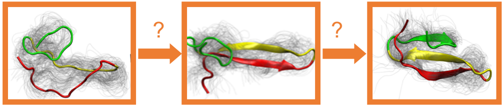
Importance sampling for rare events and some pathologies of the exit problem

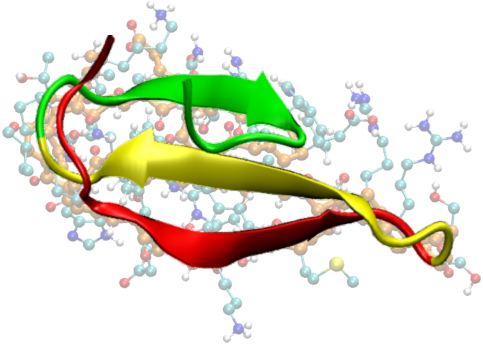
Carsten Hartmann (BTU Cottbus-Senftenberg)
Joint work with Lara Neureither (Cottbus), Omar Kebiri (Cottbus), and Lorenz Richter (Berlin)

Brin Mathematics Research Center, 27th Feb – 3rd Mar 2023

Motivation: WW domain of a protein

Protein folding





[Noé et al, PNAS, 2009]

Quantities involving random stopping times

Given a Markov process $X = (X_t)_{t \geq 0}$ in $\mathbb{R}^d$ and **first hitting times**

$$\tau_O = \inf\{t \geq 0 \colon X_t \in O\}\,, \quad O \in \{A, B, C\}$$

of some measurable subsets $A, B, C \subset \mathbb{R}^d$, we want to **estimate quantities**, such as

- committor probabilities $P(\tau_B < \tau_A)$
- transition probabilities $P(\tau_C \leq T)$
- moment generating functions $\mathbb{E}[\exp(-\alpha \tau_C)]$
- mean first passage times $\mathbb{E}[\tau_C]$.

Illustrative example I: bistable system

- Overdamped Langevin equation

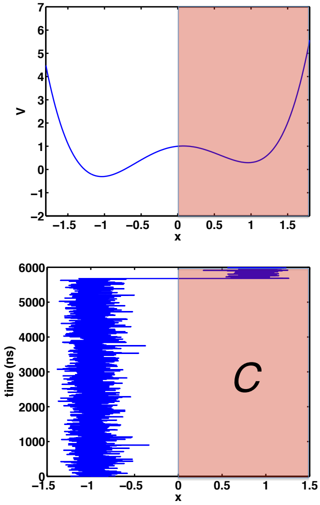
$$dX_t = -\nabla V(X_t)dt + \sqrt{2\epsilon}dB_t \,.$$

- Standard estimator of MGF $\psi = \psi_\epsilon$

$$\hat{\psi}_\epsilon^N = \frac{1}{N}\sum_{i=1}^{N} e^{-\alpha\tau_C^i} \,.$$

- Small noise asymptotics (Kramers)

$$\lim_{\epsilon\to 0} \epsilon \log \mathbb{E}[\tau_C] = \Delta V \,.$$



[Freidlin & Wentzell, 1984], [Berglund, Markov Processes Relat Fields 2013]

Illustrative example I, cont'd

- **Relative error** of the MC estimator

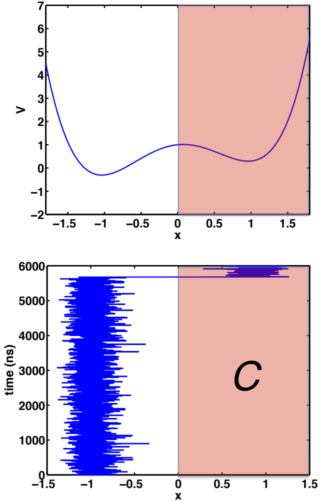
$$\delta_\epsilon = \frac{\sqrt{\mathrm{Var}[\hat{\psi}^\epsilon_N]}}{\mathbb{E}[\hat{\psi}^\epsilon_N]}$$

- Varadhan's large deviations principle

$$\mathbb{E}[(\hat{\psi}^N_\epsilon)^2] \gg (\mathbb{E}[\hat{\psi}^N_\epsilon])^2, \ \epsilon \text{ small.}$$

- Unbounded relative error as $\epsilon \to 0$

$$\limsup_{\epsilon \to 0} \delta_\epsilon = \infty$$



[Dupuis & Ellis, 1997]

Importance sampling

We may control the relative error by doing a **change of measure**, e.g.

$$\mathbb{E}\left[e^{-\alpha\tau_C}\right] = \mathbb{E}_Q\left[e^{-\alpha\tau_C}L^{-1}\right] = \mathbb{E}_Q\left[e^{-\alpha\tau_C - \log L}\right], \quad \alpha > 0$$

assuming that the likelihood ratio $L = \frac{dQ}{dP} > 0$ exists.

Key observations

1. **zero variance** change of measure from P to $Q = Q^*$ exists, with likelihood ratio

$$L^* = e^{c(\alpha) - \alpha\tau_C}, \quad c(\alpha) = -\log \mathbb{E}\left[e^{-\alpha\tau_C}\right].$$

2. variance reduction may not increase the **likelihood of the rare event**.

Importance sampling

We may control the relative error by doing a **change of measure**, e.g.

$$\mathbb{E}\left[e^{-\alpha\tau_C}\right] = \mathbb{E}_Q\left[e^{-\alpha\tau_C}L^{-1}\right] = \mathbb{E}_Q\left[e^{-\alpha\tau_C - \log L}\right], \quad \alpha > 0$$

assuming that the likelihood ratio $L = \frac{dQ}{dP} > 0$ exists.

Key observations

1. **zero variance** change of measure from P to $Q = Q^*$ exists, with likelihood ratio

$$L^* = e^{c(\alpha) - \alpha\tau_C}, \quad c(\alpha) = -\log \mathbb{E}\left[e^{-\alpha\tau_C}\right].$$

2. variance reduction may not increase the **likelihood of the rare event**.

Importance sampling

We may control the relative error by doing a **change of measure**, e.g.

$$\mathbb{E}\left[e^{-\alpha\tau_C}\right] = \mathbb{E}_Q\left[e^{-\alpha\tau_C}L^{-1}\right] = \mathbb{E}_Q\left[e^{-\alpha\tau_C - \log L}\right], \quad \alpha > 0$$

assuming that the likelihood ratio $L = \frac{dQ}{dP} > 0$ exists.

Key observations

1. **zero variance** change of measure from P to $Q = Q^*$ exists, with likelihood ratio

$$L^* = e^{c(\alpha)-\alpha\tau_C}, \quad c(\alpha) = -\log \mathbb{E}\left[e^{-\alpha\tau_C}\right].$$

2. variance reduction may not increase the **likelihood of the rare event**.

Controlling the variance in rare event simulation (non-exhaustive list)

Importance sampling (invasive)

- adaptive importance sampling based on optimal control techniques

  Glasserman & Wang; Dupuis & Wang; Vanden-Eijnden & Weare; H & Schütte; Spiliopoulos; Awad, Glynn & Rubinstein; ...

- KL divergence and cross-entropy minimisation

  Rubinstein & Kroese; Zhang & H; Kappen & Ruiz; Nüsken & Richter; ...

- Mean squared error and work-normalised variance minimisation

  Glynn & Whitt; Jourdain & Lelong; Su & Fu; Vázquez-Abad & Dufresne; ...

Splitting methods (non-invasive)

- RESTART, adaptive multilevel splitting

  Villén-Altamirano & Villén-Altamirano; Cérou & Guyader; Aristoff, Lelièvre, Mayne & Teo; ...

- checkpointing, milestoning, transition interface sampling

  Asmussen & Lipsky; Faradjian, Elber, West & Shalloway; Van Erp, Moroni & Bolhuis; Vanden-Eijnden & Venturoli; ...

- forward flux sampling, weighted ensemble method

  Allen, Valeriani and Ten Wolde; Huber & Kim; ...

Outline

A certainty-equivalence principle for importance sampling

Importance sampling of diffusions using tools from stochastic control theory

From importance sampling to control variates

- A certainty-equivalence principle for importance sampling
- Importance sampling of diffusions using tools from stochastic control theory
- From importance sampling to control variates

A wish list

Let $S(X) \geq 0$ be a non-negative functional of the process X that is of the form

$$S(X) = \int_0^\tau f(X_s)\, ds + g(X_\tau)\,,$$

for suitable functions $f, g \geq 0$ and an **a.s. finite stopping time** τ, e.g.

- $f = 0$, $g = \mathbf{1}_B$, $\tau = \min\{\tau_A, \tau_B\}$, such that $\mathbb{E}[S(X)] = P(\tau_B < \tau_A)$
- $f = 1$, $g = 0$, $\tau = \tau_C$, such that $\mathbb{E}[S(X)] = \mathbb{E}[\tau_C]$
- $f = 0$, $\tau = \min\{\tau_C, T\}$, such that $\mathbb{E}[S(X)] = P(\tau_C \leq T)$

Our aim is to find a change of measure from P to Q that both reduces the **variance** and the **average length of trajectories** (and from which we can draw samples).

Certainty-equivalence principle

Instead of $\mathbb{E}[S(X)]$, we consider the **certainty-equivalent expectation**

$$\gamma = \varphi^{-1}(\mathbb{E}[\varphi(S(X))])$$

where φ is a convex (strictly increasing or decreasing) function with inverse φ^{-1}.

Two notable special cases are

- $\varphi(s) = |s|^p$ for $p > 1$, with the property

$$\left(\mathbb{E}\left[(S(X))^p\right]\right)^{1/p} \geq \mathbb{E}\left[S(X)\right]$$

- $\varphi(s) = e^{-\alpha s}$ for $\alpha > 0$, with the property

$$-\alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right] \leq \mathbb{E}\left[S(X)\right].$$

In both cases **equality holds** iff S is a.s. constant

[Whittle, Macroecon Dyn, 2002], [Fleming & Soner, 2006]

Certainty equivalence principle, cont'd

If $\varphi(s) = |s|^p$ for $p \geq 1$, it holds

$$(\mathbb{E}[(S(X))^p])^{1/p} = \sup\left\{\mathbb{E}_Q\left[S(X)\left(\frac{dQ}{dP}\right)^{-1/p}\right] : Q \ll P\right\}.$$

where the **supremum is attained** for $\frac{dQ^*}{dP} = S^p/\mathbb{E}[S^p]$, provided that $\mathbb{E}[S^p] \in (0, \infty)$.

If $\varphi(s) = e^{-\alpha s}$ for $\alpha > 0$, then

$$-\alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right] = \inf\left\{\mathbb{E}_Q[S] + \alpha^{-1} KL(Q|P) \colon Q \ll P\right\},$$

where the **infimum is attained** for $\frac{dQ^*}{dP} = e^{c(\alpha)-\alpha S}$ if $c(\alpha) = -\log \mathbb{E}[e^{-\alpha S}]$ is finite.

[Deuschl & Stroock, 1989], [Dai Pra et al, MCSS, 1996], [H & Schütte, JSTAT, 2012], [Schütte, Klus & H, Acta Numerica, 2023]

Certainty equivalence principle, cont'd

If $\varphi(s) = |s|^p$ for $p \geq 1$, it holds

$$(\mathbb{E}[(S(X))^p])^{1/p} = \sup \left\{ \mathbb{E}_Q \left[S(X) \left(\frac{dQ}{dP} \right)^{-1/p} \right] : Q \ll P \right\} .$$

where the **supremum is attained** for $\frac{dQ^*}{dP} = S^p / \mathbb{E}[S^p]$, provided that $\mathbb{E}[S^p] \in (0, \infty)$.

If $\varphi(s) = e^{-\alpha s}$ for $\alpha > 0$, then

$$-\alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right] = \inf \left\{ \mathbb{E}_Q[S] + \alpha^{-1} KL(Q|P) \colon Q \ll P \right\} ,$$

where the **infimum is attained** for $\frac{dQ^*}{dP} = e^{c(\alpha) - \alpha S}$ if $c(\alpha) = -\log \mathbb{E}[e^{-\alpha S}]$ is finite.

[Deuschl & Stroock, 1989], [Dai Pra et al, MCSS, 1996], [H & Schütte, JSTAT, 2012], [Schütte, Klus & H, Acta Numerica, 2023]

Certainty equivalence principle, cont'd

Idea (case $\varphi = \exp$):

- Assume that $P \sim Q$, with $\log L \in L^1(Q)$, and suppose that S is bounded.
- By Jensen's inequality

$$
\begin{aligned}
-\alpha^{-1} \log \int e^{-\alpha S} dP &= -\alpha^{-1} \log \int e^{-\sigma S - \log L} dQ \\
&\leq \int \left(S + \alpha^{-1} \log L\right) dQ \\
&= \mathbb{E}_Q[S] + \alpha^{-1} KL(Q|P)
\end{aligned}
$$

- **Equality** is attained iff $S + \alpha^{-1} \log L$ is constant (Q-a.s.), i.e.

$$
L = \frac{dQ}{dP} = e^{c - \alpha S}
$$

Certainty equivalence principle, cont'd

Idea (case $\varphi = \exp$):

- Assume that $P \sim Q$, with $\log L \in L^1(Q)$, and suppose that S is bounded.
- By Jensen's inequality

$$
\begin{aligned}
-\alpha^{-1} \log \int e^{-\alpha S} dP &= -\alpha^{-1} \log \int e^{-\sigma S - \log L} dQ \\
&\leq \int \left(S + \alpha^{-1} \log L\right) dQ \\
&= \mathbb{E}_Q[S] + \alpha^{-1} KL(Q|P)
\end{aligned}
$$

- **Equality** is attained iff $S + \alpha^{-1} \log L$ is constant (Q-a.s.), i.e.

$$
L = \frac{dQ}{dP} = e^{c - \alpha S}
$$

Certainty equivalence principle, cont'd

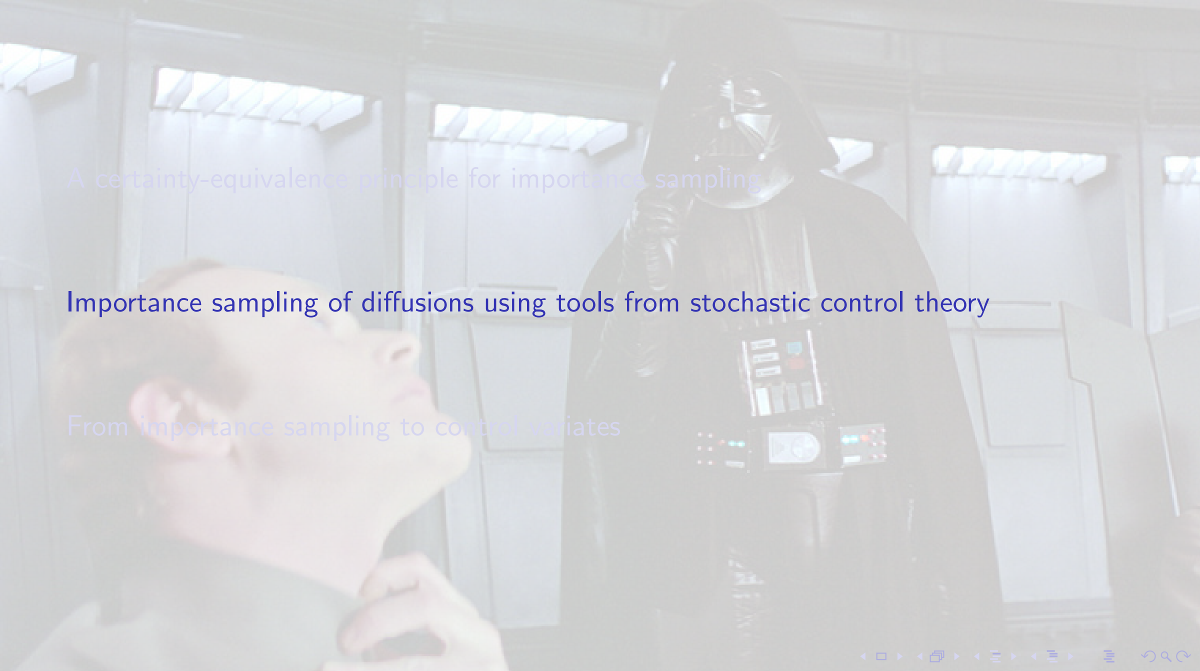
Idea (case $\varphi = \exp$):

- Assume that $P \sim Q$, with $\log L \in L^1(Q)$, and suppose that S is bounded.
- By Jensen's inequality

$$
\begin{aligned}
-\alpha^{-1} \log \int e^{-\alpha S} dP &= -\alpha^{-1} \log \int e^{-\sigma S - \log L} dQ \\
&\leq \int \left(S + \alpha^{-1} \log L\right) dQ \\
&= \mathbb{E}_Q[S] + \alpha^{-1} KL(Q|P)
\end{aligned}
$$

- **Equality** is attained iff $S + \alpha^{-1} \log L$ is constant (Q-a.s.), i.e.

$$
L = \frac{dQ}{dP} = e^{c - \alpha S}
$$



A certainty-equivalence principle for importance sampling

Importance sampling of diffusions using tools from stochastic control theory

From importance sampling to control variates

Set-up

We consider two **diffusion process** X and X^u on $[0, \infty)$ governed by

$$dX_s = b(X_s)ds + \sigma(X_s)dW_s \quad \text{and} \quad dX_s^u = b^u(X_s^u)ds + \sigma(X_s^u)dW_s$$

with

$$b^u(x) = b(x) + \sigma(x)u$$

for any **admissible control** u, such that **Girsanov's Theorem** holds

$$\mathbb{E}[\varphi(S(X^u))L^{-1}] = \mathbb{E}[\varphi(S(X))]$$

where the **likelihood ratio** is given by

$$L = \exp\left(\int_0^T u_s \cdot dW_s - \frac{1}{2}\int_0^T |u_s|^2 ds\right) .$$

Zero-variance importance sampling I

Theorem (Zero-variance estimator for $\varphi(s) = |s|$)

Let h be the classical solution to the linear parabolic boundary value problem

$$\left(\frac{\partial}{\partial t} + \frac{1}{2}\sigma\sigma^T : \nabla_x^2 + b \cdot \nabla_x\right) h = -f \quad \text{in} \quad D$$
$$h = g \quad \text{on} \quad \partial D^+ ,$$

where the precise definitions of the domains D and ∂D^+ depend on the problem at hand. Then $h(x,t) = \mathbb{E}[S(X)|X_t = x]$, and the controlled SDE with control

$$v_s^* = (\sigma(X_s^v))^T \nabla_x \log h(X_s^v, s) \, , \; s \geq t \, .$$

generates a **zero variance change of measure** Q^*.

[Awad, Glynn & Rubinstein, Math Oper Res, 2013]; cf. [Bardou, PhD Thesis, 2005]

Zero-variance importance sampling II

Theorem (Zero-variance estimator for $\varphi(s) = \exp(-\alpha s)$)

Let u^* be a minimiser of the cost functional

$$J(u) = \mathbb{E}\left[S(X^u) + \frac{1}{2\alpha}\int_t^\tau |u_s|^2\, ds\right] \quad (\alpha > 0)$$

under the **controlled dynamics**

$$dX_s^u = (b(X_s^u) + \sigma(X_s^u)u_s)ds + \sigma(X_s^u)dW_s\,, \quad X_t^u = x\,.$$

The minimiser is unique and generates a **zero variance change of measure** Q^*. Moreover

$$J(u^*) = -\alpha^{-1}\log\mathbb{E}\left[e^{-\alpha S(X)}\,\middle|\, X_t = x\right].$$

[H & Schütte, JSTAT, 2012], [H et al, Entropy, 2017]

Superficial comparison between the two cases

- In the case $\varphi = \exp$, the optimal control is again **Markovian feedback control**:

$$u_s^* = -\alpha(\sigma(X_s^u))^T \nabla_x V(X_s^u, s)$$

  where $V = \min_u J(u)$ is the **value function** of the optimal control problem.
- The last expression should be compared to the case when $\varphi = id$, viz.

$$v_s^* = (\sigma(X_s^v))^T \nabla_x \log h(X_s^v, s)$$

- The stochastic control problem for $\varphi = \exp$ is of **linear-quadratic type**, for which a variety of numerical methods exists (meshless, stochastic optimisation based, etc.), whereas the case $\varphi = id$ does not belong in any standard categoty.
- In both cases, we cannot draw directly from Q^*, because the optimal controls that generate $Q^* = Q(u^*)$ or $Q^* = Q(v^*)$ **depend on the quantity of interest**.

[Graham & Talay, 2013], [Kebiri, Neureither & H, IHP Proc, 2019],[Schütte, Klus, H, Acta Numerica, 2023]

Example I: committor probabilities $q_{AB}(x) = P_x(\tau_B < \tau_B)$

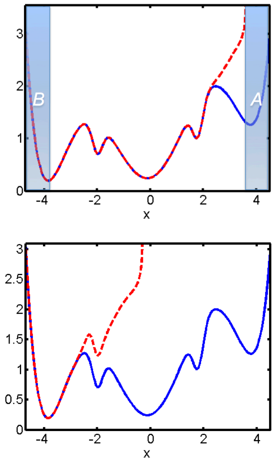
- Underdamped LD $dX_s = -\nabla V(X_s)ds + \sqrt{2}dW_t$
- **Optimally biased potential** ("h-transform") for the case $\varphi = id$ (i.e. $f = 0$, $g = \mathbf{1}_B$, $\tau = \tau_{A\cup B}$):

$$V^* = V + 2\log q_{AB}$$

- **Stochastic control formulation** ($\alpha = 1, f = 0$, $g = -\log \mathbf{1}_B$): minimise the cost

$$J(u) = \mathbb{E}\left[\frac{1}{2}\int_0^{\tau^u} |u_s|^2 ds - \log(\mathbf{1}_{X^u_{\tau^u}\in B})\right],$$

subject to $dX^u_t = (u_t - \nabla V(X^u_t))\,dt + dW_t$ (cf. Margot's poster).



[H et al, Entropy, 2014], [H et al, Nonlinearity, 2016], [Gao et al, MMS, 2023]

Example II: exit time of a Brownian motion

- Let $X_t = x + \sigma W_t$ with $X_0 = x \in (a, b)$, and set

$$\tau = \inf\{t \geq 0 \colon X_t \notin (a, b)\}$$

- **Mean first exit time** $h(x) = \mathbb{E}_x[\tau]$ is given by

$$h(x) = \frac{(b - x)(x - a)}{\sigma^2}, \quad a \leq x \leq b\,.$$

- **Moment-generating function** $\phi(x) = \mathbb{E}_x[e^{-\alpha\tau}]$ is given by

$$\phi(x) = \frac{e^{-\gamma x}\left(e^{\gamma(a+b)} + e^{2\gamma x}\right)}{e^{\gamma a} + e^{\gamma b}}, \quad \gamma = \sqrt{\frac{2\alpha}{\sigma^2}}, \quad a \leq x \leq b\,.$$

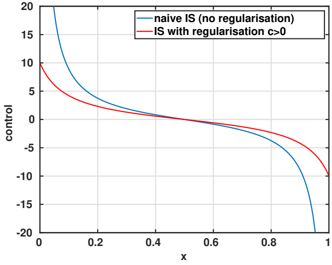
- Relative error of crude MC diverges as $\sigma \to 0$.

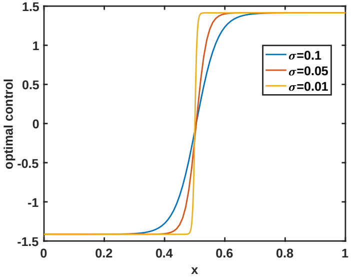
Example II, cont'd: controls

- Bias is singular at domain boundary:

$$\nabla V^*(x) = \nabla V(x) + 2\frac{\nabla h(x)}{h(x)}.$$

- $P(\tau = \infty) = 0$, but $Q(\tau = \infty) = 1$, i.e., IS has **infinite simulation time**





- Control seeks to minimise **variance** and **average simulation** time:

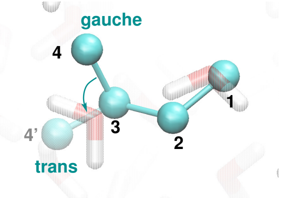
$$\min_u \mathbb{E}_x \left[\tau^u + \frac{1}{2\alpha} \int_0^{\tau^u} |u_s|^2 ds \right]$$

Example III: butane in water ($d = 16224$)

Probability of making a **gauche-trans transition** before time T:

$$-\log \mathbb{P}(\tau \le T) = \min_u \mathbb{E}\left[\frac{1}{2}\int_0^{\tau\wedge T} |u_t|^2\, dt - \log \mathbf{1}_{\partial C}(X^u_{\tau\wedge T})\right],$$

with τ denoting the first exit time from the gauche conformation "C"



T [ps]	$\mathbf{P}(\tau \le T)$	Error	Var	Accel. $\mathcal{I}$
0.1	4.30×10^{-5}	0.77×10^{-5}	3.53×10^{-6}	42.5
0.2	1.21×10^{-3}	0.11×10^{-3}	2.50×10^{-4}	26.0
0.5	6.85×10^{-3}	0.38×10^{-3}	2.88×10^{-3}	13.0
1.0	1.74×10^{-2}	0.08×10^{-2}	1.21×10^{-2}	7.0

IS of butane in a box of 900 water molecules (underdamped LD, SPC/E, GROMOS force field) using **cross-entropy minimisation**

[Zhang et al, SISC, 2014], [H et al, J Comp Dyn, 2014], [Zhang et al, PTRF, 2018], [H & Richter, arXiv:2102.09606, 2023]

Some remarks on the control formulation

- We have replaced a sampling problem by a **variational problem** that admits many formulations (e.g. KL or cross-entropy minimisation, FBSDE, ...), e.g.

$$KL(Q|Q^*) = J(u) - J(u^*)\,,$$

  that give rise to workable numerical algorithms (cf. Weiqing's talk).
- The stochastic control formulation of the sampling problem

$$-\alpha^{-1}\log\mathbb{E}\left[e^{-\alpha S(X)}\right] = \min_u \mathbb{E}\left[S(X^u) + \frac{1}{2\alpha}\int_0^{\tau^u} |u_s|^2\, ds\right]$$

  is consistent with **large deviations** (cf. Hugo's and Zach's talks).
- In many cases the optimal control becomes **stationary** (e.g. MFET, committors).

Some remarks on the control formulation

- We have replaced a sampling problem by a **variational problem** that admits many formulations (e.g. KL or cross-entropy minimisation, FBSDE, ...), e.g.

$$KL(Q|Q^*) = J(u) - J(u^*)\,,$$

  that give rise to workable numerical algorithms (cf. Weiqing's talk).
- The stochastic control formulation of the sampling problem

$$-\alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right] = \min_u \mathbb{E}\left[S(X^u) + \frac{1}{2\alpha}\int_0^{\tau^u} |u_s|^2\, ds\right]$$

  is consistent with **large deviations** (cf. Hugo's and Zach's talks).
- In many cases the optimal control becomes **stationary** (e.g. MFET, committors).

Some remarks on the control formulation

- We have replaced a sampling problem by a **variational problem** that admits many formulations (e.g. KL or cross-entropy minimisation, FBSDE, ...), e.g.

$$KL(Q|Q^*) = J(u) - J(u^*)\,,$$

  that give rise to workable numerical algorithms (cf. Weiqing's talk).

- The stochastic control formulation of the sampling problem

$$-\alpha^{-1}\log\mathbb{E}\left[e^{-\alpha S(X)}\right] = \min_u \mathbb{E}\left[S(X^u) + \frac{1}{2\alpha}\int_0^{\tau^u} |u_s|^2\, ds\right]$$

  is consistent with **large deviations** (cf. Hugo's and Zach's talks).

- In many cases the optimal control becomes **stationary** (e.g. MFET, committors).

A certainty-equivalence principle for importance sampling

Importance sampling of diffusions using tools from stochastic control theory

From importance sampling to control variates

Computation of the first moment

- Q^* generated by solving the SOC problem is in general **not a good change of measure** for other purposes, e.g. the mean, because

$$Var(\tau^u L^{-1}) \geq Var(\tau) \quad \text{and} \quad \mathbb{E}[\tau^u]\, Var(\tau^u L^{-1}) \geq \mathbb{E}[\tau]\, Var(\tau)$$

- The LHS is the scaled **cumulant-generating function** of S; for small α,

$$-\alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right] \approx \mathbb{E}[S(X)] - \frac{\alpha}{2} Var(S(X)),$$

in particular

$$\mathbb{E}[S(X)] = -\lim_{\alpha \searrow 0} \alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right]$$

- However, for small α, the control becomes **heavily penalised**, and we cannot expect the likelihood of the rare event to significantly increase.

[Badowski, PhD Thesis, 2016], [Schütte, Klus & H, Acta Numerica, 2023]

Computation of the first moment

- Q^* generated by solving the SOC problem is in general **not a good change of measure** for other purposes, e.g. the mean, because

$$Var(\tau^u L^{-1}) \geq Var(\tau) \quad \text{and} \quad \mathbb{E}[\tau^u]\, Var(\tau^u L^{-1}) \geq \mathbb{E}[\tau]\, Var(\tau)$$

- The LHS is the scaled **cumulant-generating function** of S; for small α,

$$-\alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right] \approx \mathbb{E}[S(X)] - \frac{\alpha}{2}\, Var(S(X))\,,$$

in particular

$$\mathbb{E}[S(X)] = -\lim_{\alpha \searrow 0} \alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right]$$

- However, for small α, the control becomes **heavily penalised**, and we cannot expect the likelihood of the rare event to significantly increase.

[Badowski, PhD Thesis, 2016], [Schütte, Klus & H, Acta Numerica, 2023]

Computation of the first moment

- Q^* generated by solving the SOC problem is in general **not a good change of measure** for other purposes, e.g. the mean, because

$$Var(\tau^u L^{-1}) \geq Var(\tau) \quad \text{and} \quad \mathbb{E}[\tau^u]\, Var(\tau^u L^{-1}) \geq \mathbb{E}[\tau]\, Var(\tau)$$

- The LHS is the scaled **cumulant-generating function** of S; for small α,

$$-\alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right] \approx \mathbb{E}[S(X)] - \frac{\alpha}{2} Var(S(X))\,,$$

in particular

$$\mathbb{E}[S(X)] = -\lim_{\alpha \searrow 0} \alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X)}\right]$$

- However, for small α, the control becomes **heavily penalised**, and we cannot expect the likelihood of the rare event to significantly increase.

[Badowski, PhD Thesis, 2016], [Schütte, Klus & H, Acta Numerica, 2023]

From importance sampling to control variates

Theorem (Control variate limit)

Let X denote the solution to the **uncontrolled SDE**, and let X^* be the solution under the optimal control u^*, with likelihood ratio $L = L(u^*) > 0$. Then, as $\alpha \to 0$,

$$-\alpha^{-1} \log \mathbb{E}\left[e^{-\alpha S(X^*)} L^{-1}\right] \to \mathbb{E}\left[S(X) - \int_0^\tau Z_s \cdot dW_s\right]$$

where

$$Z_s = (\sigma(X_s))^T \nabla_x h(X_s, s)$$

with h being the classical solution to the linear boundary value problem associated with the **linear expectation** for $\varphi = id$. Moreover,

$$Var\left(S(X) + \int_0^\tau Z_s \cdot dW_s\right) = 0$$

[Schütte, Klus & H, Acta Numerica, 2023]; [H et al, in preparation]; cf. [H et al, Chaos, 2019]

Some remarks on the control variates limit

- Assuming **sufficient regularity of the value function** V and its spatial derivative $\nabla_x V$, it holds that on any compact subset of $\mathbb{R}^d \times [0, \infty)$:

$$V \to h \quad \text{and} \quad \nabla_x V \to \nabla_x h\,.$$

- The Itô integral in

$$S(X) - \int_0^\tau Z_s \cdot dW_s$$

  is a **control variate** that nullifies the variance of S. Note that Z depends on $\nabla_x h$, not on $\nabla_x \log h$ like the corresponding IS estimator.
- The control variate idea is not new (see, e.g., Graham's and Talay's book), but the connection to stochastic optimal control allows for a generalisation to the case when the underlying HJB equation has only a **viscosity solution**.

[Newton, SIMA, 1994], [Boyalval & Lelièvre, CMS, 2010], [Graham & Talay, 2013], [Roussel, PhD Thesis, 2018]

Example II, cont'd: exit time of a Brownian motion

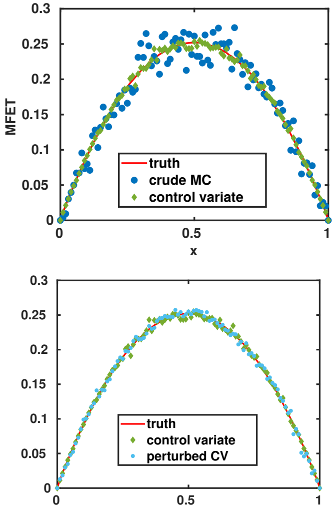
- Exit of a 1-dimensional **Brownian motion**

$$X_t = x + W_t$$

from the interval $I = (0,1)$:

$$\mathbb{E}_x[\tau] \approx \tau + \int_0^\tau (2X_s - 1)\,dW_s$$

- **Comparison** of crude MC and control variates for $n = 100$ sample points
- Estimator **robust under perturbations** of integrand (about 20% in sup-norm); observed errors are likely due to EM discretisation.



Take-home message

- Adaptive IS scheme based on **exponential averages**; resulting control problem features short trajectories with **minimum variance estimators**.
- Direct IS approach for the mean may have issues for **unbounded stopping times** that can lead to infinitely long simulation times after reweighting.
- Generally, IS may be **sensitive to bad approximations of the control**, especially in high-dimensions; for random stopping times, there is a trade-off between the average trajectory length and approximation error

$$\delta_{\mathrm{rel}} = O\left(e^{\mathrm{error}^2 \times \mathbb{E}[\tau^{(2u^* - u)}]}\right).$$

- **Control variates** can cope with the somewhat pathological exit time case.

[Bickel, Li & Bengtsson, 2008], [Agapiou et al, Stat Sci, 2015], [H & Richter, arXiv:2102.09606, 2023]

Thank you for your attention!

Acknowledgement:

David Aristoff
Annika Jöster
Omar Kebiri
Lara Neureither
Lorenz Richter
Christof Schütte
Markus Strehlau
Markus Weber

German Science Foundation (DFG)