Large deviations and variational representations for infinite dimensional systems

Paul Dupuis

Division of Applied Mathematics
Brown University

IMA

with Amarjit Budhiraja

M. Boué, J. Chen, M. Fischer, A. Ganguly, V. Maroulas, K. Spiliopoulos

January 2013

Outline

1. Techniques for process-level large deviation problems
2. Example of a representation in continuous time and an application
3. History of control ideas in large deviations
4. Representations for infinite dimensional Brownian motion
5. Derivation of the representation
6. Application to a problem of pattern theory
7. Some remarks

Process-level large deviation problems

Focus on continuous time, Markovian models.

Process-level large deviation problems

Focus on continuous time, Markovian models. Finite dimensional state example:

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

with $X^{\varepsilon}(t) \in \mathbb{R}^d$, $W(t) \in \mathbb{R}^k$, and b and σ of suitable dimensional and nice.

Process-level large deviation problems

Focus on continuous time, Markovian models. Finite dimensional state example:

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

with $X^{\varepsilon}(t) \in \mathbb{R}^d$, $W(t) \in \mathbb{R}^k$, and b and σ of suitable dimensional and nice. We consider $X^{\varepsilon}(\cdot) \in \mathcal{E} \doteq C([0,T]:\mathbb{R}^d)$. An LDP means

1. there is a *rate function* $I : \mathcal{E} \to [0,\infty]$ with compact level sets,
2. there are the bounds

$$\begin{aligned}
\liminf_{\varepsilon\to 0} \varepsilon \log P\left(X^{\varepsilon} \in A\right) &\geq -\inf_{\phi\in A^{\circ}} I(\phi) \\
\limsup_{\varepsilon\to 0} \varepsilon \log P\left(X^{\varepsilon} \in A\right) &\leq -\inf_{\phi\in \bar{A}} I(\phi).
\end{aligned}$$

Process-level large deviation problems

We use the (equivalent) Laplace formulation: for all bounded continuous $F : \mathcal{E} \to \mathbb{R}$

$$\lim_{\varepsilon \to 0} \varepsilon \log E e^{-\frac{1}{\varepsilon} F(X^\varepsilon)} = -\inf_{\phi} \left[F(\phi) + I(\phi) \right].$$

Process-level large deviation problems

We use the (equivalent) Laplace formulation: for all bounded continuous $F : \mathcal{E} \to \mathbb{R}$

$$\lim_{\varepsilon \to 0} \varepsilon \log E e^{-\frac{1}{\varepsilon} F(X^{\varepsilon})} = -\inf_{\phi} \left[F(\phi) + I(\phi) \right].$$

A control-type formulation of rate function:

$$I(\phi) = \inf \left\{ \frac{1}{2} \int_0^T \|u(t)\|^2 \, dt : \dot{\phi}(t) = b(\phi(t)) + \sigma(\phi(t)) u(t), \phi(0) = x_0 \right\}.$$

Process-level large deviation problems

We use the (equivalent) Laplace formulation: for all bounded continuous $F : \mathcal{E} \to \mathbb{R}$

$$\lim_{\varepsilon \to 0} \varepsilon \log E e^{-\frac{1}{\varepsilon} F(X^{\varepsilon})} = - \inf_{\phi} \left[F(\phi) + I(\phi) \right].$$

A control-type formulation of rate function:

$$I(\phi) = \inf \left\{ \frac{1}{2} \int_0^T \|u(t)\|^2 \, dt : \dot{\phi}(t) = b(\phi(t)) + \sigma(\phi(t)) u(t), \phi(0) = x_0 \right\}.$$

Let $\phi = \mathcal{G}^0(u)$ denote the mapping $u \to \phi$. Then the (expected) Laplace result is

$$\lim_{\varepsilon \to 0} -\varepsilon \log E e^{-\frac{1}{\varepsilon} F(X^{\varepsilon})} = \inf_{u} \left[F(\mathcal{G}^0(u)) + \frac{1}{2} \int_0^T \|u(t)\|^2 \, dt \right].$$

Process-level large deviation problems

Many of the first methods were based on super-exponentially close approximation.

Process-level large deviation problems

Many of the first methods were based on super-exponentially close approximation. Examples:

- Approach introduced by Freidlin and Wentzell considers random vector

$$X^{\varepsilon,\delta} \doteq (X^{\varepsilon}(0), X^{\varepsilon}(\delta), X^{\varepsilon}(2\delta), \ldots, X^{\varepsilon}(T)).$$

  Proves LD bounds for $X^{\varepsilon,\delta}$ and transfers them to $X^{\varepsilon}(\cdot)$ in limit $\delta \downarrow 0$.

Process-level large deviation problems

Many of the first methods were based on super-exponentially close approximation. Examples:

- Approach introduced by Freidlin and Wentzell considers random vector
$$X^{\varepsilon,\delta} \doteq (X^{\varepsilon}(0), X^{\varepsilon}(\delta), X^{\varepsilon}(2\delta), \ldots, X^{\varepsilon}(T)).$$
Proves LD bounds for $X^{\varepsilon,\delta}$ and transfers them to $X^{\varepsilon}(\cdot)$ in limit $\delta \downarrow 0$.

Process-level large deviation problems

Many of the first methods were based on super-exponentially close approximation. Examples:

- Approach introduced by Freidlin and Wentzell considers random vector
$$X^{\varepsilon,\delta} \doteq \left(X^{\varepsilon}(0), X^{\varepsilon}(\delta), X^{\varepsilon}(2\delta), \ldots, X^{\varepsilon}(T)\right).$$
Proves LD bounds for $X^{\varepsilon,\delta}$ and transfers them to $X^{\varepsilon}(\cdot)$ in limit $\delta \downarrow 0$.
- Approach introduced by Azencott based on knowing the LDP for "driving noise" $\sqrt{\varepsilon}W$ and uses
$$dX^{\varepsilon,\delta}(t) = b(X^{\varepsilon,\delta}(t))dt + \sqrt{\varepsilon}\sigma(X^{\varepsilon,\delta}(\delta\lfloor t/\delta \rfloor))dW(t), \quad X^{\varepsilon,\delta}(0) = x_0.$$
Continuous mapping theorem gives LDP for $X^{\varepsilon,\delta}$ and again transfer to X^{ε} in limit $\delta \downarrow 0$.

Process-level large deviation problems

Many of the first methods were based on super-exponentially close approximation. Examples:

- Approach introduced by Freidlin and Wentzell considers random vector

$$X^{\varepsilon,\delta} \doteq (X^{\varepsilon}(0), X^{\varepsilon}(\delta), X^{\varepsilon}(2\delta), \ldots, X^{\varepsilon}(T)).$$

  Proves LD bounds for $X^{\varepsilon,\delta}$ and transfers them to $X^{\varepsilon}(\cdot)$ in limit $\delta \downarrow 0$.

- Approach introduced by Azencott based on knowing the LDP for "driving noise" $\sqrt{\varepsilon}W$ and uses

$$dX^{\varepsilon,\delta}(t) = b(X^{\varepsilon,\delta}(t))dt + \sqrt{\varepsilon}\sigma(X^{\varepsilon,\delta}(\delta \lfloor t/\delta \rfloor))dW(t), \quad X^{\varepsilon,\delta}(0) = x_0.$$

  Continuous mapping theorem gives LDP for $X^{\varepsilon,\delta}$ and again transfer to X^{ε} in limit $\delta \downarrow 0$.

Process-level large deviation problems

Many of the first methods were based on super-exponentially close approximation. Examples:

- Approach introduced by Freidlin and Wentzell considers random vector

$$X^{\varepsilon,\delta} \doteq (X^{\varepsilon}(0), X^{\varepsilon}(\delta), X^{\varepsilon}(2\delta), \ldots, X^{\varepsilon}(T)).$$

  Proves LD bounds for $X^{\varepsilon,\delta}$ and transfers them to $X^{\varepsilon}(\cdot)$ in limit $\delta \downarrow 0$.

- Approach introduced by Azencott based on knowing the LDP for "driving noise" $\sqrt{\varepsilon}W$ and uses

$$dX^{\varepsilon,\delta}(t) = b(X^{\varepsilon,\delta}(t))dt + \sqrt{\varepsilon}\sigma(X^{\varepsilon,\delta}(\delta\lfloor t/\delta \rfloor))dW(t), \quad X^{\varepsilon,\delta}(0) = x_0.$$

  Continuous mapping theorem gives LDP for $X^{\varepsilon,\delta}$ and again transfer to X^{ε} in limit $\delta \downarrow 0$.

Approximations complicate argument, impose extraneous conditions (a significant problem in infinite dimensions).

Finite dimensional representation and an application

A representation from [Boué-D, 1998] asserts that given any bounded measurable $\Theta : C([0,T] : \mathbb{R}^k) \to \mathbb{R}$,

$$-\varepsilon \log E e^{-\frac{1}{\varepsilon}\Theta(\sqrt{\varepsilon}W)} = \inf_{u \in \mathcal{A}} E\left[\Theta\left(\int_0^{\cdot} u(t)dt + \sqrt{\varepsilon}W(\cdot)\right) + \frac{1}{2}\int_0^T \|u(t)\|^2 dt\right],$$

Finite dimensional representation and an application

A representation from [Boué-D, 1998] asserts that given any bounded measurable $\Theta : C([0,T] : \mathbb{R}^k) \to \mathbb{R}$,

$$-\varepsilon \log E e^{-\frac{1}{\varepsilon}\Theta(\sqrt{\varepsilon}W)} = \inf_{u \in \mathcal{A}} E\left[\Theta\left(\int_0^{\cdot} u(t)dt + \sqrt{\varepsilon}W(\cdot)\right) + \frac{1}{2}\int_0^T \|u(t)\|^2 dt\right],$$

where

$$\mathcal{A} \doteq \left\{u : u \text{ is } \mathcal{F}_t\text{-progressively measurable and } E\int_0^T \|u(t)\|^2 dt < \infty\right\}$$

and W is an $\mathcal{F}_t$-Brownian motion.

Finite dimensional representation and an application

Suppose that there is a unique strong solution to the SDE, with

$$X^{\varepsilon}(\cdot) = \mathcal{G}^{\varepsilon}\left[\sqrt{\varepsilon}W\right](\cdot), \quad \mathcal{G}^{\varepsilon} : C([0,T] : \mathbb{R}^{k}) \to C([0,T] : \mathbb{R}^{d}).$$

Finite dimensional representation and an application

Suppose that there is a unique strong solution to the SDE, with

$$X^\varepsilon(\cdot) = \mathcal{G}^\varepsilon\left[\sqrt{\varepsilon}W\right](\cdot), \quad \mathcal{G}^\varepsilon : C([0,T]:\mathbb{R}^k) \to C([0,T]:\mathbb{R}^d).$$

Then for bounded, continuous F automatically

$$F(X^\varepsilon(\cdot)) = F(\mathcal{G}^\varepsilon\left[\sqrt{\varepsilon}W\right](\cdot)) = \Theta(\sqrt{\varepsilon}W)$$

qualifies, so that

$$\begin{aligned} &-\varepsilon \log Ee^{-\frac{1}{\varepsilon}F(X^\varepsilon)} \\ &\quad = \inf_{u\in\mathcal{A}} E\left[F\left(\mathcal{G}^\varepsilon\left[\int_0^\cdot u(t)dt + \sqrt{\varepsilon}W\right]\right) + \frac{1}{2}\int_0^T \|u(t)\|^2\,dt\right]. \end{aligned}$$

Finite dimensional representation and an application

Suppose that there is a unique strong solution to the SDE, with

$$X^{\varepsilon}(\cdot) = \mathcal{G}^{\varepsilon}\left[\sqrt{\varepsilon}W\right](\cdot), \quad \mathcal{G}^{\varepsilon} : C([0,T]:\mathbb{R}^k) \to C([0,T]:\mathbb{R}^d).$$

Then for bounded, continuous F automatically

$$F(X^{\varepsilon}(\cdot)) = F(\mathcal{G}^{\varepsilon}\left[\sqrt{\varepsilon}W\right](\cdot)) = \Theta(\sqrt{\varepsilon}W)$$

qualifies, so that

$$\begin{aligned} &-\varepsilon \log E e^{-\frac{1}{\varepsilon}F(X^{\varepsilon})} \\ &\qquad = \inf_{u\in\mathcal{A}} E\left[F\left(\mathcal{G}^{\varepsilon}\left[\int_0^{\cdot} u(t)dt + \sqrt{\varepsilon}W\right]\right) + \frac{1}{2}\int_0^T \|u(t)\|^2\, dt\right]. \end{aligned}$$

But by strong uniquness $\mathcal{G}^{\varepsilon}\left[\int_0^{\cdot} u(t)dt + \sqrt{\varepsilon}W\right](\cdot)$ is the solution to

$$d\bar{X}^{\varepsilon}(t) = b(\bar{X}^{\varepsilon}(t))dt + \sigma(\bar{X}^{\varepsilon}(t))u(t)dt + \sqrt{\varepsilon}\sigma(\bar{X}^{\varepsilon}(t))dW(t), \bar{X}^{\varepsilon}(0) = x_0,$$

Finite dimensional representation and an application

Now becomes convergence of variational problems.

Finite dimensional representation and an application

Now becomes convergence of variational problems. Does

$$\inf_{u \in \mathcal{A}} E\left[F\left(\bar{X}^{\varepsilon}\right) + \frac{1}{2}\int_0^T \|u(t)\|^2 \, dt\right]$$

subject to

$$d\bar{X}^{\varepsilon}(t) = b(\bar{X}^{\varepsilon}(t))dt + \sigma(\bar{X}^{\varepsilon}(t))u(t)dt + \sqrt{\varepsilon}\sigma(\bar{X}^{\varepsilon}(t))dW(t), \bar{X}^{\varepsilon}(0) = x_0,$$

converge to

$$\inf_{u \in L^2[0,T]} \left[F(\phi) + \frac{1}{2}\int_0^T \|u(t)\|^2 \, dt\right]$$

subject to

$$\dot{\phi}(t) = b(\phi(t)) + \sigma(\phi(t))u(t), \quad \phi(0) = x_0?$$

Finite dimensional representation and an application

Now becomes convergence of variational problems. Does

$$\inf_{u \in \mathcal{A}} E\left[F\left(\bar{X}^{\varepsilon}\right) + \frac{1}{2}\int_0^T \|u(t)\|^2 \, dt\right]$$

subject to

$$d\bar{X}^{\varepsilon}(t) = b(\bar{X}^{\varepsilon}(t))dt + \sigma(\bar{X}^{\varepsilon}(t))u(t)dt + \sqrt{\varepsilon}\sigma(\bar{X}^{\varepsilon}(t))dW(t), \bar{X}^{\varepsilon}(0) = x_0,$$

converge to

$$\inf_{u \in L^2[0,T]} \left[F(\phi) + \frac{1}{2}\int_0^T \|u(t)\|^2 \, dt\right]$$

subject to

$$\dot{\phi}(t) = b(\phi(t)) + \sigma(\phi(t))u(t), \quad \phi(0) = x_0?$$

Under appropriate conditions on coefficients, true by LLN (weak convergence).

Finite dimensional representation and an application

A useful simplification. One can argue that the controls $u \in \mathcal{A}$ can be further restricted without loss to the *compact* Polish space

$$L^2_M[0,T] = \left\{ u \in L^2[0,T] : \int_0^T \|u(t)\|^2 \, dt \leq M \right\},$$

where $M < \infty$ depends on F. Tightness of controls then *automatic*.

Finite dimensional representation and an application

A useful simplification. One can argue that the controls $u \in \mathcal{A}$ can be further restricted without loss to the *compact* Polish space

$$L^2_M[0,T] = \left\{ u \in L^2[0,T] : \int_0^T \|u(t)\|^2 \, dt \le M \right\},$$

where $M < \infty$ depends on F. Tightness of controls then *automatic*.

Theorem. Assume (i) strong existence and uniqueness for SDE, (ii) any condition implying tightness of $\left\{ (\bar{X}^\varepsilon, \bar{u}^\varepsilon) \right\}$ when $\bar{u}^\varepsilon \in L^2_M[0,T]$, (iii) continuity of coefficients. Then the LLN limit holds, and hence the LDP is valid.

Finite dimensional representation and an application

Lower bound (LDP upper bound). For any sequence $\{\bar{u}^{\varepsilon}\}$ the pair $\{(\bar{X}^{\varepsilon}, \bar{u}^{\varepsilon})\}$ is tight. If $(\bar{X}, \bar{u})$ is subsequential limit, then w.p.1

$$d\bar{X}(t) = b(\bar{X}(t))dt + \sigma(\bar{X}(t))\bar{u}(t)dt, \quad \bar{X}(0) = x_0.$$

Lower bound (LDP upper bound). For any sequence $\{\bar{u}^{\varepsilon}\}$ the pair $\{(\bar{X}^{\varepsilon}, \bar{u}^{\varepsilon})\}$ is tight. If $(\bar{X}, \bar{u})$ is subsequential limit, then w.p.1

$$d\bar{X}(t) = b(\bar{X}(t))dt + \sigma(\bar{X}(t))\bar{u}(t)dt, \quad \bar{X}(0) = x_0.$$

Assume $\bar{u}^{\varepsilon}$ is within ε of the infimum.

$$\begin{aligned}
&\liminf_{\varepsilon\to 0} \inf_{u\in\mathcal{A}} E\left[F\left(\bar{X}^{\varepsilon}\right) + \frac{1}{2}\int_0^T \|u(t)\|^2\, dt\right] \\
&= \liminf_{\varepsilon\to 0} E\left[F\left(\bar{X}^{\varepsilon}\right) + \frac{1}{2}\int_0^T \|\bar{u}^{\varepsilon}(t)\|^2\, dt\right] \\
&\geq E\left[F\left(\bar{X}\right) + \frac{1}{2}\int_0^T \|\bar{u}(t)\|^2\, dt\right] \\
&\geq \inf_{u\in L^2[0,T]} \left[F(\phi) + \frac{1}{2}\int_0^T \|u(t)\|^2\, dt\right].
\end{aligned}$$

Upper bound (LDP lower bound). Pick $(\phi, \bar{u})$ within $\delta > 0$ of infimum, and use this $\bar{u}$ in the definition of $\bar{X}^{\varepsilon}$.

Finite dimensional representation and an application

Upper bound (LDP lower bound). Pick $(\phi, \bar{u})$ within $\delta > 0$ of infimum, and use this $\bar{u}$ in the definition of $\bar{X}^{\varepsilon}$. Then $(\bar{X}^{\varepsilon}, \bar{u})$ converges in distribution to $(\phi, \bar{u})$ and

$$\begin{aligned}
&\limsup_{\varepsilon \to 0} \inf_{u \in \mathcal{A}} E\left[F\left(\bar{X}^{\varepsilon}\right) + \frac{1}{2}\int_0^T \|u(t)\|^2 \, dt\right] \\
&\quad \leq \limsup_{\varepsilon \to 0} E\left[F\left(\bar{X}^{\varepsilon}\right) + \frac{1}{2}\int_0^T \|\bar{u}(t)\|^2 \, dt\right] \\
&\quad = \left[F\left(\phi\right) + \frac{1}{2}\int_0^T \|\bar{u}(t)\|^2 \, dt\right] \\
&\quad \leq \inf_{u \in L^2[0,T]} \left[F(\phi) + \frac{1}{2}\int_0^T \|u(t)\|^2 \, dt\right] + \delta.
\end{aligned}$$

Finite dimensional representation and an application

Key facts used: (i) the variational representation

$$-\varepsilon \log Ee^{-\frac{1}{\varepsilon}F(X^\varepsilon)}$$
$$= \inf_{u\in\mathcal{A}} E\left[F\left(\mathcal{G}^\varepsilon\left[\int_0^\cdot u(t)dt + \sqrt{\varepsilon}W\right]\right) + \frac{1}{2}\int_0^T \|u(t)\|^2\,dt\right],$$

and

Finite dimensional representation and an application

Key facts used: (i) the variational representation

$$-\varepsilon \log E e^{-\frac{1}{\varepsilon}F(X^{\varepsilon})} = \inf_{u \in \mathcal{A}} E\left[F\left(\mathcal{G}^{\varepsilon}\left[\int_0^{\cdot} u(t)dt + \sqrt{\varepsilon}W\right]\right) + \frac{1}{2}\int_0^T \|u(t)\|^2 dt\right],$$

and (ii) when $\bar{u}^{\varepsilon}$ takes values in the compact set $L^2_M[0,T]$ and converges in distribution to $\bar{u}$, we have a LLN for the small noise controlled processes

$$\mathcal{G}^{\varepsilon}\left[\int_0^{\cdot} \bar{u}^{\varepsilon}(t)dt + \sqrt{\varepsilon}W\right] \rightarrow \mathcal{G}^0\left[\int_0^{\cdot} \bar{u}(t)dt\right]$$

i.e., $\bar{X}^{\varepsilon} \rightarrow \bar{X}$, in distribution.

History of control ideas in large deviations

Use of control ideas in large deviations originates with Fleming, further developed by Fleming-Souganidis.

History of control ideas in large deviations

Use of control ideas in large deviations originates with Fleming, further developed by Fleming-Souganidis.

- Considers F of special form where

$$v^{\varepsilon}(x,t) = -\varepsilon \log E\left[e^{-\frac{1}{\varepsilon}F(X^{\varepsilon})} \middle| X^{\varepsilon}(t) = x \right]$$

  satisfies a nonlinear PDE (e.g., escape probabilities, containment probabilities).
- PDE is a Hamilton-Jacobi-Bellman equation, so has representation as value for control problem.
- Convergence of $v^{\varepsilon}(x_0, 0)$ proved by PDE arguments (in particular *uniqueness of solution to limit PDE*).

History of control ideas in large deviations

Use of control ideas in large deviations originates with Fleming, further developed by Fleming-Souganidis.

- Considers F of special form where

$$v^{\varepsilon}(x,t) = -\varepsilon \log E\left[e^{-\frac{1}{\varepsilon}F(X^{\varepsilon})} \middle| X^{\varepsilon}(t) = x \right]$$

  satisfies a nonlinear PDE (e.g., escape probabilities, containment probabilities).
- PDE is a Hamilton-Jacobi-Bellman equation, so has representation as value for control problem.
- Convergence of $v^{\varepsilon}(x_0, 0)$ proved by PDE arguments (in particular *uniqueness of solution to limit PDE*).

Similar idea behind approach of Feng and Kurtz, convergence of nonlinear semigroups.

- Applicable (and applied) to problems with infinite dimensional state, but again *requires comparison principle for limit*.

Representations for infinite dimensional Brownian motion

In the analogous infinite dimensional setting there are several models for Gaussian white noise:

- Infinite sequence of iid Brownian motions, space $C([0,T]:\mathbb{R}^\infty)$
- Q-Wiener process, space $C([0,T]:H)$, where $(H,\langle\cdot,\cdot\rangle)$ is a Hilbert space and Q is bounded, strictly positive trace class operator on H
- Brownian sheet (integral of space-time white noise), space $C([0,T]\times\mathcal{O}:\mathbb{R})$, $\mathcal{O}\subset\mathbb{R}^d$ bounded and open

Representations can be mapped from one to the other. See [Budhiraja, D 2000] for Q-Wiener process, [Budhiraja, D, Maroulas 2008] for others.

Representations for infinite dimensional Brownian motion

State here case of iid Brownian motions (used later). Topology on $\mathbb{R}^\infty$ consistent with uniform convergence on bounded index sets. Let

$$\beta(t) = (W_1(t), W_2(t), \ldots),$$

and let $\Theta : C([0,T] : \mathbb{R}^\infty) \to \mathbb{R}$ be bounded and measurable.

Representations for infinite dimensional Brownian motion

State here case of iid Brownian motions (used later). Topology on $\mathbb{R}^\infty$ consistent with uniform convergence on bounded index sets. Let

$$\beta(t) = (W_1(t), W_2(t), \ldots),$$

and let $\Theta : C([0,T] : \mathbb{R}^\infty) \to \mathbb{R}$ be bounded and measurable. Then

$$-\varepsilon \log E e^{-\frac{1}{\varepsilon}\Theta(\sqrt{\varepsilon}\beta)} = \inf_{u \in \mathcal{A}} E\left[\Theta\left(\int_0^{\cdot} u(t)dt + \sqrt{\varepsilon}\beta(\cdot)\right) + \frac{1}{2}\int_0^T \|u(t)\|_{\ell^2}^2 \, dt\right],$$

where

$$\mathcal{A} \doteq \left\{u : u \text{ is } \mathcal{F}_t\text{-predictible and } E\int_0^T \|u(t)\|_{\ell^2}^2 \, dt < \infty\right\}.$$

Representations for infinite dimensional Brownian motion

A general LDP for "small noise" models can be formulated in same way as finite dimensional case. Consider the compact sets

$$L^2_M[0,T] = \left\{ u \in L^2[0,T] : \int_0^T \|u(t)\|^2_{\ell^2}\, dt \leq M \right\}.$$

Representations for infinite dimensional Brownian motion

Assume there are measurable maps $\mathcal{G}^{\varepsilon} : C([0,T] : \mathbb{R}^{\infty}) \rightarrow \mathbb{R}$ with the following properties:

Representations for infinite dimensional Brownian motion

Assume there are measurable maps $\mathcal{G}^{\varepsilon} : C([0, T] : \mathbb{R}^{\infty}) \to \mathbb{R}$ with the following properties:

- for every $M < \infty$ the set

$$\left\{ \mathcal{G}^0 \left(\int_0^{\cdot} u(t)dt \right) : \int_0^T \|u(t)\|_{\ell^2}^2 \, dt \leq M \right\}$$

  is compact;

Representations for infinite dimensional Brownian motion

Assume there are measurable maps $\mathcal{G}^{\varepsilon} : C([0,T] : \mathbb{R}^{\infty}) \to \mathbb{R}$ with the following properties:

- for every $M < \infty$ the set

$$\left\{ \mathcal{G}^0 \left(\int_0^{\cdot} u(t)dt \right) : \int_0^T \|u(t)\|_{\ell^2}^2 \, dt \le M \right\}$$

  is compact;

- if $M < \infty$ and $u^{\varepsilon} \in \mathcal{A}$ satisfies $u^{\varepsilon} \in L^2_M[0,T]$ w.p.1 and $u^{\varepsilon} \to u$ in distribution, then

$$\mathcal{G}^{\varepsilon} \left(\int_0^{\cdot} u^{\varepsilon}(t)dt + \sqrt{\varepsilon}\beta(\cdot) \right) \to \mathcal{G}^0 \left(\int_0^{\cdot} u(t)dt \right).$$

Representations for infinite dimensional Brownian motion

Assume there are measurable maps $\mathcal{G}^{\varepsilon} : C([0,T] : \mathbb{R}^{\infty}) \to \mathbb{R}$ with the following properties:

- for every $M < \infty$ the set

$$\left\{ \mathcal{G}^0 \left(\int_0^{\cdot} u(t) dt \right) : \int_0^T \|u(t)\|_{\ell^2}^2 \, dt \leq M \right\}$$

  is compact;
- if $M < \infty$ and $u^{\varepsilon} \in \mathcal{A}$ satisfies $u^{\varepsilon} \in L^2_M[0,T]$ w.p.1 and $u^{\varepsilon} \to u$ in distribution, then

$$\mathcal{G}^{\varepsilon} \left(\int_0^{\cdot} u^{\varepsilon}(t) dt + \sqrt{\varepsilon}\beta(\cdot) \right) \to \mathcal{G}^0 \left(\int_0^{\cdot} u(t) dt \right).$$

Representations for infinite dimensional Brownian motion

Assume there are measurable maps $\mathcal{G}^{\varepsilon} : C([0,T] : \mathbb{R}^{\infty}) \to \mathbb{R}$ with the following properties:

- for every $M < \infty$ the set

$$\left\{ \mathcal{G}^0 \left(\int_0^{\cdot} u(t)dt \right) : \int_0^T \|u(t)\|_{\ell^2}^2 \, dt \leq M \right\}$$

  is compact;
- if $M < \infty$ and $u^{\varepsilon} \in \mathcal{A}$ satisfies $u^{\varepsilon} \in L_M^2[0,T]$ w.p.1 and $u^{\varepsilon} \to u$ in distribution, then

$$\mathcal{G}^{\varepsilon} \left(\int_0^{\cdot} u^{\varepsilon}(t)dt + \sqrt{\varepsilon}\beta(\cdot) \right) \to \mathcal{G}^0 \left(\int_0^{\cdot} u(t)dt \right).$$

Theorem. Under these conditions, $\mathcal{G}^{\varepsilon}\left(\sqrt{\varepsilon}\beta(\cdot)\right)$ satisfies the LDP with rate function $I(\phi) = \inf \left\{ \frac{1}{2} \int_0^T \|u(t)\|_{\ell^2}^2 \, dt : \phi = \mathcal{G}^0 \left(\int_0^{\cdot} u(t)dt \right) \right\}$.

Representations for infinite dimensional Brownian motion

Applications of this result and the variations:

- Stochastic evolution equations with multiplicative noise
- 2D Navier-Stokes with multiplicative noise
- Stochastic tamed 3D Navier-Stokes
- SPDEs with reflection
- SPDEs with varying boundary conditions
- Randomly perturbed Boussinesq equations
- 3D stochastic wave equation
- 2D stochastic hydrodynamical systems
- Stochastic shell models
- Stochastic Ginzburg-Landau equation with multipicative noise
- Stochastic Volterra equations in Banach spaces

Derivation of the representation

Starting point–relative entropy duality formula. Let $\varepsilon = 1$ and μ denote measure of $\beta(\cdot)$ on $\mathcal{E} = C([0, T] : \mathbb{R}^\infty)$. Then for bounded measurable Θ

$$-\log \int_{\mathcal{E}} e^{-\Theta(\phi)} \mu(d\phi) = \inf_{\nu \in \mathcal{P}(\mathcal{E})} \left[\int_{\mathcal{E}} \Theta(\phi) \nu(d\phi) + R\left(\nu \,\|\mu\right) \right],$$

where

$$R\left(\nu \,\|\mu\right) = \int_{\mathcal{E}} \log\left(\frac{d\nu}{d\mu}(\phi)\right) \nu(d\phi) \text{ if } \nu \ll \mu$$

and $R\left(\nu \,\|\mu\right) = \infty$ else.

Derivation of the representation

Starting point–relative entropy duality formula. Let $\varepsilon = 1$ and μ denote measure of $\beta(\cdot)$ on $\mathcal{E} = C([0,T] : \mathbb{R}^{\infty})$. Then for bounded measurable Θ

$$-\log \int_{\mathcal{E}} e^{-\Theta(\phi)} \mu(d\phi) = \inf_{\nu \in \mathcal{P}(\mathcal{E})} \left[\int_{\mathcal{E}} \Theta(\phi) \nu(d\phi) + R(\nu \| \mu) \right],$$

where

$$R(\nu \| \mu) = \int_{\mathcal{E}} \log \left(\frac{d\nu}{d\mu}(\phi) \right) \nu(d\phi) \text{ if } \nu \ll \mu$$

and $R(\nu \| \mu) = \infty$ else. For "nice" control processes u one can consider ν defined by

$$\frac{d\nu}{d\mu} = \exp \left[\int_0^T \left(\langle u, d\beta \rangle_{\ell^2} - \frac{1}{2} \|u\|_{\ell^2}^2 \, dt \right) \right].$$

Derivation of the representation

Then with dP^{ν}/dP as this RN derivative, Girsanov's Theorem gives

$$\begin{aligned}
-\log \int_{\mathcal{E}} e^{-\Theta(\phi)} \mu(d\phi) &\leq \int_{\mathcal{E}} \Theta(\phi)\nu(d\phi) + R(\nu \,\|\mu) \\
&= E^{\nu}\left[\Theta\left(\int_0^{\cdot} u(t)dt + \tilde{\beta}(\cdot)\right) + \frac{1}{2}\int_0^T \|u(t)\|_{\ell^2}^2 \, dt\right]
\end{aligned}$$

Derivation of the representation

Then with dP^{ν}/dP as this RN derivative, Girsanov's Theorem gives

$$\begin{aligned}
-\log \int_{\mathcal{E}} e^{-\Theta(\phi)} \mu(d\phi) &\leq \int_{\mathcal{E}} \Theta(\phi)\nu(d\phi) + R\left(\nu \,\|\mu\right) \\
&= E^{\nu}\left[\Theta\left(\int_0^{\cdot} u(t)dt + \tilde{\beta}(\cdot)\right) + \frac{1}{2}\int_0^T \|u(t)\|_{\ell^2}^2 \, dt\right]
\end{aligned}$$

Also, have equality

$$-\log \int_{\mathcal{E}} e^{-\Theta(\phi)} \mu(d\phi) = \int_{\mathcal{E}} \Theta(\phi)\nu_0(d\phi) + R\left(\nu_0 \,\|\mu\right)$$

for

$$\nu_0(d\phi) \doteq e^{-\Theta(\phi)} \mu(d\phi) \Big/ Z_{\Theta}.$$

Derivation of the representation

Then with dP^{ν}/dP as this RN derivative, Girsanov's Theorem gives

$$\begin{aligned} -\log \int_{\mathcal{E}} e^{-\Theta(\phi)} \mu(d\phi) &\leq \int_{\mathcal{E}} \Theta(\phi)\nu(d\phi) + R\left(\nu \,\|\mu\right) \\ &= E^{\nu}\left[\Theta\left(\int_0^{\cdot} u(t)dt + \tilde{\beta}(\cdot)\right) + \frac{1}{2}\int_0^T \|u(t)\|_{\ell^2}^2 \, dt\right] \end{aligned}$$

Also, have equality

$$-\log \int_{\mathcal{E}} e^{-\Theta(\phi)} \mu(d\phi) = \int_{\mathcal{E}} \Theta(\phi)\nu_0(d\phi) + R\left(\nu_0 \,\|\mu\right)$$

for

$$\nu_0(d\phi) \doteq e^{-\Theta(\phi)} \mu(d\phi) \Big/ Z_{\Theta}.$$

Then for "nice" Θ, $d\, [\nu_0]|_{\mathcal{F}_t} \big/ d\, [\mu]|_{\mathcal{F}_t}$ defines a martingale, and by representation theorem construct control u_0 such that

$$-\log \int_{\mathcal{E}} e^{-\Theta(\phi)} \mu(d\phi) = E^{\nu_0}\left[\Theta\left(\int_0^{\cdot} u_0(t)dt + \tilde{\beta}(\cdot)\right) + \frac{1}{2}\int_0^T \|u_0(t)\|_{\ell^2}^2 \, dt\right].$$

Derivation of the representation

Issues:

- Remove restrictions that infimum only over "nice" processes u and extend from "nice" Θ to all bounded and measurable

Derivation of the representation

Issues:

- Remove restrictions that infimum only over "nice" processes u and extend from "nice" Θ to all bounded and measurable
- Want to rewrite so on single (original) probability space. Given $(u, \tilde{\beta})$ on $(\Omega, \mathcal{F}, P^{\nu})$ must *construct* $(\tilde{u}, \beta)$ on $(\Omega, \mathcal{F}, P)$ with same (joint) distribution.

Derivation of the representation

Issues:

- Remove restrictions that infimum only over "nice" processes u and extend from "nice" Θ to all bounded and measurable
- Want to rewrite so on single (original) probability space. Given $(u, \tilde{\beta})$ on $(\Omega, \mathcal{F}, P^{\nu})$ must *construct* $(\tilde{u}, \beta)$ on $(\Omega, \mathcal{F}, P)$ with same (joint) distribution.

Derivation of the representation

Issues:

- Remove restrictions that infimum only over "nice" processes u and extend from "nice" Θ to all bounded and measurable
- Want to rewrite so on single (original) probability space. Given $(u, \tilde{\beta})$ on $(\Omega, \mathcal{F}, P^{\nu})$ must *construct* $(\tilde{u}, \beta)$ on $(\Omega, \mathcal{F}, P)$ with same (joint) distribution.

Second item requires time discretization, backward dynamic programming, various approximations.

Application to a problem of pattern theory

Problem of interest (from [Budhiraja, D, Maroulas 2010]): given noisy, low dimensional observations, infer a structured but unobserved mapping. Well known Bayesian approach via image registration (or matching) due to Grenander called "deformable templates."

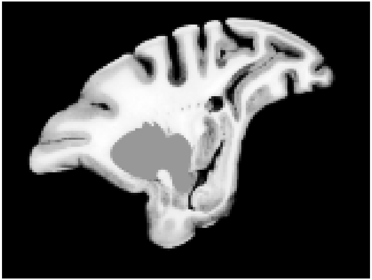
Application to a problem of pattern theory

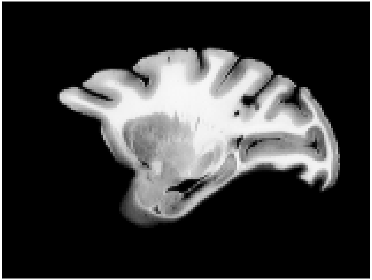
Problem of interest (from [Budhiraja, D, Maroulas 2010]): given noisy, low dimensional observations, infer a structured but unobserved mapping. Well known Bayesian approach via image registration (or matching) due to Grenander called "deformable templates." Ingredients:

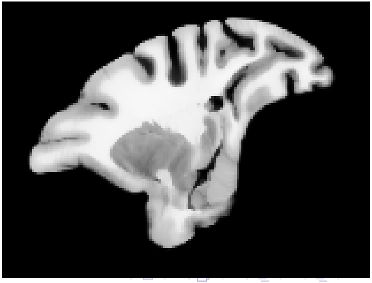
- a "template" $T : \mathcal{O} \subset \mathbb{R}^3 \to \mathbb{R}$ (could also have range $\mathbb{R}^d$), a stored canonical instance of the structured object of interest;
- a random diffeomorphism $h : \mathcal{O} \to \mathcal{O}$, thereby inducing a prior distribution via $T(h(\cdot))$;
- for a partition $\cup_{i \in \mathcal{L}} \mathcal{O}_i = \mathcal{O}$, noisy observations d_i of $\int_{\mathcal{O}_i} T(h(x))dx$.

A typical example: shape and intensity information in computational anatomy.

Application to pattern theory (images courtesy M. Miller)







Application to a problem of pattern theory

Questions: (i) convenient and natural construction of random diffeomorphisms, and (ii) construction and intepretation of an estimator.

Application to a problem of pattern theory

Questions: (i) convenient and natural construction of random diffeomorphisms, and (ii) construction and intepretation of an estimator.

Based on models from continuum mechanics, stochastic flows

$$d\psi_t^{\varepsilon}(x) = \sqrt{\varepsilon}\sum_{j=1}^{\infty}\lambda_j(\psi_t^{\varepsilon}(x))dW_j(t), \quad \psi_0^{\varepsilon}(x) = x,$$

are natural, and $x \to \psi_t^{\varepsilon}(x)$ defines a diffeomorphism with spatial correlation $a(x,y) = \sum_{j=1}^{\infty}\lambda_j(x)\lambda_j(y)$ structure when the λ_j are othonormal in an appropriate HS of functions (requires appropriate smoothness of a).

Application to a problem of pattern theory

Questions: (i) convenient and natural construction of random diffeomorphisms, and (ii) construction and intepretation of an estimator.

Based on models from continuum mechanics, stochastic flows

$$d\psi_t^\varepsilon(x) = \sqrt{\varepsilon}\sum_{j=1}^{\infty} \lambda_j(\psi_t^\varepsilon(x))dW_j(t), \quad \psi_0^\varepsilon(x) = x,$$

are natural, and $x \to \psi_t^\varepsilon(x)$ defines a diffeomorphism with spatial correlation $a(x,y) = \sum_{j=1}^{\infty} \lambda_j(x)\lambda_j(y)$ structure when the λ_j are othonormal in an appropriate HS of functions (requires appropriate smoothness of a). Then $Y^\varepsilon(x) = \psi_1^\varepsilon(x)$ gives the prior, with observations

$$D_i^\varepsilon = \int_{\mathcal{O}_i} T(Y^\varepsilon(x))dx + \sqrt{\varepsilon}\xi_i,$$

and $\xi_i, i \in \mathcal{L}$ independent and normal.

Application to a problem of pattern theory

Applying the theorem one obtains an LDP for $\psi_t^{\varepsilon}(x)$ (and hence $Y^{\varepsilon}(x)$) in an appropriate space of diffeomorphisms, with rate

$$I(h) = \inf \left\{ \frac{1}{2} \int_0^T \|u(t)\|_{\ell^2}^2 \, dt : \dot{h}_t(x) = \sum_{j=1}^{\infty} \lambda_j(h_t(x)) u_j(t), h_0(x) = x \right\}.$$

Application to a problem of pattern theory

Applying the theorem one obtains an LDP for $\psi^{\varepsilon}_t(x)$ (and hence $Y^{\varepsilon}(x)$) in an appropriate space of diffeomorphisms, with rate

$$I(h) = \inf\left\{\frac{1}{2}\int_0^T \|u(t)\|^2_{\ell^2}\, dt : \dot{h}_t(x) = \sum_{j=1}^{\infty} \lambda_j(h_t(x))u_j(t), h_0(x) = x\right\}.$$

One then obtains LDP for the conditional distribution of Y^{ε}, given observation $\{D_i^{\varepsilon} = d_i, i \in \mathcal{L}\}$:

$$J(y) = \hat{J}(y) - \inf_{\hat{y}} \hat{J}(\hat{y}),$$

$$\hat{J}(y) = \inf\left\{I(h) + \frac{1}{2}\sum_{i\in\mathcal{L}}\left|d_i - \int_{\mathcal{O}_i} T(y(x))dx\right|^2 : y(x) = h_1(x)\right\}$$

Interpretation of minimizer as approximate MAP estimator. Coincides with and gives probabilistic interpretation for variational problem introduced in [D,Grenander, Miller 1998].

Some remarks

- Other (not "small noise") applications also of interest, difficult or not depending on difficulty of LLN analysis. E.g., problems with homogenization quite difficult (see [D, Spiliopolous 2012] for some results in finite dimensions). Other applications to analysis of Monte Carlo schemes.

Some remarks

- Other (not "small noise") applications also of interest, difficult or not depending on difficulty of LLN analysis. E.g., problems with homogenization quite difficult (see [D, Spiliopolous 2012] for some results in finite dimensions). Other applications to analysis of Monte Carlo schemes.
- Other important category of continuous time problems: processes with jumps. Topic of talk tomorrow by A. Budhiraja.

Some remarks

- Other (not "small noise") applications also of interest, difficult or not depending on difficulty of LLN analysis. E.g., problems with homogenization quite difficult (see [D, Spiliopolous 2012] for some results in finite dimensions). Other applications to analysis of Monte Carlo schemes.
- Other important category of continuous time problems: processes with jumps. Topic of talk tomorrow by A. Budhiraja.
- Representations for discrete time analogues is some sense simpler, but less elegant. Representations based on chain rule for relative entropy, but less structured than those in continuous time which use "infinitely divisible" driving noises.