_n are i.i.d. Gaussian.
For a particle filter the weights for each particle will look like:

$$w(x) = e^{-\frac{1}{\epsilon}(y-r(x))^2} = e^{-\frac{1}{\epsilon}g_y(x)}$$

The effective sample size of the ensemble is related to the quantity

$$\frac{\mathbf{E}\left[e^{-\frac{1}{\epsilon}g_y(X^\epsilon)}\right]^2}{\mathbf{E}\left[e^{-\frac{2}{\epsilon}g_y(X^\epsilon)}\right]} = \frac{1}{R^\epsilon}$$

Other methods suffer related problems.

Importance sampling.

What's the problem? We generate a huge number of samples that result in near zero values of $e^{-\frac{1}{\epsilon}g(X_j^\epsilon)}$ and if we're lucky we get one sample with a relatively large value of $e^{-\frac{1}{\epsilon}g(X_j^\epsilon)}$.

Solution: Try to "**pull**" the process toward the region where $e^{-\frac{1}{\epsilon}g(X_j^\epsilon)}$ is relatively large.

Instead of sampling the solution, X^ϵ, of

$$dX^\epsilon(t) = b(X^\epsilon(t))\,dt + \sqrt{\epsilon}\,\sigma(X^\epsilon(t))\,dW(t)$$

sample the solution, $\hat{X}^\epsilon$, of

$$d\hat{X}^\epsilon(t) = \left(b(\hat{X}^\epsilon(t)) + \sigma(\hat{X}^\epsilon(t))v(t,\hat{X}^\epsilon(t))\right)\,dt + \sqrt{\epsilon}\,\sigma(\hat{X}^\epsilon(t))\,dW(t).$$

and assign a weight Z^ϵ to each sample so that

$$\mathbf{E}\left[e^{-\frac{1}{\epsilon}g\left(\hat{X}^\epsilon\right)}Z^\epsilon\right] = \mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right].$$

The importance sampling estimator is:

$$\delta^\epsilon = \frac{1}{M} \sum_{j=1}^{M} e^{-\frac{1}{\epsilon} g(\hat{X}_j^\epsilon)} Z_j^\epsilon$$

where $(\hat{X}_j^\epsilon, W_j)$ are independent samples of $(\hat{X}^\epsilon, W)$ and, from Girsanov's formula,

$$Z_j^\epsilon = \exp\left(-\frac{1}{\sqrt{\epsilon}} \int_0^T v(t, \hat{X}_j^\epsilon(t))\, dW_j(t) - \frac{1}{2\epsilon} \int_0^T v(t, \hat{X}_j^\epsilon(t))^2\, dt \right).$$

Now we have

$$\text{rel err} = \frac{1}{\sqrt{M}} \sqrt{R^\epsilon - 1} \quad \text{and} \quad R^\epsilon = \frac{\mathbf{E}\left[e^{-\frac{2}{\epsilon} g(\hat{X}^\epsilon)} (Z^\epsilon)^2 \right]}{\mathbf{E}\left[e^{-\frac{1}{\epsilon} g(X^\epsilon)} \right]^2}$$

We want to choose the function v to make $R^\epsilon \approx 1$.

The importance sampling estimator is:

$$\delta^\epsilon = \frac{1}{M}\sum_{j=1}^{M} e^{-\frac{1}{\epsilon}g(\hat{X}_j^\epsilon)} Z_j^\epsilon$$

where $(\hat{X}_j^\epsilon, W_j)$ are independent samples of $(\hat{X}^\epsilon, W)$ and, from Girsanov's formula,

$$Z_j^\epsilon = \exp\left(-\frac{1}{\sqrt{\epsilon}}\int_0^T v(t, \hat{X}_j^\epsilon(t))\, dW_j(t) - \frac{1}{2\epsilon}\int_0^T v(t, \hat{X}_j^\epsilon(t))^2\, dt\right).$$

Now we have

$$\text{rel err} = \frac{1}{\sqrt{M}}\sqrt{R^\epsilon - 1} \quad \text{and} \quad R^\epsilon = \frac{\mathbf{E}\left[e^{-\frac{2}{\epsilon}g\left(\hat{X}^\epsilon\right)}\left(Z^\epsilon\right)^2\right]}{\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon)}\right]^2}$$

We want to choose the function v to make $R^\epsilon \approx 1$.

Now

$$R^\epsilon \sim \exp\left(\frac{2\gamma_1 - \gamma_2 + o(1)}{\epsilon}\right)$$

where, as before

$$\gamma_1 = \inf_{\substack{\varphi \in \mathcal{AC}([0,T]),\\ \varphi(0)=x_0}} \left\{ \int_0^T \frac{1}{2} \|\sigma^{-1}(\dot{\varphi} - b)\|^2 \, ds + g(\varphi) \right\},$$

and γ_2 is now given by (at least when v is smooth),

$$\gamma_2 = \inf_{\substack{\varphi \in \mathcal{AC}([0,T]),\\ \varphi(0)=x_0}} \left\{ \int_0^T L(s, \varphi(s), \dot{\varphi}(s)) \, ds + 2g(\varphi) \right\}$$

where

$$L(t,x,\beta) = \|\sigma^{-1}(x)(\beta - b(x))\|^2 - \frac{1}{2}\|\sigma^{-1}(x)(\beta - b(x) - \sigma(x)\, v(t,x))\|^2$$

We want to choose the v that makes γ_2 as large as possible.

If we want $R^\epsilon \approx 1$ we had better have that $\gamma_2 = 2\gamma_1$.
Such an estimator is called **log-efficient.**

Note: log-efficiency only implies that $R^\epsilon \sim e^{o(1)/\epsilon}$.

We'll see that it is possible to do much better.

The common "**optimal twist**" method corresponds to the choice

$$v(t,x) = \sigma(x)^{-1}(\dot{\hat{\varphi}}_{0,x_0}(t) - b(x))$$

where

$$\hat{\varphi}_{0,x_0} \in \arg \min_{\substack{\varphi \in \mathcal{AC}([0,T]),\\ \varphi(0)=x_0}} \left\{ \int_0^T \frac{1}{2} \|\sigma^{-1}(\dot{\varphi} - b)\|^2 \, ds + g(\varphi) \right\}.$$

In this case

$$\hat{X}^\epsilon(t) = \hat{\varphi}_{0,x_0}(t) + \sqrt{\epsilon} \int_0^t \sigma(X^\epsilon(s))\, dW(s),$$

One can think of $\hat{\varphi}_{0,x_0}$ as the most likely path of X^ϵ (in the small ϵ limit) when all possible trajectories are reweighted by $e^{-\frac{1}{\epsilon}g(\cdot)}$.

Unfortunately this method is typically not even log-efficient.

The common "**optimal twist**" method corresponds to the choice

$$v(t,x) = \sigma(x)^{-1}(\dot{\hat{\varphi}}_{0,x_0}(t) - b(x))$$

where

$$\hat{\varphi}_{0,x_0} \in \arg \min_{\substack{\varphi \in \mathcal{AC}([0,T]),\\ \varphi(0)=x_0}} \left\{ \int_0^T \frac{1}{2} \|\sigma^{-1}(\dot{\varphi} - b)\|^2 \, ds + g(\varphi) \right\}.$$

In this case

$$\hat{X}^\epsilon(t) = \hat{\varphi}_{0,x_0}(t) + \sqrt{\epsilon} \int_0^t \sigma(X^\epsilon(s)) \, dW(s),$$

One can think of $\hat{\varphi}_{0,x_0}$ as the most likely path of X^ϵ (in the small ϵ limit) when all possible trajectories are reweighted by $e^{-\frac{1}{\epsilon} g(\cdot)}$.

Unfortunately this method is typically not even log-efficient.

Focusing on the case that $g(\varphi) = g(\varphi(T))$, consider the function

$$\Phi^\epsilon(t, x) = \mathbf{E}_{t,x}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon(T))}\right].$$

It's easy to see that the importance sampling estimator of $\mathbf{E}_{0,x_0}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon(T))}\right] = \Phi^\epsilon(0, x_0)$ with

$$v^\epsilon = -\epsilon\frac{\sigma^T\Phi^\epsilon_x}{\Phi^\epsilon}$$

has zero variance, i.e. $R^\epsilon = 1$.

Φ^ϵ solve a linear second order parabolic PDE with terminal condition $\Phi^\epsilon(T, x) = \mathrm{e}^{-\frac{1}{\epsilon}g(x)}$.

Of course there's no hope of finding a global solution of the PDE in more than a few dimensions.

Focusing on the case that $g(\varphi) = g(\varphi(T))$, consider the function

$$\Phi^\epsilon(t,x) = \mathbf{E}_{t,x}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon(T))}\right].$$

It's easy to see that the importance sampling estimator of $\mathbf{E}_{0,x_0}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon(T))}\right] = \Phi^\epsilon(0,x_0)$ with

$$v^\epsilon = -\epsilon\frac{\sigma^T\Phi^\epsilon_x}{\Phi^\epsilon}$$

has zero variance, i.e. $R^\epsilon = 1$.

Φ^ϵ solve a linear second order parabolic PDE with terminal condition $\Phi^\epsilon(T,x) = \mathrm{e}^{-\frac{1}{\epsilon}g(x)}$.

Of course there's no hope of finding a global solution of the PDE in more than a few dimensions.

Instead we'll consider the $\epsilon \to 0$ limit of the log transform of Φ^ϵ,

$$G^\epsilon = -\epsilon \log \Phi^\epsilon$$

which solves the second order Hamilton-Jacobi Equation

$$-G_t^\epsilon - bG_x^\epsilon + \frac{1}{2}\left(\sigma^T G_x^\epsilon\right)^2 - \frac{\epsilon}{2}\sigma\sigma^T G_{xx}^\epsilon = 0, \qquad G^\epsilon(T,x) = g(x) \tag{3}$$

In terms of G^ϵ

$$v^\epsilon = -\sigma^T G_x^\epsilon.$$

So we can set

$$v^0 = -\sigma^T G_x$$

where G is the viscosity solution of

$$-G_t - bG_x + \frac{1}{2}\left(\sigma^T G_x\right)^2 = 0, \qquad G(T,x) = g(x)$$

and hope for the best.

Instead we'll consider the $\epsilon \to 0$ limit of the log transform of Φ^ϵ,

$$G^\epsilon = -\epsilon \log \Phi^\epsilon$$

which solves the second order Hamilton-Jacobi Equation

$$-G^\epsilon_t - bG^\epsilon_x + \frac{1}{2}\left(\sigma^T G^\epsilon_x\right)^2 - \frac{\epsilon}{2}\sigma\sigma^T G^\epsilon_{xx} = 0, \qquad G^\epsilon(T,x) = g(x) \tag{3}$$

In terms of G^ϵ

$$v^\epsilon = -\sigma^T G^\epsilon_x.$$

So we can set

$$v^0 = -\sigma^T G_x$$

where G is the viscosity solution of

$$-G_t - bG_x + \frac{1}{2}\left(\sigma^T G_x\right)^2 = 0, \qquad G(T,x) = g(x)$$

and hope for the best.

G has the control representation

$$G(t,x) = \inf_{\substack{\varphi \in \mathcal{AC}([t,T]),\\ \varphi(t)=x}} \left\{ \int_t^T \frac{1}{2} \|\sigma^{-1}(\dot{\varphi} - b)\|^2 \, ds + g\left(\varphi(T)\right) \right\}.$$

Notice that $\gamma_1 = G(0, x_0)$.

G is the rate appearing in the Laplace Principle.

Furthermore, where G is differentiable,

$$b(t,x) + \sigma(t,x)\, v^0(t,x) = \dot{\hat{\varphi}}_{t,x}(t)$$

where

$$\hat{\varphi}_{t,x} = \arg \min_{\substack{\varphi \in \mathcal{AC}([t,T]),\\ \varphi(t)=x}} \left\{ \int_t^T \frac{1}{2} \|\sigma^{-1}(\dot{\varphi} - b)\|^2 \, ds + g\left(\varphi(T)\right) \right\}.$$

For v^0 we have that

$$d\hat{X}^\epsilon(t) = \dot{\hat{\varphi}}_{t,\hat{X}^\epsilon(t)}\, dt + \sqrt{\epsilon}\,\sigma(\hat{X}^\epsilon(t))\, dW(t).$$

An analogue of this estimator, in a discrete time setup, first appears in

- DUPUIS, P. and WANG, H. (2004). Importance sampling, large deviations, and differential games. *Stochastics.* **76** 481–508.

for problems in which one can compute G by hand.

Our approach: Solve the optimization problem for $\hat{\varphi}_{t,\hat{X}^\epsilon(t)}$ on-the-fly at each point along the trajectory of $\hat{X}^\epsilon$.

This procedure can be carried out at reasonable cost and, as we prove, the estimator has very favorable error properties.

For v^0 we have that

$$d\hat{X}^\epsilon(t) = \dot{\hat{\varphi}}_{t,\hat{X}^\epsilon(t)}\, dt + \sqrt{\epsilon}\,\sigma(\hat{X}^\epsilon(t))\, dW(t).$$

An analogue of this estimator, in a discrete time setup, first appears in

- DUPUIS, P. and WANG, H. (2004). Importance sampling, large deviations, and differential games. *Stochastics.* **76** 481–508.

for problems in which one can compute G by hand.

Our approach: Solve the optimization problem for $\hat{\varphi}_{t,\hat{X}^\epsilon(t)}$ on-the-fly at each point along the trajectory of $\hat{X}^\epsilon$.

This procedure can be carried out at reasonable cost and, as we prove, the estimator has very favorable error properties.

Illustration of vanishing error :

$$X^\epsilon = \sqrt{\epsilon}\, W, \qquad x_0 = 0.1$$

Estimate $\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon(1))}\right]$ where $g(x) = \begin{cases} \frac{1}{2}(1-x)^2, & x \geq 0, \\ \frac{1}{2}(1+x)^2, & x < 0. \end{cases}$

ϵ	standard MC R^ϵ	optimal twist R^ϵ	new R^ϵ	our estimate	$\mathbf{E}\left[e^{-\frac{1}{\epsilon}g(X^\epsilon(1))}\right]$
1	1.0340	1.2564	1.1746	0.8368	0.8369
2^{-1}	1.0800	1.6636	1.3494	0.7225	0.7227
2^{-2}	1.3084	3.5982	1.6971	0.4848	0.4852
2^{-3}	2.2672	25.526	2.2903	0.1983	0.1986
2^{-4}	7.7807	977.66	2.5990	0.3316×10^{-1}	0.3323×10^{-1}
2^{-5}	81.266	–	1.5193	0.1127×10^{-2}	0.1129×10^{-2}
2^{-6}	6008.4	–	1.0200	0.1666×10^{-5}	0.1666×10^{-5}

$M = 10^9$

What happens at the discontinuities of v^0?

$$X^\epsilon = \sqrt{\epsilon}\, W, \qquad x_0 = 0$$

Estimate $\mathbf{E}\left[e^{-\frac{1}{\epsilon} g(X^\epsilon(1))}\right]$ where $g(x) = \begin{cases} \frac{1}{2}(1-x)^2, & x \geq 0, \\ \frac{1}{2}(1+x)^2, & x < 0. \end{cases}$

ϵ	standard MC R^ϵ	optimal twist R^ϵ	new R^ϵ	our estimate	$\mathbf{E}\left[e^{-\frac{1}{\epsilon} g(X^\epsilon(1))}\right]$
1	1.0336	1.3034	1.1762	0.8372	0.8373
2^{-1}	1.0800	1.8775	1.3714	0.7212	0.7217
2^{-2}	1.3110	5.2088	1.7604	0.4805	0.4793
2^{-3}	2.2918	65.397	2.4711	0.1868	0.1870
2^{-4}	8.3531	8805.1	3.3540	0.2607×10^{-1}	0.2584×10^{-1}
2^{-5}	121.08	–	4.7635	0.4715×10^{-3}	0.4744×10^{-3}
2^{-6}	18596	–	5.6141	0.1574×10^{-6}	0.1591×10^{-6}

$M = 10^9$

Theoretical Issues:

For any chosen discretization scheme

$$\text{relative error} \approx \sqrt{\begin{matrix}\text{statistical}\\ \text{error}\end{matrix} + \begin{matrix}\text{relative}\\ \text{bias}\end{matrix}}$$

For the Euler discretization we prove

Theorem

If G is smooth on $[0, T]$, *then*

$$\begin{matrix}\textit{statistical}\\ \textit{error}\end{matrix} \sim \mathcal{O}\left(\frac{\triangle}{\epsilon}\right) + \mathcal{O}(\epsilon) \quad \textit{and} \quad \begin{matrix}\textit{relative}\\ \textit{bias}\end{matrix} \sim \mathcal{O}\left(\frac{\triangle}{\epsilon}\right)$$

* *If G is smooth on* $[0, T)$, *then*

$$\begin{matrix}\textit{statistical}\\ \textit{error}\end{matrix} \sim C + \mathcal{O}\left(\frac{\triangle}{\epsilon}\right) + \mathcal{O}(\epsilon) \quad \textit{and} \quad \begin{matrix}\textit{relative}\\ \textit{bias}\end{matrix} \sim \mathcal{O}\left(\frac{\triangle}{\epsilon}\right)$$

One can decrease the step size $\triangle$ algebraically with ϵ instead of exponentially.

A nontrivial test problem:

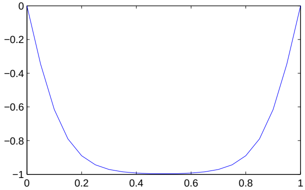
$$u_t^\epsilon = \nu u_{xx}^\epsilon - \frac{1}{\nu} V'(u^\epsilon) + \sqrt{\epsilon}\,\eta$$

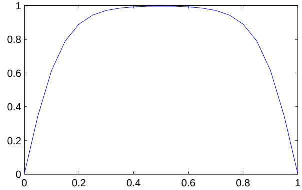
where η is a space-time white noise, $\nu > 0$ is a small parameter, and

$$V(u) = (1 - u^2)^2.$$

The deterministic equation has two steady states:

$$u_- \approx -1 \qquad \text{and} \qquad u_+ \approx +1$$





We'll try to approximate

$$p^\epsilon = P\left(\int_0^1 u^\epsilon(T,x)\,dx \geq 0\right)$$

A classic paper by Faris and Jona-Lasinio shows that the Large Deviations action functional for u^ϵ is

$$I(u) = \int_0^T \int_0^1 \left(u_t - \nu u_{xx} + \frac{1}{\nu}V'(u)\right)^2 \, dxdt,$$

i.e. that

$$-\epsilon \log p^\epsilon \longrightarrow \inf_{\substack{u:\, u(0,\cdot)=u_- \\ \int_0^1 u(T,x)\,dx \geq 0}} I(u).$$

Outline of a continuation strategy:

1. Compute the first few local minimizers of $I(u)$ for the initial state $\hat{X}^\epsilon(0) = u_-$ with time horizon T.
2. Compute $v^0(0, \hat{X}^\epsilon(0))$ using the local minimizer with lowest value of $I(u)$.
3. Compute $\hat{X}^\epsilon(\triangle)$.
4. Using the local minimizers computed in step 1 as initial conditions find new local minimizers given the state $\hat{X}^\epsilon(\triangle)$ and a time horizon of $T - \triangle$.
5. ...

We track each local min by continuation and assume that the global minimizer is one of these local min.

number of local min tracked	relative error	our estimate
2	0.28	9.509×10^{-9}
3	0.28	9.346×10^{-9}
4	0.28	9.462×10^{-9}

$M = 100$, $\nu = 0.05$, and $\epsilon = 0.1$

Revisiting the optimal twist

If we assume that the initial condition is $N\left(a, \frac{1}{\epsilon}\Gamma\Gamma^T\right)$, the Large Deviations rate functional becomes

$$\mathcal{F}(\varphi) = \frac{1}{2}\|\Gamma^{-1}(\varphi(0) - a)\|^2 + \int_0^T \frac{1}{2}\|\sigma^{-1}(\dot{\varphi} - b(\varphi))\|^2 \, ds + g(\varphi)$$

If $\mathcal{F}$ is convex then the optimal twist **is** log-efficient.

Is this convexity reasonable for weather/climate data assimilation problems?

Note that this doesn't require that the predictive distribution is log-convex.

Revisiting the optimal twist

If we assume that the initial condition is $N\left(a, \frac{1}{\epsilon}\Gamma\Gamma^{T}\right)$, the Large Deviations rate functional becomes

$$\mathcal{F}(\varphi) = \frac{1}{2}\|\Gamma^{-1}(\varphi(0) - a)\|^2 + \int_0^T \frac{1}{2}\|\sigma^{-1}(\dot{\varphi} - b(\varphi))\|^2\, ds + g\left(\varphi\right)$$

If $\mathcal{F}$ is convex then the optimal twist **is** log-efficient.

Is this convexity reasonable for weather/climate data assimilation problems?

Note that this doesn't require that the predictive distribution is log-convex.

Incorporating optimal twist in ensemble schemes

For example we could

1. Approximate the current posterior distribution by an $N(a, \Gamma\Gamma^T)$.
2. Minimize $\mathcal{F}$.
3. Generate samples using the control found in step 2 and calculate the first few moments of the next posterior.
4. Weight and resample (as in a particle filter) or transform the samples (as in an ensemble Kalman filter).
5. Set a to be the new posterior mean and $\Gamma\Gamma^T$ to be the new covariance matrix... repeat.