Subsolutions of a Hamilton-Jacobi-Bellman Equation and the Design of Rare Event Simulation Methods

Paul Dupuis

Division of Applied Mathematics
Brown University

IMA

(A. Budhiraja, A. Buijsrogge, T. Dean, D. Johnson, K. Leder, D. Sezer, M. Snarski, K. Spiliopoulos, H. Wang)

May 2018

Outline

- Examples
- Monte Carlo and rare event considerations
- Two methods: Importance sampling and splitting
- Importance Functions
- Some heuristics
- Why Importance Functions should be subsolutions
- Performance for schemes based on subsolutions
- Remarks (construction of subsolutions, proofs, ...)
- A situation where importance sampling and splitting differ greatly

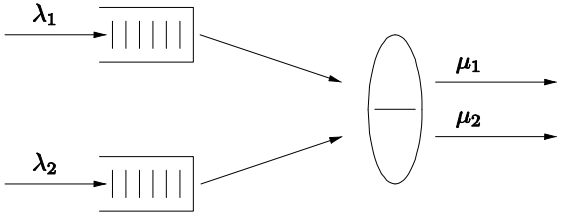
Detailed exposition in Chapters 14-17 of

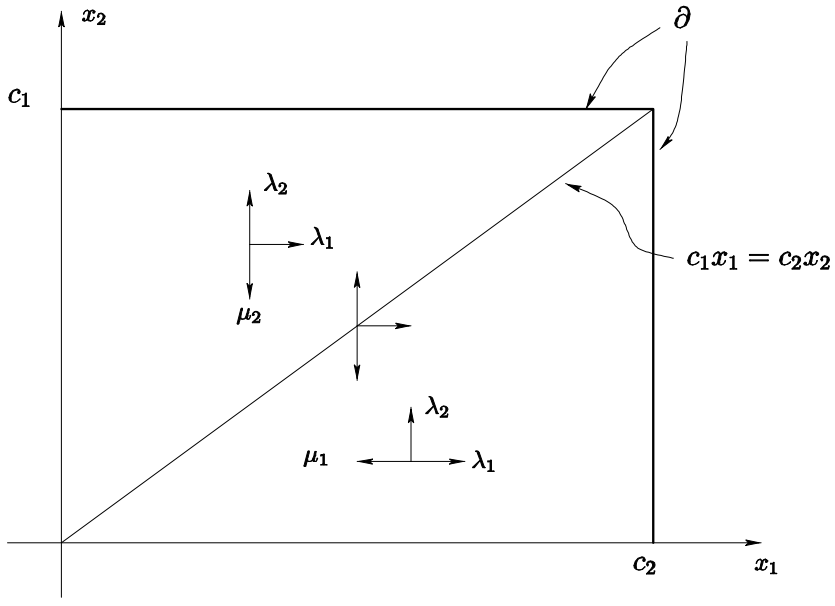
Representations and Weak Convergence Methods for the Analysis and Approximation of Rare Events, A. Budhiraja and D, Springer-Verlag, 2018.

Examples: General framework

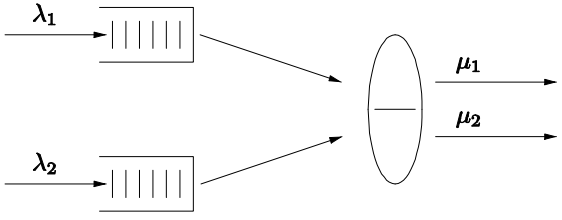
- Monte Carlo estimation of probabilities and expected values largely determined by rare events.
- Stochastic processes with light-tailed random variables.
- Typical examples: exit problems, risk-sensitive functionals, functionals of invariant distributions with simple structure.
- Exploit a law of large numbers (LLN) scaling, continuous-time limit.

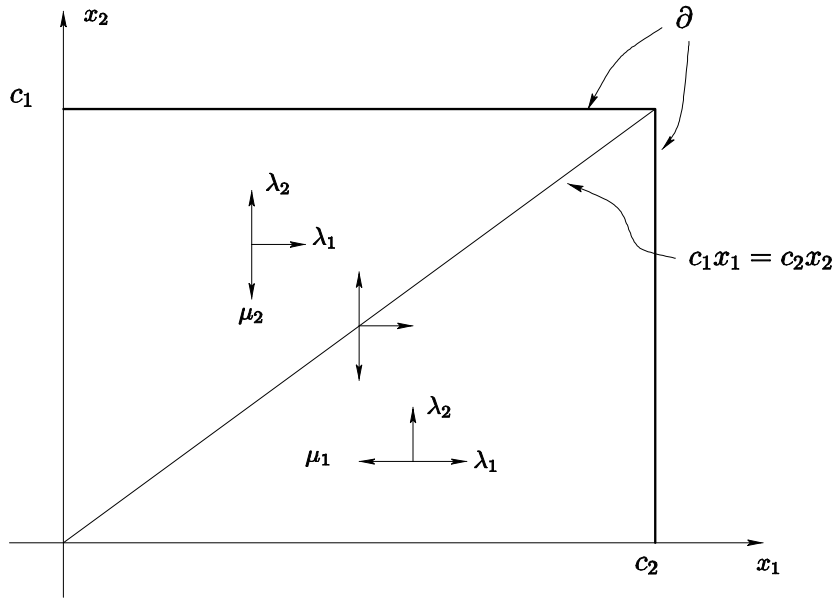
Example 1: Weighted serve-the-longer policy (wireless)





Example 1: Weighted serve-the-longer policy (wireless)





$$p_n = P\{Q_i \text{ exceeds } c_i n \text{ some } i = 1,2 \text{ before } Q = (0,0) | Q(0) = (1,0)\}.$$

Standard large deviation scaling:

$$X^n(t) = \frac{1}{n} Q(nt)$$

$$p_n = P\{X_i^n \text{ exceeds } c_i \text{ some } i = 1,2 \text{ before } X^n = (0,0) | X^n(0) = (1/n, 0)\}.$$

Example 2: Chemical reaction network and metastability

Rates for molecule types to react:

$$A \xrightarrow{r_1} B, \quad B \xrightarrow{r_2} A, \quad 2B + A \xrightarrow{r_3} 3B.$$

Suppose n molecules. If $(C_A^n(t), C_B^n(t)) =$ (fraction type A, fraction type B) at t, then $C_B^n(t) = 1 - C_A^n(t)$,

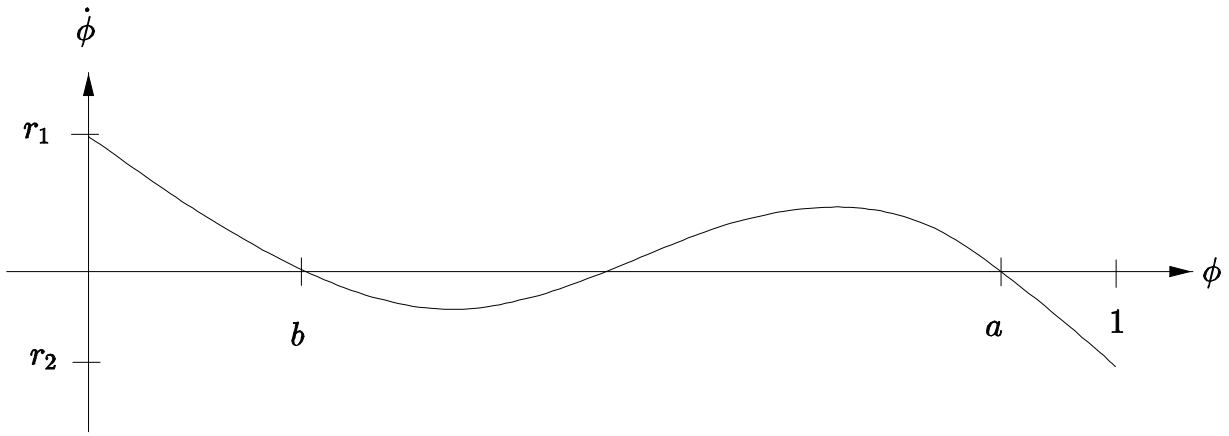
$$\begin{aligned} C_A^n(t) &\to C_A^n(t) - \frac{1}{n} \text{ at rate } n\left[r_1 C_A^n(t) + r_3 C_A^n(t) C_B^n(t)^2\right] \\ C_A^n(t) &\to C_A^n(t) + \frac{1}{n} \text{ at rate } n\left[r_2 C_B^n(t)\right]. \end{aligned}$$

Example 2: Chemical reaction network and metastability

LLN limit for C_A^n as $n \to \infty$:

$$\dot{\phi} = -r_1\phi + r_2(1-\phi) - r_3\phi(1-\phi)^2$$

with two stable equilibria when $r_3 > 3(r_1 + r_2)$:



$$p_n = P\{C_A^n(T) \text{ near stable point } a | C_A^n(0) \text{ near stable point } b\}$$

and reverse rate.

Example 3: Not-so-rare but high cost per sample–SPDE

Spread of pollutant, with concentration $u^n(x,t), x \in D \subset \mathbb{R}^d, t \in [0,T]$,

$$u_t^n(x,t) = cu_{xx}^n(x,t) + \langle v(x), u_x^n(x,t)\rangle - \alpha u^n(x,t) + \frac{1}{n}\cdot N(dx,dt)$$

with boundary and initial conditions and $N(dx,dt)$ spatial-temporal Poisson noise. Quantity of interest

$$p_n = P\left\{u^n(x_0,T) \geq u^*\right\}.$$

However, issue with sampling is high cost and p_n small but not exceedingly so.

Example 3: Not-so-rare but high cost per sample–SPDE

Spread of pollutant, with concentration $u^n(x,t), x \in D \subset \mathbb{R}^d, t \in [0,T]$,

$$u_t^n(x,t) = cu_{xx}^n(x,t) + \langle v(x), u_x^n(x,t)\rangle - \alpha u^n(x,t) + \frac{1}{n}\cdot N(dx,dt)$$

with boundary and initial conditions and $N(dx,dt)$ spatial-temporal Poisson noise. Quantity of interest

$$p_n = P\left\{u^n(x_0,T) \geq u^*\right\}.$$

However, issue with sampling is high cost and p_n small but not exceedingly so. Here a *moderate deviation approximation* may be more useful. Similar issues with other systems with high sampling cost (e.g., mean field models).

Canonical model and quantities to be estimated

As a general discrete time Markov model consider iid random vector fields $\{v_i(x), x \in \mathbb{R}^d\}$, with

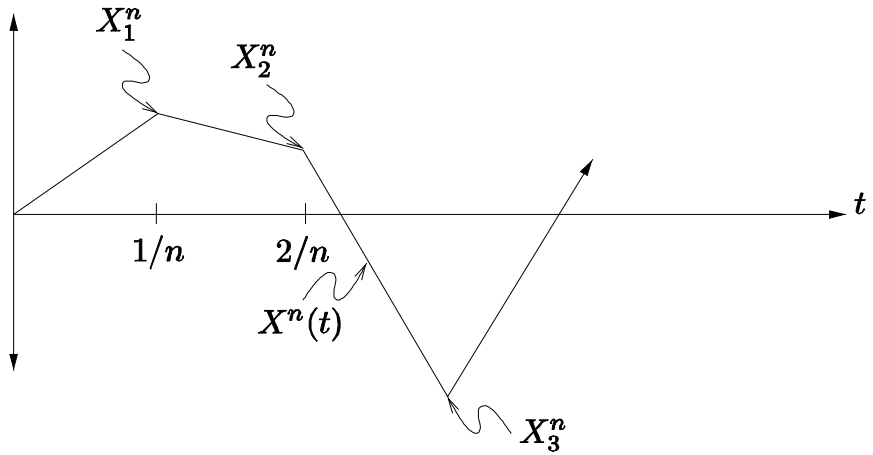
$$P\{v_i(x) \in A\} = \theta(A|x)$$

and the process

$$X_{i+1}^n = X_i^n + \frac{1}{n} v_i(X_i^n), \quad X_0^n = x.$$

Continuous time interpolation:

$$X^n(i/n) = X_i^n, \quad \text{piecewise linear interpolation for } t \neq i/n.$$



Canonical model and quantities to be estimated

Continuous time models:

- diffusion processes such as

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

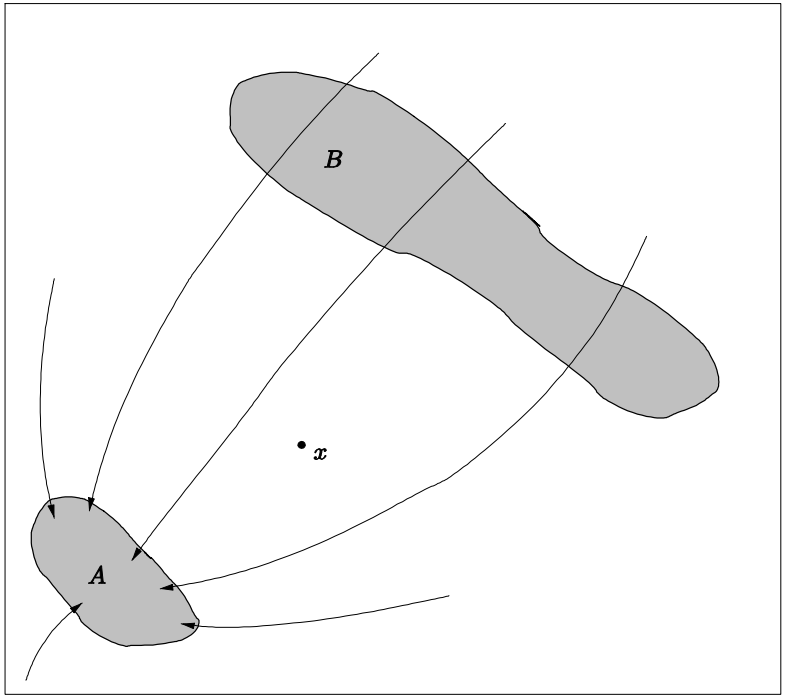
  corresponds to canonical model with

$$\theta(\cdot|x) = N(b(x), \sigma(x)\sigma^{T}(x)),$$

  (i.e., Euler approximation) with $\varepsilon = 1/n$.
- continuous time pure jump (e.g., queueing model) do not need time discretization, have development entirely analogous to discrete time theory.

Canonical model and quantity to be estimated

Hitting probability: assume LLN trajectories attracted to a point in A



and estimate

$$p_n(x) = P\{X^n \text{ hits } B \text{ before } A | X^n(0) = x\}$$

for $x \in (A \cup B)^c$. Essentially a finite time problem.

Monte Carlo and rare event considerations

Define

$$H(y,\alpha) = \log E \exp\langle \alpha, v_i(y)\rangle\,, \quad L(y,\beta) = \sup_{\alpha\in\mathbb{R}^d} \left[\langle \alpha,\beta\rangle - H(y,\alpha)\right],$$

assume $H(y,\alpha) < \infty$ all $\alpha \in \mathbb{R}^d$.

Under conditions $\{X^n(\cdot)\}$ satisfies a Large Deviation Principle with rate function

$$I_T(\phi) = \int_0^T L(\phi,\dot{\phi})dt$$

if ϕ is AC and $\phi(0) = x$, and $I_T(\phi) = \infty$ else.

Monte Carlo and rare event considerations

Define

$$H(y,\alpha) = \log E \exp\langle \alpha, v_i(y)\rangle\,, \quad L(y,\beta) = \sup_{\alpha\in\mathbb{R}^d} \left[\langle \alpha,\beta\rangle - H(y,\alpha)\right],$$

assume $H(y,\alpha) < \infty$ all $\alpha \in \mathbb{R}^d$.

Under conditions $\{X^n(\cdot)\}$ satisfies a Large Deviation Principle with rate function

$$I_T(\phi) = \int_0^T L(\phi,\dot{\phi})dt$$

if ϕ is AC and $\phi(0) = x$, and $I_T(\phi) = \infty$ else. Heuristically, for $T < \infty$, given ϕ, small $\delta > 0$ and large n

$$P\left\{\sup_{0\le t\le T} \|X^n(t) - \phi(t)\| \le \delta\right\} \approx e^{-nI_T(\phi)}.$$

Monte Carlo and rare event considerations

Hitting probability:

$$-\frac{1}{n}\log p_n(x)$$
$$\to \inf\{I_T(\phi) : \phi(0) = x, \phi \text{ enters } B \text{ prior to } A \text{ before } T, T < \infty\}.$$

Let

$$\mathcal{T}_{B,A} = \{ \text{ trajectories that hit } B \text{ prior to } A \}$$

$$r(x) = \inf\{I_T(\phi) : \phi(0) = x, \phi \text{ enters } B \text{ prior to } A \text{ before } T, T < \infty\}$$

Monte Carlo and rare event considerations

- For standard Monte Carlo we average iid copies of $1_{\{X^n \in \mathcal{T}_{B,A}\}}$. One needs $K \approx e^{nr(x)}$ samples for bounded relative error [std dev/$p_n(x)$].

Monte Carlo and rare event considerations

- For standard Monte Carlo we average iid copies of $1_{\{X^n \in \mathcal{T}_{B,A}\}}$. One needs $K \approx e^{nr(x)}$ samples for bounded relative error [std dev/$p_n(x)$].
- Alternative approach: construct iid random variables $s_1^n, \ldots, s_K^n$ with $Es_1^n = p_n(x)$ and use the unbiased estimator

$$\hat{q}_{n,K}(x) \doteq \frac{s_1^n + \cdots + s_K^n}{K}.$$

Monte Carlo and rare event considerations

- For standard Monte Carlo we average iid copies of $1_{\{X^n \in \mathcal{T}_{B,A}\}}$. One needs $K \approx e^{nr(x)}$ samples for bounded relative error [std dev/$p_n(x)$].
- Alternative approach: construct iid random variables $s_1^n, \ldots, s_K^n$ with $Es_1^n = p_n(x)$ and use the unbiased estimator

$$\hat{q}_{n,K}(x) \doteq \frac{s_1^n + \cdots + s_K^n}{K}.$$

- Performance determined by variance of s_1^n, and since unbiased by $E\,(s_1^n)^2$.

Monte Carlo and rare event considerations

- For standard Monte Carlo we average iid copies of $1_{\{X^n \in \mathcal{T}_{B,A}\}}$. One needs $K \approx e^{nr(x)}$ samples for bounded relative error [std dev/$p_n(x)$].
- Alternative approach: construct iid random variables $s_1^n, \ldots, s_K^n$ with $Es_1^n = p_n(x)$ and use the unbiased estimator

$$\hat{q}_{n,K}(x) \doteq \frac{s_1^n + \cdots + s_K^n}{K}.$$

- Performance determined by variance of s_1^n, and since unbiased by $E\,(s_1^n)^2$.
- By Jensen's inequality

$$-\frac{1}{n} \log E\,(s_1^n)^2 \leq -\frac{2}{n} \log Es_1^n = -\frac{2}{n} \log p_n(x) \to 2r(x).$$

Monte Carlo and rare event considerations

- For standard Monte Carlo we average iid copies of $1_{\{X^n \in \mathcal{T}_{B,A}\}}$. One needs $K \approx e^{nr(x)}$ samples for bounded relative error [std dev/$p_n(x)$].
- Alternative approach: construct iid random variables $s_1^n, \ldots, s_K^n$ with $Es_1^n = p_n(x)$ and use the unbiased estimator

$$\hat{q}_{n,K}(x) \doteq \frac{s_1^n + \cdots + s_K^n}{K}.$$

- Performance determined by variance of s_1^n, and since unbiased by $E\,(s_1^n)^2$.
- By Jensen's inequality

$$-\frac{1}{n} \log E\,(s_1^n)^2 \leq -\frac{2}{n} \log Es_1^n = -\frac{2}{n} \log p_n(x) \to 2r(x).$$

- An estimator is called *asymptotically efficient* if

$$\liminf_{n\to\infty} -\frac{1}{n} \log E\,(s_1^n)^2 \geq \ 2r(x).$$

Two methods: Importance sampling and splitting

- Two main methods (to date) for construction of random variables s_i^n with $Es_i^n = p_n(x)$: *importance sampling* (IS) and *splitting*.

Two methods: Importance sampling and splitting

- Two main methods (to date) for construction of random variables s_i^n with $Es_i^n = p_n(x)$: *importance sampling* (IS) and *splitting*.
- **Idea of importance sampling.** Simulate under a different distribution for which the event is not rare, correct using likelihood ratio to make unbiased.

Two methods: Importance sampling and splitting

- Two main methods (to date) for construction of random variables s_i^n with $Es_i^n = p_n(x)$: *importance sampling* (IS) and *splitting*.
- **Idea of importance sampling.** Simulate under a different distribution for which the event is not rare, correct using likelihood ratio to make unbiased.
- **Idea of splitting.** Many variations. Simplest is to trigger splits in such a way that rare event is encouraged. Divide unit mass of original particle among descendants to maintain unbiasedness.

Importance sampling

Tilted distributions. Recall

$$X_{i+1}^n = X_i^n + \frac{1}{n} v_i(X_i^n), \quad X_0^n = x$$

and $P\{v_i(y) \in dz\} = \theta(dz|y)$. Consider the *exponential tilt* with *tilt parameter* α and

$$\theta^\alpha(dz|y) = e^{\langle \alpha, y \rangle - H(y,\alpha)} \theta(dz|y).$$

Importance sampling

Tilted distributions. Recall

$$X_{i+1}^n = X_i^n + \frac{1}{n} v_i(X_i^n), \quad X_0^n = x$$

and $P\{v_i(y) \in dz\} = \theta(dz|y)$. Consider the *exponential tilt* with *tilt parameter* α and

$$\theta^\alpha(dz|y) = e^{\langle \alpha, y\rangle - H(y,\alpha)} \theta(dz|y).$$

Construct $\bar{X}_i^n$ recursively by setting

$$\bar{X}_{i+1}^n = \bar{X}_i^n + \frac{1}{n} \bar{Z}_i^n, \quad \bar{X}_0^n = x,$$

where $P\left\{\bar{Z}_i^n \in dz | \text{ data till time } i\right\} = \theta^{\bar{\alpha}_i^n}(dz|\bar{X}_i^n), \bar{\alpha}_i^n = F_i^n(\bar{X}_0^n, \ldots, \bar{X}_i^n)$.

Importance sampling

Tilted distributions. Recall

$$X_{i+1}^n = X_i^n + \frac{1}{n} v_i(X_i^n), \quad X_0^n = x$$

and $P\{v_i(y) \in dz\} = \theta(dz|y)$. Consider the *exponential tilt* with *tilt parameter* α and

$$\theta^\alpha(dz|y) = e^{\langle \alpha, y\rangle - H(y,\alpha)} \theta(dz|y).$$

Construct $\bar{X}_i^n$ recursively by setting

$$\bar{X}_{i+1}^n = \bar{X}_i^n + \frac{1}{n} \bar{Z}_i^n, \quad \bar{X}_0^n = x,$$

where $P\left\{\bar{Z}_i^n \in dz | \text{ data till time } i\right\} = \theta^{\bar{\alpha}_i^n}(dz|\bar{X}_i^n), \bar{\alpha}_i^n = F_i^n(\bar{X}_0^n, \ldots, \bar{X}_i^n)$. Likelihood ratio up to time n, new distribution with respect to old:

$$\prod_{i=0}^{n-1} e^{\left\langle \bar{\alpha}_i^n, \bar{Z}_i^n \right\rangle - H(\bar{X}_i^n, \bar{\alpha}_i^n)}$$

Importance sampling

Let $\bar{N}^n = \min\left\{i : \bar{X}_i^n \in A \cup B\right\}$ and set

$$s^n = 1_{\left\{\bar{X}^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{\bar{N}^n - 1} e^{-\left\langle \bar{\alpha}_i^n, \bar{Z}_i^n \right\rangle + H(\bar{X}_i^n, \bar{\alpha}_i^n)}.$$

Importance sampling

Let $\bar{N}^n = \min\left\{i : \bar{X}_i^n \in A \cup B\right\}$ and set

$$s^n = 1_{\left\{\bar{X}^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{\bar{N}^n - 1} e^{-\left\langle \bar{\alpha}_i^n, \bar{Z}_i^n \right\rangle + H(\bar{X}_i^n, \bar{\alpha}_i^n)}.$$

Recall that performance is measured by

$$E\left(s^n\right)^2 = E_x\left[1_{\left\{\bar{X}^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{\bar{N}^n - 1} e^{-2\left\langle \bar{\alpha}_i^n, \bar{Z}_i^n \right\rangle + 2H(\bar{X}_i^n, \bar{\alpha}_i^n)}\right],$$

Importance sampling

Let $\bar{N}^n = \min\left\{i : \bar{X}_i^n \in A \cup B\right\}$ and set

$$s^n = 1_{\left\{\bar{X}^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{\bar{N}^n-1} e^{-\left\langle \bar{\alpha}_i^n, \bar{Z}_i^n \right\rangle + H(\bar{X}_i^n, \bar{\alpha}_i^n)}.$$

Recall that performance is measured by

$$E\left(s^n\right)^2 = E_x\left[1_{\left\{\bar{X}^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{\bar{N}^n-1} e^{-2\left\langle \bar{\alpha}_i^n, \bar{Z}_i^n \right\rangle + 2H(\bar{X}_i^n, \bar{\alpha}_i^n)}\right],$$

which in terms of *original* random variables with $\alpha_i^n = F_i^n(X_0^n, \ldots, X_{i-1}^n)$ is

$$E\left(s^n\right)^2 = E_x\left[1_{\left\{X^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{N^n-1} e^{-\left\langle \alpha_i^n, v_i(X_i^n) \right\rangle + H(X_i^n, \alpha_i^n)}\right].$$

Importance sampling

Let $\bar{N}^n = \min\left\{i : \bar{X}_i^n \in A \cup B\right\}$ and set

$$s^n = 1_{\left\{\bar{X}^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{\bar{N}^n - 1} e^{-\left\langle \bar{\alpha}_i^n, \bar{Z}_i^n \right\rangle + H(\bar{X}_i^n, \bar{\alpha}_i^n)}.$$

Recall that performance is measured by

$$E\left(s^n\right)^2 = E_x\left[1_{\left\{\bar{X}^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{\bar{N}^n - 1} e^{-2\left\langle \bar{\alpha}_i^n, \bar{Z}_i^n \right\rangle + 2H(\bar{X}_i^n, \bar{\alpha}_i^n)}\right],$$

which in terms of *original* random variables with $\alpha_i^n = F_i^n(X_0^n, \ldots, X_{i-1}^n)$ is

$$E\left(s^n\right)^2 = E_x\left[1_{\left\{X^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{N^n - 1} e^{-\left\langle \alpha_i^n, v_i(X_i^n) \right\rangle + H(X_i^n, \alpha_i^n)}\right].$$

Potential problem: the exponential in the last display is dangerous.

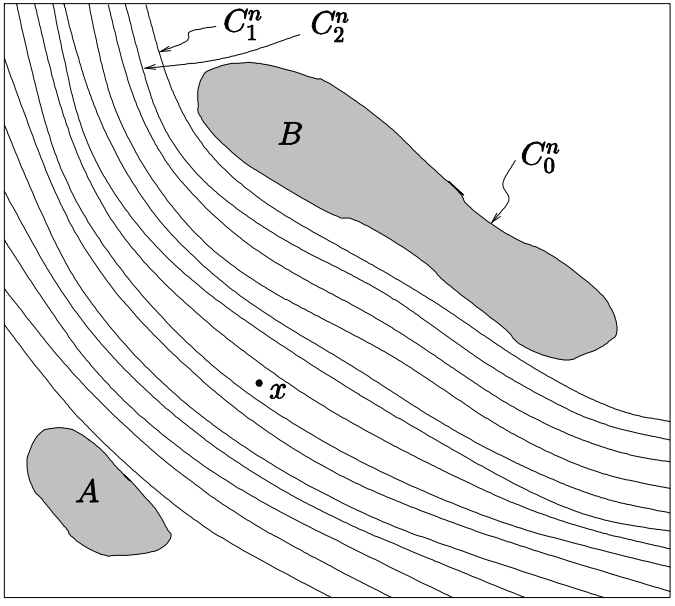
Splitting

Start a particle at initial condition of interest, then branch in a way that encourages outcome of rare event.
Key issues:

- When to branch?
- How much to branch?
- How to form an unbiased estimate?

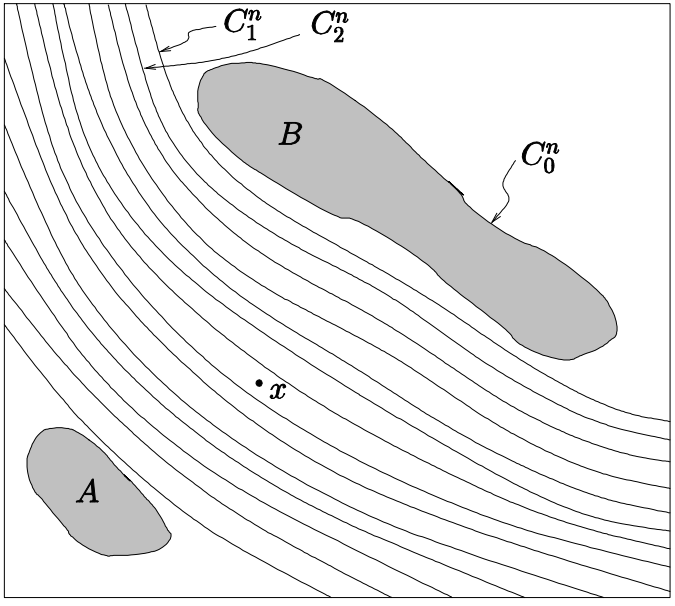
Splitting

A certain number [proportional to n] of *splitting thresholds* C_j^n are defined which enhance migration, e.g.,



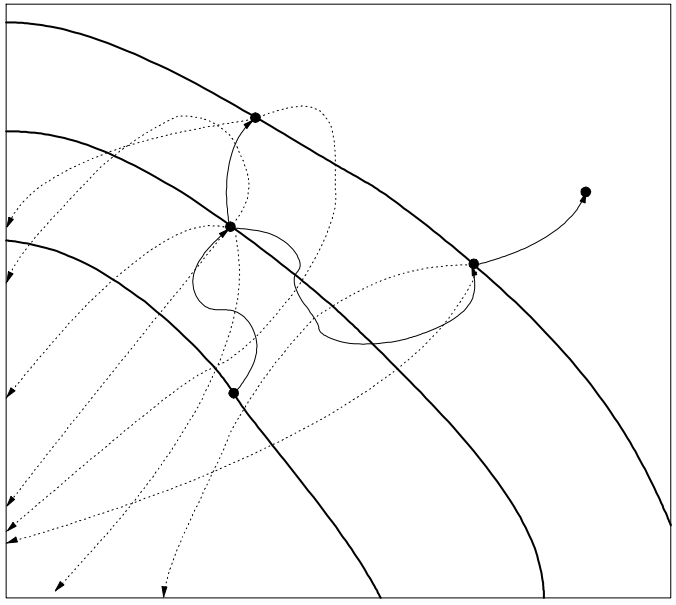
Splitting

A certain number [proportional to n] of *splitting thresholds* C_j^n are defined which enhance migration, e.g.,

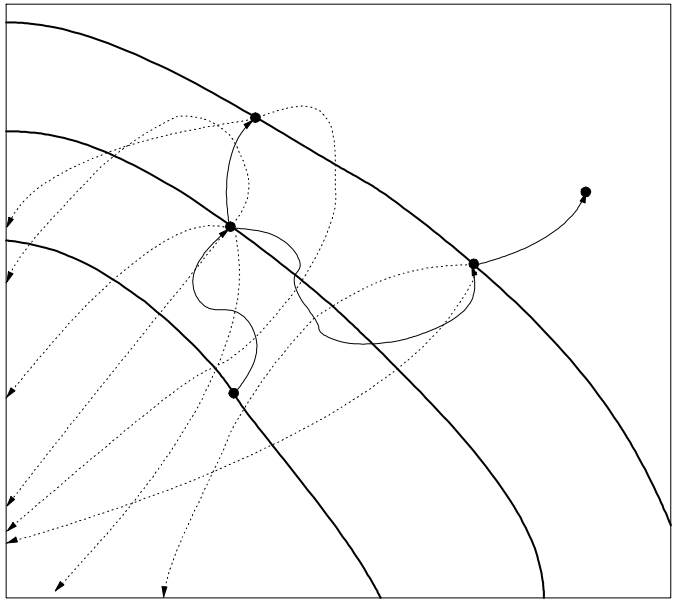


A single particle is started at x that follows the same law as X^n, but branches into a number of independent copies each time a new level is reached.

Splitting



Splitting



For simplicity assume the number of new particles M is deterministic. At each branching a multiplicative weight $1/M$ is assigned to each descendent.

Splitting

Evolution continues until every particle has reached either A or B. Let

$$\begin{aligned} R_x^n &= \text{total number of particles generated} \\ X_j^n(t) &= \text{trajectory of } j\text{th particle,} \\ W^n &= \text{product of weights assigned to } j \text{ along path} \end{aligned}$$

Then

$$s^n = \sum_{j=1}^{R_x^n} 1_{\{X_j^n \in \mathcal{T}_{B,A}\}} W^n.$$

Splitting

Evolution continues until every particle has reached either A or B. Let

$$\begin{aligned} R_x^n &= \text{total number of particles generated} \\ X_j^n(t) &= \text{trajectory of } j\text{th particle,} \\ W^n &= \text{product of weights assigned to } j \text{ along path} \end{aligned}$$

Then

$$s^n = \sum_{j=1}^{R_x^n} 1_{\{X_j^n \in \mathcal{T}_{B,A}\}} W^n.$$

If Kn thresholds, then

$$s^n = W^n \cdot (\#\text{ of particles reaching } B \text{ before } A) = \frac{1}{M^{Kn}} \sum_{j=1}^{R_x^n} 1_{\{X_j^n \in \mathcal{T}_{B,A}\}}.$$

Splitting

Evolution continues until every particle has reached either A or B. Let

$$\begin{aligned} R_x^n &= \text{total number of particles generated} \\ X_j^n(t) &= \text{trajectory of } j\text{th particle,} \\ W^n &= \text{product of weights assigned to } j \text{ along path} \end{aligned}$$

Then

$$s^n = \sum_{j=1}^{R_x^n} 1_{\{X_j^n \in \mathcal{T}_{B,A}\}} W^n.$$

If Kn thresholds, then

$$s^n = W^n \cdot (\#\text{ of particles reaching } B \text{ before } A) = \frac{1}{M^{Kn}} \sum_{j=1}^{R_x^n} 1_{\{X_j^n \in \mathcal{T}_{B,A}\}}.$$

How to choose thresholds C_r^n, M, and weights?

Remarks

- First use of IS in rare event context was Seigmund (1976).

Remarks

- First use of IS in rare event context was Seigmund (1976).
- *Fixed rate* splitting schemes begin with Kahn and Harris (1951), further developed in Booth and Hendricks (1984). Schemes have different names in different communities (e.g., *forward flux* in chemistry).

Remarks

- First use of IS in rare event context was Seigmund (1976).
- *Fixed rate* splitting schemes begin with Kahn and Harris (1951), further developed in Booth and Hendricks (1984). Schemes have different names in different communities (e.g., *forward flux* in chemistry).
- Some obvious inefficiencies lead to the RESTART variant Villén-Altamirano and Villén-Altamirano (1994), which allows particles of little likely use to be terminated *without introducing bias*. Harder to analyze but leads to same theoretical bounds and preferable for several reasons.

Importance Functions

Natural conceptual framework for design is through *importance functions*. Let $V : \mathbb{R}^d \to \mathbb{R}$ be continuous and satisfy

$$V(x) \leq 0 \text{ for } x \in B.$$

*The first examples showing that state feedback was essential for even good (better than naive MC) performance are in Ref 5.

Importance Functions

Natural conceptual framework for design is through *importance functions*. Let $V : \mathbb{R}^d \to \mathbb{R}$ be continuous and satisfy

$$V(x) \leq 0 \text{ for } x \in B.$$

Design of scheme:

- for importance sampling, and assuming V continuously differentiable, the change of measure if the simulated trajectory is at $\bar{X}_i^n$ is*

$$\theta^{-DV(\bar{X}_i^n)}(dz|y) = e^{\left\langle -DV(\bar{X}_i^n), z\right\rangle - H(y,-DV(\bar{X}_i^n))}\theta(dz|y).$$

- for splitting, assuming V is continuous, thresholds are defined by

$$C_i^n = \{y : V(y) \leq i\,(\log M)\,/n\}\,,$$

  if M descendents at each time of splitting.

*The first examples showing that state feedback was essential for even good (better than naive MC) performance are in Ref 5.

Some heuristics

Recall

$$
\begin{aligned}
r(x) &= \lim_{n\to\infty} -\frac{1}{n}\log p_n(x) \\
&= \inf\left\{\int_0^T L(\phi(s),\dot{\phi}(s))ds : \phi(0)=x, \phi \text{ hits } B \text{ before } A,\ T<\infty\right\}.
\end{aligned}
$$

Some heuristics

Recall

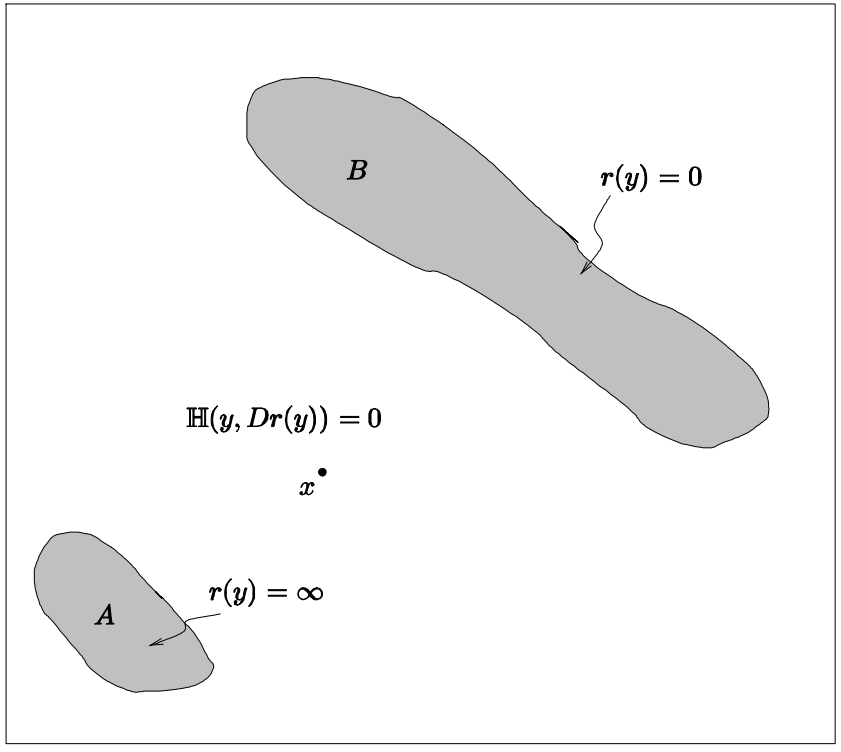
$$\begin{aligned} r(x) &= \lim_{n\to\infty} -\frac{1}{n}\log p_n(x) \\ &= \inf\left\{\int_0^T L(\phi(s),\dot{\phi}(s))ds : \phi(0)=x, \phi \text{ hits } B \text{ before } A,\ T<\infty\right\}. \end{aligned}$$

Generalize to *arbitrary* starting position y. Then $r(y)$ satisfies a dynamic programming relation: for $\Delta > 0$

$$r(y) = \inf\left\{\int_0^\Delta L(\phi(s),\dot{\phi}(s))ds + r(y+\phi(\Delta)) : \phi(0)=y\right\}.$$

Some heuristics

This implies $r(y)$ is a weak sense (viscosity) solution to



where $\mathbb{H}(y, \alpha) = -H(y, -\alpha)$.

Some heuristics

If we could use r as the importance function, problems with splitting, importance sampling essentially solved!

Splitting: With thresholds defined by r,

$$\begin{aligned} P\{\text{particle born at threshold } j \text{ reaches } j-1\} &\approx e^{-n \inf I_\Delta(\phi)} \\ &\approx e^{-n\left[r(C_j^n)-r(C_{j-1}^n)\right]} \\ &= e^{-n[\log M/n]} \\ &= \frac{1}{M}, \end{aligned}$$

where inf is over paths connecting any point in C_j^n to C_{j-1}^n. Critical (borderline) growth, and weights $M^{-Kn} = M^{-n[r(x)/(\log M)]} = e^{-nr(x)}$.

Some heuristics

Importance sampling: With tilts defined by

$$\theta^{-Dr(\bar{X}_i^n)}(dz|y) = e^{\left\langle -Dr(\bar{X}_i^n), z\right\rangle - H(y,-Dr(\bar{X}_i^n))}\theta(dz|y),$$

likelihood ratio along *any trajectory* satisfies

$$\begin{aligned}
&1_{\{\bar{X}^n_{\bar{N}^n}\in B\}} \prod_{i=0}^{\bar{N}^n-1} e^{\left\langle Dr(\bar{X}_i^n), \bar{Z}_i^n\right\rangle + H(\bar{X}_i^n, -Dr(\bar{X}_i^n))} \\
&\quad = 1_{\{\bar{X}^n_{\bar{N}^n}\in B\}} \prod_{i=0}^{\bar{N}^n-1} e^{n\left\langle Dr(\bar{X}_i^n), [\bar{X}_{i+1}^n - \bar{X}_i^n]\right\rangle - \mathbb{H}(\bar{X}_i^n, Dr(\bar{X}_i^n))} \\
&\quad = 1_{\{\bar{X}^n_{\bar{N}^n}\in B\}} e^{n\sum_{i=1}^{\bar{N}^n-1}\left\langle Dr(\bar{X}_i^n), [\bar{X}_{i+1}^n - \bar{X}_i^n]\right\rangle} \\
&\quad \approx 1_{\{\bar{X}^n_{\bar{N}^n}\in B\}} e^{nr(\bar{X}^n_{\bar{N}^n}) - nr(x)} = 1_{\{\bar{X}^n_{\bar{N}^n}\in B\}} e^{-nr(x)}.
\end{aligned}$$

Some heuristics

While possible in special cases, r is typically not available. However, it turns out that the critical properties of an importance function V are

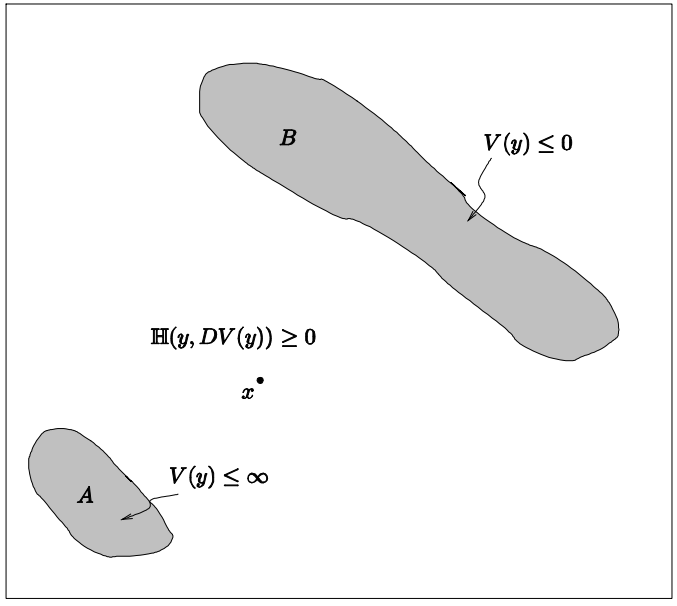
- it should be a *subsolution* to the PDE,
- its value at the starting point,
- it should be C^1 for important sampling and a viscosity subsolution for splitting.

Why Importance Functions should be subsolutions

Let

$$\mathbb{H}(y, \alpha) = -H(y, -\alpha).$$

A *classical sense* subsolution V for the escape probability is smooth (C^1) and satisfies

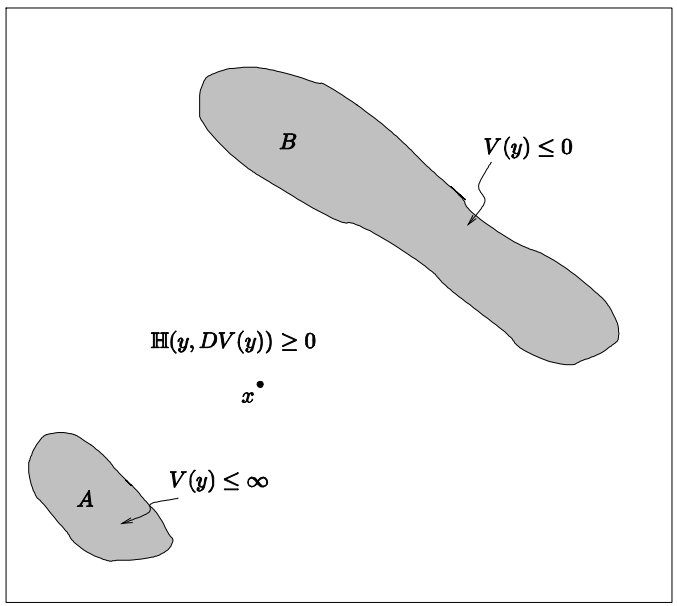


Why Importance Functions should be subsolutions

Let

$$\mathbb{H}(y, \alpha) = -H(y, -\alpha).$$

A *classical sense* subsolution V for the escape probability is smooth (C^1) and satisfies



A *viscosity sense* subsolution V satisfies the boundary inequalities and $\mathbb{H}(y, p) \ge 0$ for any superdifferential p of V at $y \in (A \cup B)^c$ (need not be smooth). Analogous definitions for other problems (e.g., finite time probabilities, functionals).

Why Importance Functions should be subsolutions

A constraint on importance functions. If V is used as an importance function and V is not a subsolution, bad (exponentially bad) things happen.

- For importance sampling we have expression for second moment when V used:

$$E\left(s^n\right)^2 = E_x\left[1_{\left\{X^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{N^n-1} e^{\left\langle DV(X_i^n), v_i(X_i^n)\right\rangle + H(X_i^n, -DV(X_i^n))}\right].$$

  In regions where $H(x, -DV(x)) > 0$, large deviation analysis shows decay rate strictly smaller than twice that of $p_n(x)$. Asymptotic efficiency impossible.

Why Importance Functions should be subsolutions

A constraint on importance functions. If V is used as an importance function and V is not a subsolution, bad (exponentially bad) things happen.

- For importance sampling we have expression for second moment when V used:

$$E\left(s^n\right)^2 = E_x\left[1_{\left\{X^n \in \mathcal{T}_{B,A}\right\}} \prod_{i=0}^{N^n-1} e^{\left\langle DV(X_i^n), v_i(X_i^n)\right\rangle + H(X_i^n, -DV(X_i^n))}\right].$$

  In regions where $H(x, -DV(x)) > 0$, large deviation analysis shows decay rate strictly smaller than twice that of $p_n(x)$. Asymptotic efficiency impossible.

- For splitting in regions where $H(x, -DV(x)) > 0$ the mean number of particles reaching next threshold strictly larger than 1, exponentially many particles, and a "work-normalized" asymptotic efficiency impossible.

Performance for schemes based on subsolutions

Theorem

Let s^n be the splitting estimate for the escape probability $p_n(x)$ based on V using either standard splitting or RESTART. Then the number of particles generated grows subexponentially in n if and only if V is a viscosity subsolution, in which case we also have

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\left(s^n\right)^2 \geq \ V(x) + r(x).$$

Under additional regularity $\liminf$ *and* $\geq$ *become* $\lim$ *and* $=$.

Performance for schemes based on subsolutions

Theorem

Let s^n be the splitting estimate for the escape probability $p_n(x)$ based on V using either standard splitting or RESTART. Then the number of particles generated grows subexponentially in n if and only if V is a viscosity subsolution, in which case we also have

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\left(s^n\right)^2 \geq \ V(x) + r(x).$$

Under additional regularity $\liminf$ *and* $\geq$ *become* $\lim$ *and* $=$.

With a *strict* subsolution, one can bound the mean number of particles uniformly in n.

Performance for schemes based on subsolutions

Theorem

Let V be a classical subsolution and s^n be the importance sampling estimate for the escape probability $p_n(x)$ based on V. Then

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\,(s^n)^2 \geq \; V(x) + r(x).$$

Under additional regularity $\liminf$ *becomes* $\lim$ *with a RHS bounded below by* $V(x) + r(x)$.

Performance for schemes based on subsolutions

Theorem

Let V be a classical subsolution and s^n be the importance sampling estimate for the escape probability $p_n(x)$ based on V. Then

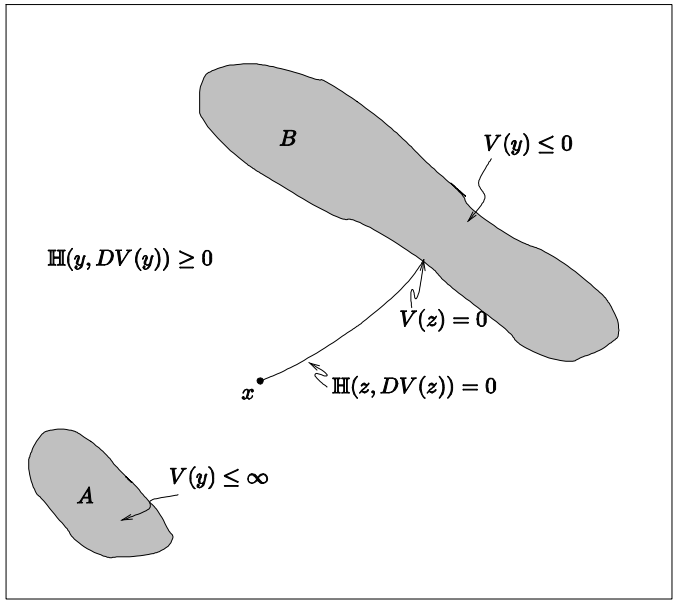
$$\liminf_{n\to\infty} -\frac{1}{n}\log E\,(s^n)^2 \geq \; V(x) + r(x).$$

Under additional regularity lim inf *becomes* lim *with a RHS bounded below by $V(x) + r(x)$.*

Proofs of these and other results found in Springer book.

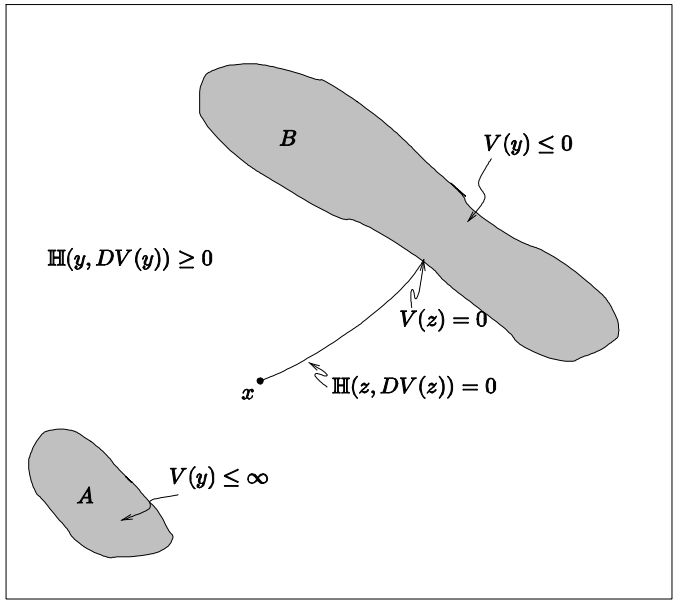
Performance for schemes based on subsolutions

Requirements for asymptotic optimality. Asymptotic optimality means $V(x) = r(x)$ (note $V(x) \leq r(x)$ automatic). Equation holds with equality along conditional most likely path starting at x:



Performance for schemes based on subsolutions

Requirements for asymptotic optimality. Asymptotic optimality means $V(x) = r(x)$ (note $V(x) \leq r(x)$ automatic). Equation holds with equality along conditional most likely path starting at x:



Analogous results for finite time problems, expected values, and other (but not all) problem formulations.

Construction of subsolutions

Various methods depending on structure of problem (see the references):

Construction of subsolutions

Various methods depending on structure of problem (see the references):

- For systems with $\mathbb{H}(y, \alpha)$ piecewise constant structure (e.g., queueing networks)–a direct construction based on pointwise minimum of affine functions using critical roots of $\mathbb{H}(y, \alpha) = 0$.

Construction of subsolutions

Various methods depending on structure of problem (see the references):

- For systems with $\mathbb{H}(y, \alpha)$ piecewise constant structure (e.g., queueing networks)–a direct construction based on pointwise minimum of affine functions using critical roots of $\mathbb{H}(y, \alpha) = 0$.
- Construction as pointwise minimum of solutions to boundary/terminal conditions admitting explicit solutions (useful for occupancy and related combinatorial problems).

Construction of subsolutions

Various methods depending on structure of problem (see the references):

- For systems with $\mathbb{H}(y, \alpha)$ piecewise constant structure (e.g., queueing networks)–a direct construction based on pointwise minimum of affine functions using critical roots of $\mathbb{H}(y, \alpha) = 0$.
- Construction as pointwise minimum of solutions to boundary/terminal conditions admitting explicit solutions (useful for occupancy and related combinatorial problems).
- Construction in terms of solution to linear/quadratic/regulator (in particular useful for *moderate deviations* approximations).

Construction of subsolutions

Various methods depending on structure of problem (see the references):

- For systems with $\mathbb{H}(y, \alpha)$ piecewise constant structure (e.g., queueing networks)–a direct construction based on pointwise minimum of affine functions using critical roots of $\mathbb{H}(y, \alpha) = 0$.
- Construction as pointwise minimum of solutions to boundary/terminal conditions admitting explicit solutions (useful for occupancy and related combinatorial problems).
- Construction in terms of solution to linear/quadratic/regulator (in particular useful for *moderate deviations* approximations).
- For finite time problems one can be constructed in terms of Freidlin-Wentsell quasipotential. Available in explicit form for reversible or (asymptotically reversible). Useful when time interval is large for splitting, and for importance sampling if no "rest point".

Remarks on proofs

Importance sampling. Estimate

$$1_{\{\bar{X}^n_{\bar{N}^n} \in B\}} \prod_{i=0}^{\bar{N}^n - 1} e^{\left\langle DV(\bar{X}^n_i), \bar{Z}^n_i \right\rangle + H(\bar{X}^n_i, -DV(\bar{X}^n_i))}.$$

- Express second moment for estimator as an exponential integral with respect to original distributions
- Use same method as that of LD analysis to write a *stochastic control* representation for log of second moment
- Use classical verification argument to bound representation (hence decay of second moment)

Remarks on proofs

Splitting. Estimate based on splitting rate M and Kn thesholds

$$s^n = \frac{\#\text{ particles reach } B \text{ before } A}{M^{Kn}}, \quad E(s^n)^2 = E\left[\sum_{j=1}^{R_x^n} 1_{\{X_j^n \in \mathcal{T}_{B,A}\}} M^{-Kn}\right]^2.$$

- Partition expectation into contributions from pairs of particles with last common ancestor at threshold κ.
- Bound each such using large deviation bounds (one particle gets to C_κ^n, two independent particles go from C_κ^n to B), decay of M^{-Kn}, dynamic programming inequalities on r and V to get upper bound of form

$$e^{-n[r(x)+V(x)]}.$$

Summary–distinctions between importance sampling and splitting

Subsolutions allow one to characterize sufficient (and also necessary) conditions for performance measures for important sampling and splitting, but there are important differences.

- The differences are largely related to the regularity required from the subsolution [classical for importance sampling, viscosity for splitting], and how it is used to define the scheme [$D\bar{V}$ versus $\bar{V}$].
- Subsolutions are often naturally constructed as the pointwise min of smooth subsolutions. It turns out that if $V = \min_{k=1,\ldots K} V_k$ satisfies the boundary condition $V(x) \leq 0$ for $x \in B$ and if each V_k is a subsolution to the PDE, $\mathbb{H}(y, DV_k(y)) \geq 0$, then can use exponential mollification to produce a classical subsolution: for $\delta > 0$

$$V^\delta(x) = -\delta \log \left(\sum_{k=1}^{K} e^{-\frac{1}{\delta} V_k(x)} \right),$$

with only small change in value at starting point.

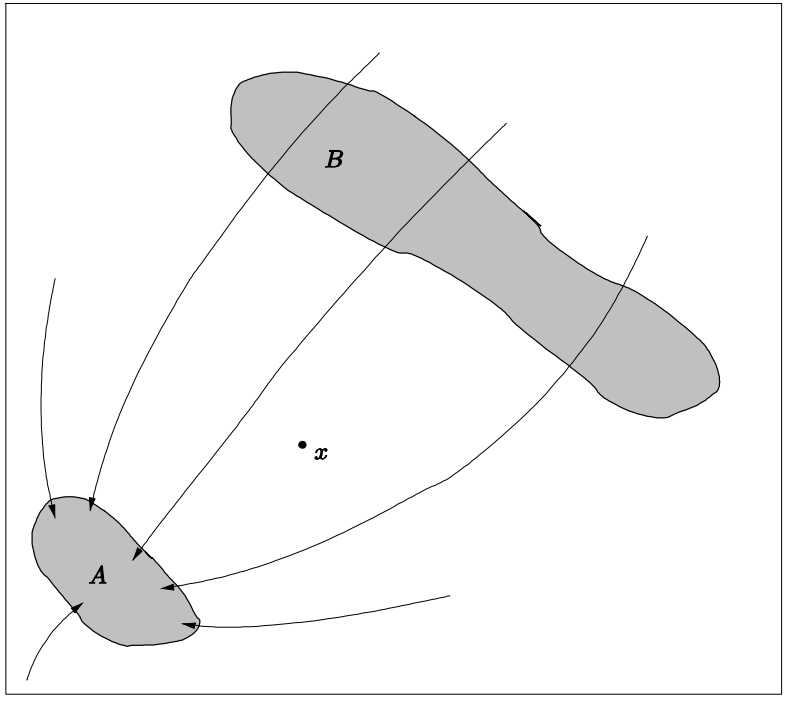
Summary–distinctions between importance sampling and splitting

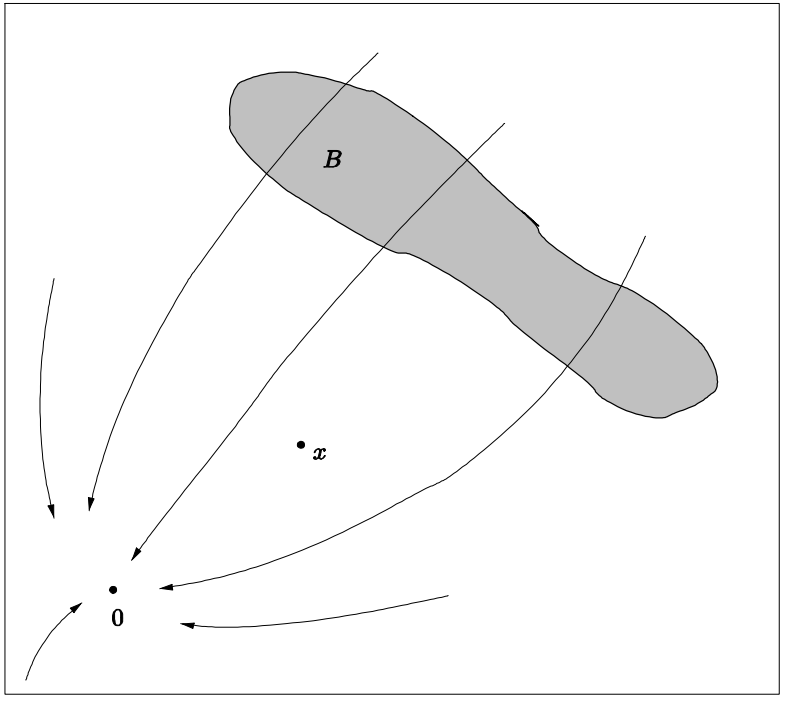
- When one goes to the trouble to make importance sampling work, generally (but not always) performs somewhat better than splitting.
- For Markov modulated noise (not covered in talk) and other multiscale problems IS requires solution to an eigenvalue for each tilt parameter used. Although splitting does not explicitly, often needed to evaluate Hamiltonian. Possible advantage to splitting.
- Under the condition $\sup_x E e^{\sigma \|v_i(x)\|^2} < \infty$ for some $\sigma > 0$, one can prove *non-asymptotic* bounds for importance sampling. Analogue not known for for splitting.

A situation where the two differ greatly

Although often comparable, recent results show can be truly significant differences. In particular, when

- **metastable point is in the domain of simulation.**





$P_x\{X^n \text{ hits } B \text{ before } A \text{ by } T\}$ vs $P_x\{X^n \text{ hits } B \text{ by } T\}$

A situation where the two differ greatly

Treatment in neighborhoods of metastable points difficult.

When T is large (e.g., transition rate calculations) the Freidlin-Wentsell quasipotential can be used to define a subsolution with nearly optimal value at starting point (optimal in limit $T \to \infty$). With

$$Q(x) = \inf \left\{ \int_0^T L(\phi(s), \dot{\phi}(s)) ds : \phi(0) = 0, \phi(T) = x, T < \infty \right\},$$

$Q + c$ is a viscosity subsolution for any c.

A situation where the two differ greatly

For problem such as $P\{X^n \text{ hits } B \text{ by } T(n) | X^n(0) = x\}$, with $T(n) \to \infty$ as $n \to \infty$, qualitative differences such as

$$\lim_{n\to\infty} -\frac{1}{n} \log E \left(s^n_{\text{splitting}}\right)^2 = 2\ V(0)$$

versus

$$\lim_{n\to\infty} -\frac{1}{n} \log E \left(s^n_{\text{IS}}\right)^2 = -\infty.$$

A situation where the two differ greatly

For problem such as $P\{X^n \text{ hits } B \text{ by } T(n) | X^n(0) = x\}$, with $T(n) \to \infty$ as $n \to \infty$, qualitative differences such as

$$\lim_{n\to\infty} -\frac{1}{n} \log E \left(s^n_{\text{splitting}}\right)^2 = 2\ V(0)$$

versus

$$\lim_{n\to\infty} -\frac{1}{n} \log E \left(s^n_{\text{IS}}\right)^2 = -\infty.$$

Source of variance in latter are paths under the change of measure which stay near 0 for long times.

References

Papers introducing methods

- Importance sampling in the Monte Carlo study of sequential tests (D. Siegmund), Ann. Statist., 4, (1976), 673–684.
- Estimation of particle transmission by random sampling (H. Kahn and T.E. Harris), National Bureau of Standards Applied Mathematics Series, 12, (1951), 27–30.
- Importance estimation in forward Monte Carlo calculations, (T. Booth and J. Hendricks), Nucl. Tech./Fusion, 6, (1984), 90–100.

Paper introducing RESTART

- RESTART: A method for accelerating rare event simulations, (M. Villen-Altamirano and J. Villen-Altamirano), Proc. of the 13th International Teletraffic Congress, Queueing, Performance and Control in ATM, (1991), 71–76.

Paper showing state feedback essential for importance sampling

- Counter examples in importance sampling for large deviations probabilities, (P. Glasserman and Y. Wang), Ann. Appl. Prob., 7, (1997), 731–746..

References

Papers developing subsolutions for importance sampling

- Importance sampling, large deviations and differential games (D. and H. Wang), Stochastics and Stochastics Reports, 76, (2004), 481–508.
- Subsolutions of an Isaacs equation and efficient schemes for importance sampling (D. and H. Wang), Math. of OR, 32, (2007), 1–35.
- Importance sampling for Jackson networks (D. and H. Wang), Queueing Systems, 62, (2009), 113–157.
- Large deviations and importance sampling for a tandem network with slow-down (D., K. Leder and H. Wang), QUESTA, 57, (2007), 71–83.

Papers developing subsolutions for splitting

- Splitting for rare event simulation: A large deviations approach to design and analysis (T. Dean and D.), SPA, 119, (2009), 562–587.
- The design and analysis of a generalized RESTART/DPR algorithm for rare event simulation (T. Dean and D.), Annals of OR, 189, (2011), 63–102.

References

Papers that analyse neighborhoods of metastable points

- Escaping from an attractor: Importance sampling and rest points I, (D., K. Spiliopoulos and X. Zhou), Annals of Applied Probability, 25, (2015), 2909–2958.
- Splitting algorithms for rare event simulation over long time intervals (A. Buijsrogge, D. and M. Snarski), preprint.