Gaussian Approximation of Transition Paths

Andrew Stuart

Mathematics Institute, University of Warwick

COSMOS, Paris, February 2nd – 5th 2016

Collaboration with Yulong Lu and Hendrik Weber
Also Frank Pinski, Gideon Simpson, Florian Theil

Funded by EPSRC, ERC and ONR

Outline

Outline

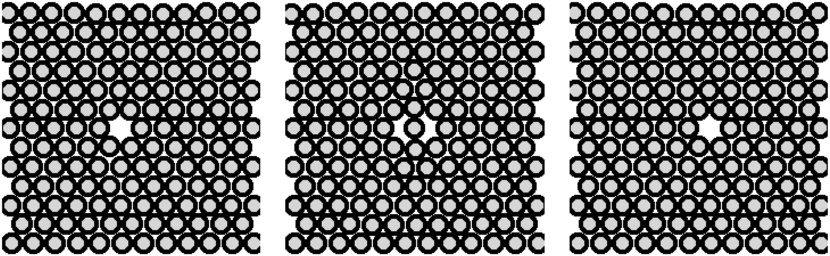
Mathematical Statement

Brownian Dynamics

- $V : \mathbb{R}^d \to \mathbb{R}^+$.
- $\mathcal{E} = \{x \in \mathbb{R}^d : \nabla V(x) = 0\}$.
- $x^\pm \in \mathcal{E}$.
- $\varepsilon > 0$ temperature.
- $T > 0$ transition time.

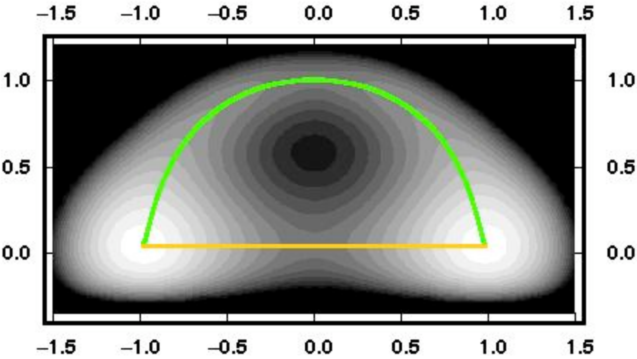
$$\frac{dx}{dt} = -\nabla V(x) + \sqrt{2\varepsilon}\frac{dW}{dt},$$
$$x(0) = x^-, \quad x(T) = x^+.$$

Example I: Vacancy Diffusion



[2] F. Pinski and A.M. Stuart,
Transition paths in molecules at finite temperature,
J. Chem. Phys., 2010.

Example II: Two Routes



[2] F. Pinski and A.M. Stuart,
Transition paths in molecules at finite temperature,
J. Chem. Phys., 2010.

Outline

Probabilistic Formulation

$$\frac{dx}{dt} = -\alpha \nabla V(x) + \sqrt{2\epsilon}\frac{dW}{dt},$$
$$x(0) = x^-, \quad x(T) = x^+.$$

Path Space Measures

- μ_α the induced probability meausure on $C([0,T];\mathbb{R}^d)$.
- μ_0 is Brownian bridge.
- $\mu := \mu_1$ is measure of interest.

Orientation

Key Questions

- What is the most likely path connecting x_- and x_+ at finite temperature $\varepsilon > 0$?
- In the low temperature limit $\varepsilon \to 0$ what happens to the most likely path?

Our Approach

- Characterize best Gaussian approximation to μ.
- Use $\Gamma-$ convergence to study the small temperature limit.

Freidlin-Wentzell Theory [1]

Large Deviations

- Let $B_\delta(\varphi)$ be a ball of radius δ in $C([0,T];\mathbb{R}^n)$ centred at φ.
- For fixed T:

$$\mathbb{P}^\mu\left(B_\delta(\varphi)\right) \approx \exp\left(-\frac{1}{2\varepsilon}\overline{S}_T(\varphi)\right).$$

- The action functional is

$$\overline{S}_T(\varphi) := \frac{1}{2}\int_0^T \left(\varphi'(t) + \nabla V(\varphi)\right)^2 dt.$$

- The most likely paths are thus minimizers of

$$S_T(\varphi) := \frac{1}{2}\int_0^T \left|\varphi'(t)\right|^2 + \left|\nabla V\left(\varphi(t)\right)\right|^2 dt.$$

Onsager-Machlup Theory [3]

Maximizing Small Ball Probabilities

- Let $B_\delta(\varphi)$ be a ball of radius δ in $C([0,T];\mathbb{R}^n)$ centred at φ.
- Then $\lim_{\delta\to 0} \frac{\mathbb{P}\big(B_\delta(\varphi_1)\big)}{\mathbb{P}\big(B_\delta(\varphi_2)\big)} = \exp\big(I_{\varepsilon,T}(\varphi_2) - I_{\varepsilon,T}(\varphi_1)\big)$.
- Here $I_{\varepsilon,T}$ is the Onsager-Machlup functional defined by

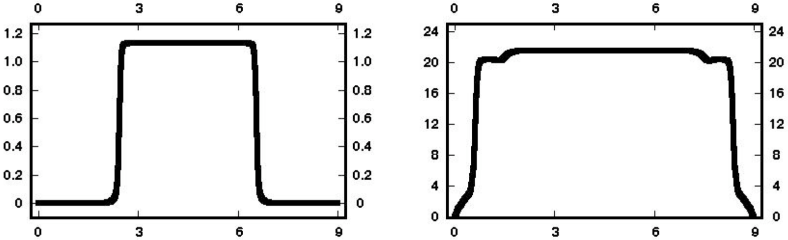
$$I_{\varepsilon,T}(\varphi) := S_T(\varphi) - \int_0^T \Big(\varepsilon \Delta V\big(\varphi(t)\big)\Big) dt.$$

- Probability of the small ball is maximized when centred at minimizers of $I_{\varepsilon,T}$ (MAP estimator in statistics).
- The Itô correction $\varepsilon \Delta V$ may produce non-physical transition paths, in contrast to large deviation approach.

Example I – Vacancy Diffusion

Onsager-Machlup Minimizers [2]

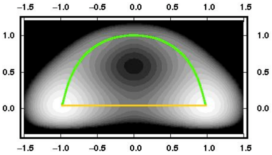
- Left: zero temperature.
- Right: finite temperture.
- Saddle is preferred.

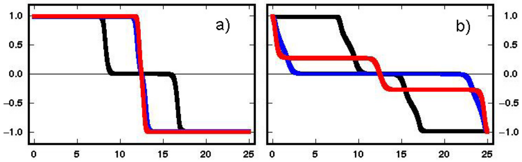


Example II – Two Routes

Onsager-Machlup Minimizers [2]

- Left: two routes.
- Middle: straight route.
- Right: curved route.
- $\varepsilon = 0$ (black), $\varepsilon = 10^{-2}$ (blue), $\varepsilon = 10^{-1}$ (red).
- Low entropy route preferred.





Outline

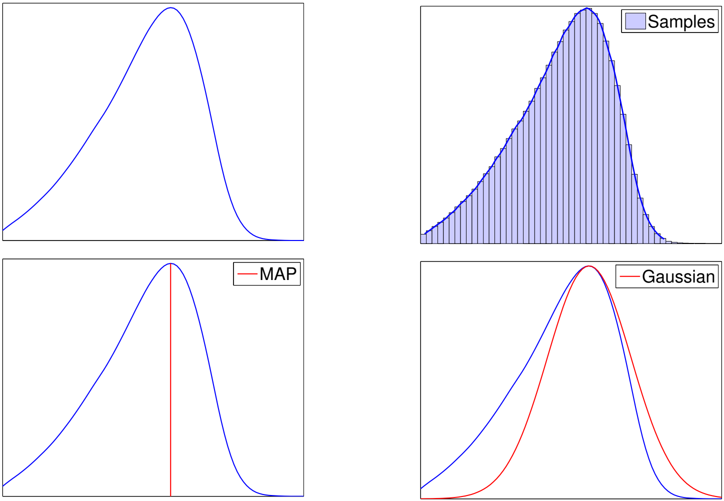
Formulation With $T = \varepsilon^{-1}, t \mapsto \varepsilon t$ [3]

$$\frac{dx}{dt} = -\frac{\alpha}{\epsilon}\nabla V(x) + \sqrt{2}\frac{dW}{dt},$$
$$x(0) = x^-, \quad x(1) = x^+.$$

Thus measure of interest μ has density with respect to the Brownian bridge μ_0 which is given by:

$$\mu(dx) = Z^{-1}\exp\Big(-\frac{1}{2\varepsilon^2}\Phi_\epsilon(x)\Big)\mu_0(dx),$$
$$Z = \mathbb{E}^{\mu_0}\exp\Big(-\frac{1}{2\varepsilon^2}\Phi_\epsilon(x)\Big),$$
$$\Phi_\epsilon(x) = \int_0^1 \Big(\frac{1}{2}|\nabla V\big(x(t)\big)|^2 - \varepsilon\triangle V\big(x(t)\big)\Big)dt.$$

Approximating Measures



Kullback-Leibler Approximation

For ν, μ probability measures define the K-L divergence

$$D_{\mathsf{KL}}(\nu\|\mu) = \mathbb{E}^{\nu} \log\left(\frac{d\nu}{d\mu}\right) \text{ if } \nu \ll \mu$$
$$= \infty \text{ otherwise.}$$

Theorem [4]

- Let $\mu \ll \mu_0 = N(m, C)$.
- Let $\mathcal{A}$ denote the set of all Gaussians equivalent to μ_0.
- Then there is $\nu \in \mathcal{A}$ that minimizes $D_{\mathsf{KL}}(\nu\|\mu)$.

[4] F. Pinski, G. Simpson, A.M. Stuart and H. Weber,
Kullback-Leibler Approximation for Probability Measures on Infinite Dimensional Spaces,
SIAM J. Math. Analysis, **47**(2015), 4091–4122.

Approximation by Inhomogeneous OU Processes [6]

$$dz(t) = -\varepsilon^{-1}A(t)z(t) + \sqrt{2}dW(t),$$
$$z(0) = z(1) = 0.$$

Approximation Class (in dimension $d = 1$ for exposition)

- If $C = 2(-\Delta + B_\varepsilon)^{-1}$, where C^{-1} has domain $H^2 \cap H_0^1$ and $B_\varepsilon = \varepsilon^{-2}A^2 - \varepsilon^{-1}A'$, then $z \sim N(0, C)$.
- Here we assume that, for some $a > 0$,

$$m \in H^1_\pm(0,1) := \{x \in H^1(0,1) : x(0) = x^-, x(1) = x^+\},$$
$$A \in H^1_a := \{u \in H^1(0,1) : u \geq a \text{ a.e.}\}.$$

- Define $\mathcal{A} := \{x = m + z, m \in H^1_\pm(0,1), z \sim N(0, C)\}$.
- We aim to find $\nu \in \mathcal{A}$ to minimize $D_{\mathrm{KL}}(\nu \| \mu)$.

Calculation of KL Divergence

Let $\overline{A}(t) = \int_t^1 A(s)ds$. Then, in dimension $d = 1$ for exposition,

$$
\begin{aligned}
D_{KL}(\nu\|\mu) &= \frac{1}{2\epsilon} F_\epsilon(m, A) + \text{const} \\
F_\epsilon(m, A) &= \frac{\varepsilon}{2}\int_0^1 |m'(t)|^2 dt + \frac{1}{\varepsilon}\mathbb{E}^\nu \Phi_\varepsilon(m+z) \\
&\quad - \frac{1}{4}\mathbb{E}^\nu \int_0^1 B_\varepsilon(t)(z(t))^2 dt + \frac{1}{2\varepsilon}\int_0^1 A(t)dt \\
&\quad + \frac{1}{2}\log\left(\int_0^1 e^{-2\overline{A}(t)/\varepsilon} dt\right)
\end{aligned}
$$

F_ε is an entropically fattened large deviations rate function.

Calculation of KL Divergence

Let $\overline{A}(t) = \int_t^1 A(s)ds$. Then, in dimension $d = 1$ for exposition,

$$
\begin{aligned}
D_{KL}(\nu\|\mu) =& \frac{1}{2\epsilon} F_\epsilon(m, A) + \text{const} \\
F_\epsilon(m, A) =& \frac{\varepsilon}{2}\int_0^1 |m'(t)|^2 dt + \frac{1}{\varepsilon}\mathbb{E}^\nu \Phi_\varepsilon(m+z) \\
& - \frac{1}{4}\mathbb{E}^\nu \int_0^1 B_\varepsilon(t)\big(z(t)\big)^2 dt + \frac{1}{2\varepsilon}\int_0^1 A(t)dt \\
& + \frac{1}{2}\log\left(\int_0^1 e^{-2\overline{A}(t)/\varepsilon} dt\right)
\end{aligned}
$$

F_ε is an entropically fattened large deviations rate function.

Calculus of Variations

Theorem (Regularized D_{KL} Minimization) [6]

For $m \in H^1_\pm, A \in H^1_a$, and for any $\gamma > 0$, define

$$J_\varepsilon(m, A) = F_\varepsilon(m, A) + \varepsilon^\gamma \|A\|^2_{H^1}.$$

Assume that $\inf_{\nu \in \mathcal{A}} D_{KL}(\nu \| \mu) < \infty$, then there exists $m_\varepsilon \in H^1_\pm, A_\varepsilon \in H^1_a$ that minimizes $J_\varepsilon(m, A)$.

Remarks

- Regularization is applied to obtain compactness w.r.t. A.
- Optimal transition paths charaterized by a Gaussian tube centered at m with variance of order ε.

Outline

Γ Convergence

Questions

- When $\varepsilon \to 0$:
- what is limit (m, A) of $(m_\varepsilon, A_\varepsilon)$?
- what is the limiting functional F of F_ε?
- does (m, A) minimize F?

Answer

- Study convergence of functionals by Γ-convergence.
- Compactness + Γ-convergence implies convergence of minima.
- Compactness means

$$\limsup_{\varepsilon \to 0} F_\varepsilon(u_\varepsilon) < \infty \quad \text{implies} \quad \exists u \text{ such that } u_\varepsilon \to u.$$

$\Gamma-$limit of F_ε

Theorem [6]

Under conditions on V and if $\gamma \in (0, \frac{1}{2})$, then the Γ-limit of F_ε on $L^1(0,1) \times L^1_a(0,1)$ is

$$F(m, A) := E(m) + \int_0^1 \frac{1}{4A(t)} \left(V''(m(t)) - A(t) \right)^2 dt$$

where

$$E(m) := \begin{cases} \sum_{\tau \in J(m)} \Psi(m(\tau-), m(\tau+)) & \text{if } m \in BV_\pm(0,1;\mathcal{E}), \\ +\infty & \text{otherwise in } L^1(0,1). \end{cases}$$

Here $J(m)$ is the jump set of m and Ψ is the energy cost for each jump, namely for $x_1, x_2 \in \mathcal{E}$,

$$\Psi(x_1, x_2) = |V(x_2) - V(x_1)|.$$

Sketch Proof

Expanding in ε and using $\mathbb{E}z^2(t) = \mathbb{E}[x(t) - m(t)]^2 \approx \varepsilon / A(t)$:

$$F_\varepsilon(m_\varepsilon, A_\varepsilon) \approx \frac{\varepsilon}{4}\int_0^1 m_\varepsilon'(t)^2 dt + \frac{1}{4\varepsilon}\int_0^1 V'(m_\varepsilon(t))^2 dt$$
$$+ \int_0^1 \frac{1}{4A_\varepsilon(t)}\Big((V''(m_\varepsilon(t)) - A_\varepsilon(t))^2 + V'(m_\varepsilon(t))V^{(3)}(m_\varepsilon(t))\Big)\,dt.$$

Note that:

$$F_\varepsilon(m_\varepsilon, A_\varepsilon) \approx \frac{1}{2}S_{\epsilon^{-1}}(m_\varepsilon) + \int_0^1 \frac{1}{4A_\varepsilon(t)}\big(V''(m_\varepsilon(t)) - A_\varepsilon(t)\big)^2$$
$$+ V'(m_\varepsilon(t))V^{(3)}(m_\varepsilon(t))dt.$$

Sketch Proof

$$F_\varepsilon(m_\varepsilon, A_\varepsilon) \approx \frac{1}{2} S_{\epsilon^{-1}}(m_\varepsilon) + \int_0^1 \frac{1}{4A_\varepsilon(t)} \left(V''(m_\varepsilon(t)) - A_\varepsilon(t)\right)^2 + V'(m_\varepsilon(t)) V^{(3)}(m_\varepsilon(t)) dt.$$

Remarks

- The first term reflects large deviations, depends only on the mean, and minimizers are well understood, using $\Gamma-$convergence, in the small temperature limit.
- The Itô correction term can be killed by setting $A_\varepsilon = V''(m_\varepsilon)$: fluctation determined by linearization of the original dynamics around mean path.
- $V'(m_\varepsilon)V^{(3)}(m_\varepsilon)$ vanishes in the limit since the large deviations energy functional forces $V'(m_\varepsilon) \approx 0$.

Outline

Conclusions

- Large deviations predicts physical transition paths.
- Onsager-Machlup theory (MAP estimator) produces non-physical transition paths.
- Entropy is key to understanding this dichotomy.
- "Best Gaussian" approximation w.r.t. Kullback-Leibler.
- Gaussian approximation explicitly incorporates entropy.
- Characterize most likely transition paths as optimal Gaussian tubes around large deviation mean.
- Gaussian tube is defined by OU fluctuations involving linearization at critical points.
- In the low temperature limit, Kullback-Leibler approximation approach removes undersirable Itô correction in Onsager-Machlup and recovers large deviations.

References

[1] "Random perturbations of dynamical systems",
M.I.Freidlin and A.D.Wentzell.
Springer, 2012 (Third Edition).

[2] "Transition paths in molecules at finite temperature",
F. Pinski and A.M. Stuart.
J. Chem. Phys. **32**, 184104 (2010).

[3] "Γ-limit for transition paths of maximal probability",
F. Pinski, A.M. Stuart and F. Theil.
J. Stat. Phys. **146**, 955-974 (2012).

[4] "Kullback-Leibler approximation for probability measures on infinite dimensional spaces" ,
F. Pinski, G. Simpson, A. M. Stuart and H. Weber.
SIAM J. Math. Analysis, **47**:4091-4122 (2015).

[5] "Algorithms for Kullback-Leibler approximation for probability measures in infinite dimensions." ,
F. Pinski, G. Simpson, A. M. Stuart and H. Weber.
SIAM J. Sci. Comp., to appear `http://arxiv.org/abs/1408.1920`

[6] "Gaussian approximations for transition paths in molecular dynamics",
Y. Lu, A. M. Stuart and H. Weber.
In Preparation (2016).