Importance sampling in the neighborhood of a stable equilibrium point

Paul Dupuis

Division of Applied Mathematics
Brown University

(Kostas Spiliopoulos, Xiang Zhou)

RESIM

26 June 2012

Problem of interest

Consider a process model such as

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

and assume $x = 0$ is a stable equilibrium for $\dot{\phi} = b(\phi)$ with domain of attraction $\mathcal{D}$.

Problem of interest

Consider a process model such as

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

and assume $x = 0$ is a stable equilibrium for $\dot{\phi} = b(\phi)$ with domain of attraction $\mathcal{D}$. We want to estimate quantities related to metastability.

Problem of interest

Consider a process model such as

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

and assume $x = 0$ is a stable equilibrium for $\dot{\phi} = b(\phi)$ with domain of attraction $\mathcal{D}$. We want to estimate quantities related to metastability.

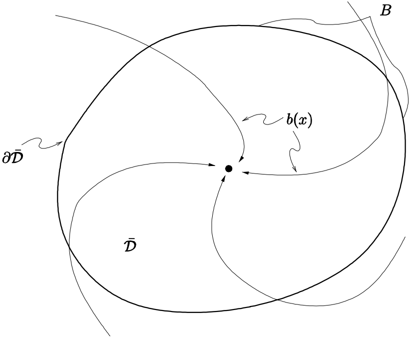
Example.

$$p_T^{\varepsilon} \doteq P\left\{X^{\varepsilon}(t) \text{ exits } \bar{\mathcal{D}} \text{ through } B \text{ before time } T \,|X^{\varepsilon}(0) = 0\right\},$$

where $B \subset \partial\bar{\mathcal{D}}$ and $\bar{\mathcal{D}} \subset \mathcal{D}$, and T is large.

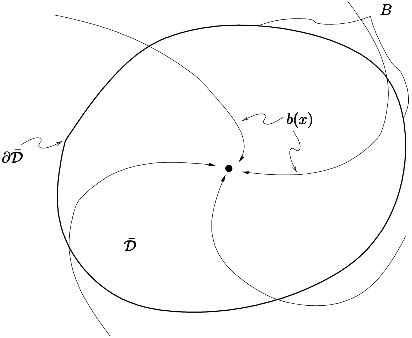
Problem of interest

$$p_T^\varepsilon \doteq P\left\{X^\varepsilon(t) \text{ exits } \bar{\mathcal{D}} \text{ through } B \text{ before time } T \,|X^\varepsilon(0)=0\right\}$$



Problem of interest

$$p_T^\varepsilon \doteq P\left\{X^\varepsilon(t) \text{ exits } \bar{\mathcal{D}} \text{ through } B \text{ before time } T \,|X^\varepsilon(0) = 0\right\}$$



Also of interest: other process models, e.g., jump Markov. We focus on diffusion.

Problem of interest

This type of problem is ubiquitous, and arises whenever metastability issues are important.

Problem of interest

This type of problem is ubiquitous, and arises whenever metastability issues are important.

Examples:

- chemical reaction rates,
- genetic switch models in biology,
- performance analysis of recursive stochastic algorithms (stochastic approximation),
- ...

Problem of interest

This type of problem is ubiquitous, and arises whenever metastability issues are important.

Examples:

- chemical reaction rates,
- genetic switch models in biology,
- performance analysis of recursive stochastic algorithms (stochastic approximation),
- ...

Note the distinction in comparison to canonical problem studied in rare event simulation: the problem is time dependent and the stable point is in the domain of the event.

One possible approach

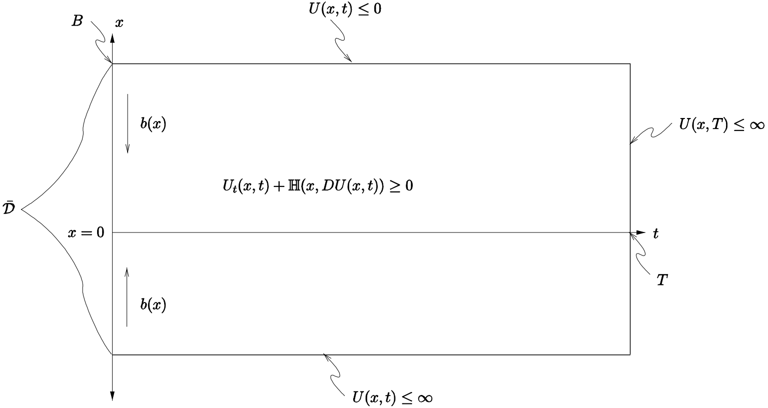
PDE/subsolutions approach to the design of importance sampling.

One possible approach

PDE/subsolutions approach to the design of importance sampling. Let

$$\mathbb{H}(x,p) = \langle p, b(x) \rangle - \frac{1}{2} \|\sigma(x)^T p\|^2$$

and let $U(x,t)$ satisfy



One possible approach

Theorem. Let $U(x,t)$ be a subsolution. Simulate

$$d\bar{X}^{\varepsilon}(t) = b(\bar{X}^{\varepsilon}(t))dt - \sigma(\bar{X}^{\varepsilon}(t))^{T} DU(\bar{X}^{\varepsilon}(t),t)dt + \sqrt{\varepsilon}\sigma(\bar{X}^{\varepsilon}(t))dW(t)$$

with $\bar{X}^{\varepsilon}(0) = 0$ and use samples

$$1\left\{\bar{\tau}^{\varepsilon} \leq T, \bar{X}^{\varepsilon}(\bar{\tau}^{\varepsilon}) \in B\right\} R^{\varepsilon}(\bar{X}^{\varepsilon}),$$

where $\bar{\tau}^{\varepsilon}$ is time of escape from $\bar{\mathcal{D}}$ and $R^{\varepsilon}(\bar{X}^{\varepsilon})$ is likelihood ratio of distribution of X^{ε} w.r.t $\bar{X}^{\varepsilon}$.

One possible approach

Theorem. Let $U(x,t)$ be a subsolution. Simulate

$$d\bar{X}^{\varepsilon}(t) = b(\bar{X}^{\varepsilon}(t))dt - \sigma(\bar{X}^{\varepsilon}(t))^{T} DU(\bar{X}^{\varepsilon}(t),t)dt + \sqrt{\varepsilon}\sigma(\bar{X}^{\varepsilon}(t))dW(t)$$

with $\bar{X}^{\varepsilon}(0) = 0$ and use samples

$$1\left\{\bar{\tau}^{\varepsilon} \le T, \bar{X}^{\varepsilon}(\bar{\tau}^{\varepsilon}) \in B\right\} R^{\varepsilon}(\bar{X}^{\varepsilon}),$$

where $\bar{\tau}^{\varepsilon}$ is time of escape from $\bar{\mathcal{D}}$ and $R^{\varepsilon}(\bar{X}^{\varepsilon})$ is likelihood ratio of distribution of X^{ε} w.r.t $\bar{X}^{\varepsilon}$. If $p^{\varepsilon}_{T} \approx e^{-\frac{1}{\varepsilon}\gamma_T}$, then

$$\limsup_{\varepsilon \to 0} \varepsilon \log E\left[1\left\{\bar{\tau}^{\varepsilon} \le T, \bar{X}^{\varepsilon}(\bar{\tau}^{\varepsilon}) \in B\right\} R^{\varepsilon}(\bar{X}^{\varepsilon})\right]^{2} \le -U(0,0) - \gamma_T.$$

One possible approach

Theorem. Let $U(x,t)$ be a subsolution. Simulate

$$d\bar{X}^{\varepsilon}(t)=b(\bar{X}^{\varepsilon}(t))dt-\sigma(\bar{X}^{\varepsilon}(t))^{T}DU(\bar{X}^{\varepsilon}(t),t)dt+\sqrt{\varepsilon}\sigma(\bar{X}^{\varepsilon}(t))dW(t)$$

with $\bar{X}^{\varepsilon}(0)=0$ and use samples

$$1\left\{\bar{\tau}^{\varepsilon}\leq T,\bar{X}^{\varepsilon}(\bar{\tau}^{\varepsilon})\in B\right\}R^{\varepsilon}(\bar{X}^{\varepsilon}),$$

where $\bar{\tau}^{\varepsilon}$ is time of escape from $\bar{\mathcal{D}}$ and $R^{\varepsilon}(\bar{X}^{\varepsilon})$ is likelihood ratio of distribution of X^{ε} w.r.t $\bar{X}^{\varepsilon}$. If $p_{T}^{\varepsilon}\approx e^{-\frac{1}{\varepsilon}\gamma_{T}}$, then

$$\limsup_{\varepsilon\rightarrow 0}\varepsilon\log E\left[1\left\{\bar{\tau}^{\varepsilon}\leq T,\bar{X}^{\varepsilon}(\bar{\tau}^{\varepsilon})\in B\right\}R^{\varepsilon}(\bar{X}^{\varepsilon})\right]^{2}\leq -U(0,0)-\gamma_{T}.$$

First question: Where to get good subsolutions?

One possible approach

When T is large the Freidlin-Wentsell quasipotential provides a *nearly optimal, time independent* subsolution. Let

$$Q(x) = \inf \left\{ \int_0^\tau \frac{1}{2} \left(\dot{\phi} - b(\phi) \right)^T \left[\sigma(\phi) \sigma(\phi)^T \right] \left(\dot{\phi} - b(\phi) \right) dt \right.$$

$$\left. : \phi(0) = 0, \phi(\tau) = x, \tau < \infty \right\}.$$

One possible approach

When T is large the Freidlin-Wentsell quasipotential provides a *nearly optimal, time independent* subsolution. Let

$$Q(x) = \inf \left\{ \int_0^\tau \frac{1}{2} \left(\dot{\phi} - b(\phi) \right)^T \left[\sigma(\phi)\sigma(\phi)^T \right] \left(\dot{\phi} - b(\phi) \right) dt \right.$$
$$\left. : \phi(0) = 0, \phi(\tau) = x, \tau < \infty \right\}.$$

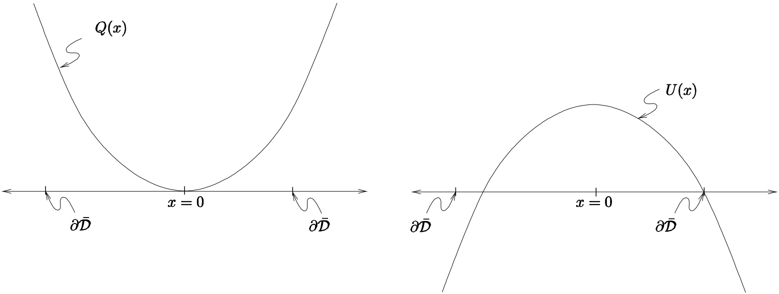
Then Q is available in explicit form for important classes of problems, and

$$U(x) = -Q(x) + \inf_{y \in B} Q(y)$$

is a subsolution with $U(0)$ slightly smaller than γ_T for T large.

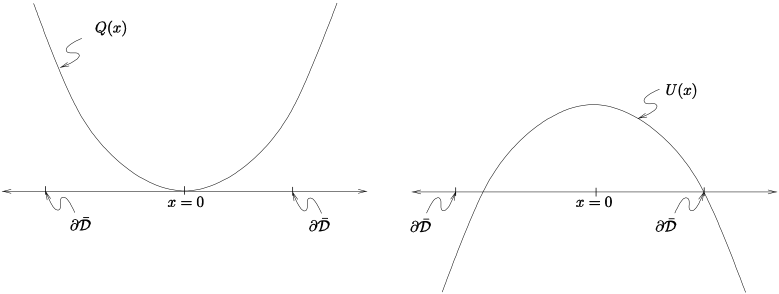
One possible approach

Easy to check that $Q(0) = 0$ and $\mathbb{H}(x, -DQ(x)) = 0$. Then adding $\inf_{y \in B} Q(y)$ maximizes $U(0)$ subject to boundary conditions:



One possible approach

Easy to check that $Q(0) = 0$ and $\mathbb{H}(x, -DQ(x)) = 0$. Then adding $\inf_{y \in B} Q(y)$ maximizes $U(0)$ subject to boundary conditions:



$$\gamma_T - U(0) = (\text{inf over } [0, T]) - (\text{inf over } [0, \infty)) \,.$$

Problem with the quasipotential subsolution

The induced importance sampling scheme has the proper decay rate $U(0) + \gamma_T$, but performance not as good as one expects. Reason for poor performance: dual exponential scalings, one in $1/\varepsilon$, one in T.

Problem with the quasipotential subsolution

The induced importance sampling scheme has the proper decay rate $U(0) + \gamma_T$, but performance not as good as one expects. Reason for poor performance: dual exponential scalings, one in $1/\varepsilon$, one in T.

Example. 1D Gauss-Markov with $\bar{\mathcal{D}} = [-1, 1], B = \partial\bar{\mathcal{D}}, c = 1, \bar{\sigma} = 1$:

$$dX^{\varepsilon}(t) = -cX^{\varepsilon}(t)dt + \sqrt{\varepsilon}\bar{\sigma}dW(t).$$

Relative errors:

T vs ε	0.25	0.5	1	1.5	2.5	10	14	18
0.20	91	7	2	1	1	10	51	179
0.16	253	10	2	1	1	10	48	139
0.13	748	16	3	1	1	9	48	378
0.11	1594	26	3	1	1	10	42	272
0.09	–	49	4	2	1	9	43	357
0.07	–	127	5	2	1	8	47	251
0.05	–	714	8	2	1	8	42	145

Problem with the quasipotential subsolution

A non-asymptotic representation for the 2nd moment. Given control process v let $\hat{X}^{\varepsilon}(0) = 0$ and

$$d\hat{X}^{\varepsilon}(t) = -c\hat{X}^{\varepsilon}(t)dt + \bar{\sigma}DU(\hat{X}^{\varepsilon}(t))dt + v(t)dt + \sqrt{\varepsilon}\bar{\sigma}dW(t).$$

Problem with the quasipotential subsolution

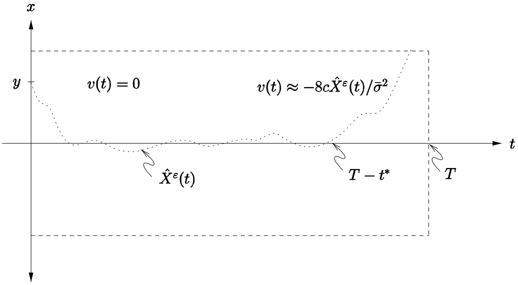
A non-asymptotic representation for the 2nd moment. Given control process v let $\hat{X}^{\varepsilon}(0)=0$ and

$$d\hat{X}^{\varepsilon}(t)=-c\hat{X}^{\varepsilon}(t)dt+\bar{\sigma}DU(\hat{X}^{\varepsilon}(t))dt+v(t)dt+\sqrt{\varepsilon}\bar{\sigma}dW(t).$$

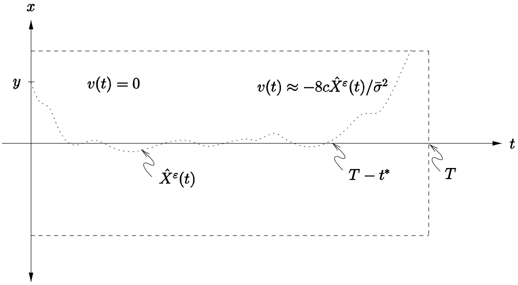
Based on a representation theorem (Boué-D, *Annals of Probability*, 1998) we have the *non-asymptotic* bound

$$\begin{aligned}&-\varepsilon\log E\left[1\left\{\bar{\tau}^{\varepsilon}\leq T,\bar{X}^{\varepsilon}(\bar{\tau}^{\varepsilon})\in B\right\}R^{\varepsilon}(\bar{X}^{\varepsilon})\right]^{2}\\ &\quad=\inf_{v}E\left[\frac{1}{2}\int_{0}^{\hat{\tau}^{\varepsilon}}\left|v(s)\right|^{2}ds-\int_{0}^{\hat{\tau}^{\varepsilon}}\left|DU(\hat{X}^{\varepsilon}(s))\right|^{2}ds+\infty 1\left\{\hat{\tau}^{\varepsilon}>T\right\}\right].\end{aligned}$$

Problem with the quasipotential subsolution



Problem with the quasipotential subsolution



Using the control above one obtains

$$E\left[\frac{1}{2}\int_0^{\hat{\tau}^\varepsilon}|v(s)|^2\,ds-\int_0^{\hat{\tau}^\varepsilon}|DU(X^\varepsilon(s))|^2\,ds+\infty 1\left\{\hat{\tau}^\varepsilon>T\right\}\right]\leq-\varepsilon C_1[T-t^*]+C_2,$$

and thus

$$E\left[1\left\{\bar{\tau}^\varepsilon\leq T,\bar{X}^\varepsilon(\bar{\tau}^\varepsilon)\in B\right\}R^\varepsilon(\bar{X}^\varepsilon)\right]^2\geq e^{C_1[T-t^*]}e^{-\frac{1}{\varepsilon}C_2}.$$

Resolution of the problem

- One can show same problem for any *time independent* importance sampling scheme.

Resolution of the problem

- One can show same problem for any *time independent* importance sampling scheme.
- Bounds that are uniform in T can be obtained when rest point excluded from domain.

Resolution of the problem

- One can show same problem for any *time independent* importance sampling scheme.
- Bounds that are uniform in T can be obtained when rest point excluded from domain.
- Difficulty occurs only in small neighborhood (vanishing as $\varepsilon \to 0$) of rest point.

Resolution of the problem

- One can show same problem for any *time independent* importance sampling scheme.
- Bounds that are uniform in T can be obtained when rest point excluded from domain.
- Difficulty occurs only in small neighborhood (vanishing as $\varepsilon \to 0$) of rest point.
- Can obtain uniform bounds if we can construct a subsolution to the PDE

Resolution of the problem

- One can show same problem for any *time independent* importance sampling scheme.
- Bounds that are uniform in T can be obtained when rest point excluded from domain.
- Difficulty occurs only in small neighborhood (vanishing as $\varepsilon \to 0$) of rest point.
- Can obtain uniform bounds if we can construct a subsolution to the PDE

Resolution of the problem

- One can show same problem for any *time independent* importance sampling scheme.
- Bounds that are uniform in T can be obtained when rest point excluded from domain.
- Difficulty occurs only in small neighborhood (vanishing as $\varepsilon \to 0$) of rest point.
- Can obtain uniform bounds if we can construct a subsolution to the PDE

$$U_t^\varepsilon(x,t) + \langle DU^\varepsilon(x,t), b(x)\rangle - \frac{1}{2}\left\|\sigma(x)DU^\varepsilon(x,t)\right\|^2 + \frac{\varepsilon}{2}\mathrm{tr}\left[\sigma(x)\sigma^T(x)D^2U^\varepsilon(x,t)\right] = 0.$$

Resolution of the problem

Resolution:

Resolution of the problem

Resolution:

- Combine quasipotential subsolution away from rest point with explicit solution to an approximating problem near rest point (linear quadratic regulator).

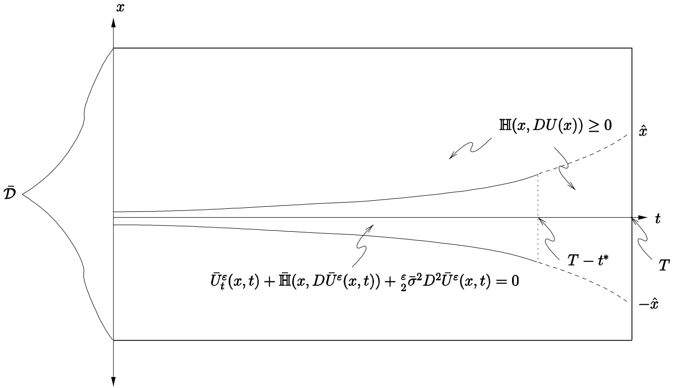
Resolution of the problem

Resolution:

- Combine quasipotential subsolution away from rest point with explicit solution to an approximating problem near rest point (linear quadratic regulator).
- Construct time and ε-dependent subsolution with bounds that are (proveably) uniform in T.

Construction of a time-dependent subsolution

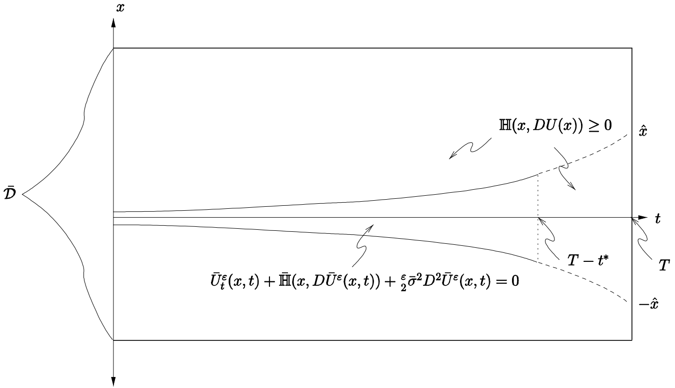
Basic form of the partition:



$\bar{\mathbb{H}}$ is the Hamiltonian of the *linearized* system.

Construction of a time-dependent subsolution

Basic form of the partition:



$\bar{\mathbb{H}}$ is the Hamiltonian of the *linearized* system. A heuristic motivation: better approximation to the zero-variance scheme for the time dependent problem, which previously was poorly approximated near $x = 0$.

Construction of a time-dependent subsolution

Consider one-dimensional problem, with linearization $c = b^{'}(0)$, $\bar{\sigma} = \sigma(0)$. For $0 \leq x \leq e^{c(t-T)}$, let $\bar{U}^{\varepsilon}(x,t)$ be solution to linear quadratic regulator with terminal condition $\bar{U}^{\varepsilon}(x,T) = M\left(x - \hat{x}\right)^{2}/2 + U\left(\hat{x}\right)$.

Construction of a time-dependent subsolution

Consider one-dimensional problem, with linearization $c = b^{'}(0)$, $\bar{\sigma} = \sigma(0)$. For $0 \leq x \leq e^{c(t-T)}$, let $\bar{U}^{\varepsilon}(x,t)$ be solution to linear quadratic regulator with terminal condition $\bar{U}^{\varepsilon}(x,T) = M(x-\hat{x})^2/2 + U(\hat{x})$. Thus

$$\begin{aligned} \bar{U}^{\varepsilon}(x,t) &= a^M(t)\left(x - \hat{x}e^{-c(t-T)}\right)^2 + U(\hat{x}) \\ &\quad -\frac{\varepsilon}{2}\log\left((2c/M+\bar{\sigma}^2) - \bar{\sigma}^2 e^{2c(t-T)}\right), \\ a^M(t) &= \frac{ce^{2c(t-T)}}{(2c/M+\bar{\sigma}^2) - \bar{\sigma}^2 e^{2c(t-T)}} > 0. \end{aligned}$$

Construction of a time-dependent subsolution

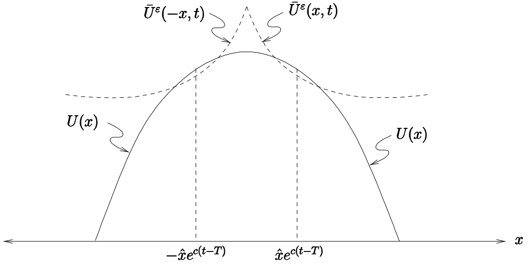
Consider one-dimensional problem, with linearization $c = b^{'}(0)$, $\bar{\sigma} = \sigma(0)$. For $0 \leq x \leq e^{c(t-T)}$, let $\bar{U}^{\varepsilon}(x,t)$ be solution to linear quadratic regulator with terminal condition $\bar{U}^{\varepsilon}(x,T) = M\left(x - \hat{x}\right)^{2}/2 + U\left(\hat{x}\right)$. Thus

$$
\begin{aligned}
\bar{U}^{\varepsilon}(x,t) &= a^{M}(t)\left(x - \hat{x}e^{-c(t-T)}\right)^{2} + U\left(\hat{x}\right) \\
&\quad -\frac{\varepsilon}{2}\log\left(\left(2c/M + \bar{\sigma}^{2}\right) - \bar{\sigma}^{2}e^{2c(t-T)}\right), \\
a^{M}(t) &= \frac{ce^{2c(t-T)}}{\left(2c/M + \bar{\sigma}^{2}\right) - \bar{\sigma}^{2}e^{2c(t-T)}} > 0.
\end{aligned}
$$

For $-e^{c(t-T)} \leq x \leq 0$ have $\bar{U}^{\varepsilon}(x,t) = \bar{U}^{\varepsilon}(-x,t)$. Here $\hat{x}$ and M parameters used in the design to control region where $\bar{U}^{\varepsilon}(x,t)$ will define subsolution.

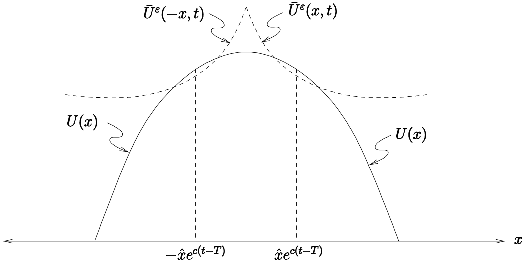
Construction of a time-dependent subsolution

Number of details in merging various functions. *Most important issue:* need to assemble time dependent subsolution as minima of several functions.



Construction of a time-dependent subsolution

Number of details in merging various functions. *Most important issue:* need to assemble time dependent subsolution as minima of several functions.



Second derivatives are $-\infty$ at handoff between functions. Causes problems and needs to be avoided.

Construction of a time-dependent subsolution

We use the *exponential mollification*

$$\bar{U}^{\delta}(t,x) = -\delta \log \left(\sum_{i=1}^{K} e^{-\frac{1}{\delta} F^{i}(t,x)} \right)$$

for various F^i ($= \bar{U}^{\varepsilon}(x,t), \bar{U}^{\varepsilon}(-x,t)$, etc.).

Construction of a time-dependent subsolution

We use the *exponential mollification*

$$\bar{U}^{\delta}(t,x) = -\delta \log \left(\sum_{i=1}^{K} e^{-\frac{1}{\delta} F^{i}(t,x)} \right)$$

for various F^i ($= \bar{U}^{\varepsilon}(x,t), \bar{U}^{\varepsilon}(-x,t)$, etc.).

First introduced for importance sampling in (D. and Wang, *Mathematics of Operations Research*, 2007), and to obtain non-aymptotic bounds in (Blanchet, Glynn and Leder, *TOMACS*, 2011).

Construction of a time-dependent subsolution

Lemma. We have the bounds

$$\wedge_{i=1}^{K} F^i(t,x) \geq \bar{U}^\delta(t,x) \geq \wedge_{i=1}^{K} F^i(t,x) - \delta \log K.$$

Suppose

$$F_t^i(x,t) + \left\langle DF^i(x,t), b(x) \right\rangle - \frac{1}{2} \left\| \sigma(x) DF^i(x,t) \right\|^2 + \frac{\varepsilon}{2} \mathrm{tr} \left[\sigma(x)\sigma^T(x) D^2 F^i(x,t) \right] \\ \geq \gamma^i(x,t,\varepsilon)$$

and let

$$\rho^i(x,t;\delta) = e^{-\frac{1}{\delta}F^i(t,x)} \Big/ \sum_{j=1}^{K} e^{-\frac{1}{\delta}F^j(t,x)}.$$

Then for $\delta \geq \varepsilon$

$$\bar{U}_t^\delta(x,t) + \left\langle D\bar{U}^\delta(x,t), b(x) \right\rangle - \frac{1}{2} \left\| \sigma(x) D\bar{U}^\delta(x,t) \right\|^2 + \frac{\varepsilon}{2} \mathrm{tr} \left[\sigma(x)\sigma^T(x) D^2 \bar{U}^\delta(x,t) \right] \\ \geq \sum_{i=1}^{K} \rho^i(x,t;\delta)\gamma^i(x,t,\varepsilon).$$

Properties of the time-dependent subsolution

Using the mollification with $\delta = \varepsilon$ to construct the subsolution and then applying a verification argument, we get the *non-asymptotic* bound

$$\left(1 - \frac{\varepsilon\bar{\sigma}^2}{4c\hat{x}^2}e^{2ct^*}\right)\left[U(0) - \frac{4c}{\bar{\sigma}^2}\hat{x}^2 e^{-2ct^*}\right] - \varepsilon\log 2$$
$$+ \frac{\varepsilon}{4\bar{\sigma}}\left[\log\left(\frac{2c}{M} + \sigma^2 - \frac{\sigma^2 c}{2c + M\bar{\sigma}^2}\right) - \log\left(\frac{2c}{M} + \sigma^2\right)\right]$$
$$+ \frac{\varepsilon}{4}\log\left(\frac{c}{2\left(2c + M\bar{\sigma}^2\right)}\right).$$

Properties of the time-dependent subsolution

Previous example with $M = 1$, $t^* = 0.9$, $\delta = \varepsilon$, and $\hat{x} \approx .5$:

T vs ε	1	1.5	2.5	10	14	18
0.20	2	2	2	1	1	1
0.16	2	2	2	1	1	1
0.13	3	3	2	1	1	1
0.11	3	4	2	1	1	1
0.09	4	4	3	2	1	1
0.07	5	5	3	2	2	1
0.05	9	8	4	2	2	2

Ongoing work

- Are there better ways to put the components together to build subsolutions?

Ongoing work

- Are there better ways to put the components together to build subsolutions?
- Extension to the non-Gaussian case.