Optimal and Learning Control

for

Autonomous Robots

Lecture 11

ADRL

Farbod Farshidian
Agile & Dexterous Robotics Lab

FNSNF

robotics Swiss National Centre of Competence in Research

ETH Zürich

Evaluation!

Please fill in the course evaluation and use the opportunity to make free text comments to give us useful feedback!

Script Erratum

Algorithm 6 ε-soft, On-Policy Monte Carlo Algorithm

choose a constant learning rate, ω
choose a positive $\varepsilon \in (0, 1]$
$Q^\pi(x,u) \leftarrow$ arbitrary
$\pi \leftarrow$ an arbitrary ε-soft policy

Repeat forever:

(a) generate an episode using π

(b) Policy Evaluation

for each pair (x,u) appearing in the episode
$R \leftarrow$ return following the first occurrence of (x,u)
$Q^\pi(x,u) \leftarrow Q^\pi(x,u) + \omega\left(R - Q^\pi(x,u)\right)$

(c) Policy Improvement

for each: x in the episode:
$u^* \leftarrow \arg\max_u Q^\pi(x,u)$
For all $a \in \mathcal{U}(x)$:

$$\pi(x,u) \leftarrow \begin{cases} \frac{\varepsilon}{|\mathcal{U}(x)|} & \text{if } u \neq u^* \\ 1 - \varepsilon\left(1 - \frac{1}{|\mathcal{U}(x)|}\right) & \text{if } u = u^* \end{cases}$$

(d) (*optional*) decrease ε.

Recap

Brownian Motion

It is stochastic process.

$$\mathbb{P}_{\mathbf{w}}(t,w) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(w-\mu t)^2}{2\sigma^2 t}\right)$$

$$\mathbb{E}\{w(t)\} = \mu t$$

$$\mathbb{V}ar\{w(t)\} = \sigma^2 t$$

Brownian Motion (cnt)

$$dw(t) = \lim_{\Delta t \to 0} w(t + \Delta t) - w(t)$$

1. The increment process, $dw(t)$, has a Gaussian distribution with the mean and the variance, $\mu \Delta t$ and $\sigma^2 \Delta t$ respectively.
2. The increment process, $dw(t)$, is statistically independent of $w(s)$ for any $s \leq t$.

Stochastic Differential Equation

$$d\mathbf{x} = \mathbf{f}(t, \mathbf{x})dt + \mathbf{g}(t, \mathbf{x})d\mathbf{w}$$

Drift Coefficient

Diffusion Coefficient

Brownian Motion
$\mathcal{N}(\mathbf{0}, \mathbf{I}dt)$

The conditional PDF is Gaussian

$$\mathbb{P}_{\mathbf{x}}(t + \Delta t, \mathbf{x} \mid t, \mathbf{y}) = \mathcal{N}\Big(\mathbf{y} + \mathbf{f}(t, \mathbf{y})\Delta t, \mathbf{g}(t, \mathbf{y})\mathbf{g}^T(t, \mathbf{y})\Delta t\Big)$$

Fokker Planck Equation

- Extracting samples: SDE

$$d\mathbf{x} = \mathbf{f}(t,\mathbf{x})dt + \mathbf{g}(t,\mathbf{x})d\mathbf{w}, \qquad \mathcal{N}(\mathbf{0}, \mathbf{I}dt)$$

- The PDF of process: Fokker Planck equation

$$\mathbb{P}_{\mathbf{x(t)}}(t, \mathbf{x} \mid s, \mathbf{y})$$

$$\partial_t \mathbb{P} = -\nabla_x^T(\mathbf{f}\mathbb{P}) + \frac{1}{2}\mathrm{Tr}\left[\nabla_{xx}(\mathbf{g}\mathbf{g}^T\mathbb{P})\right]$$

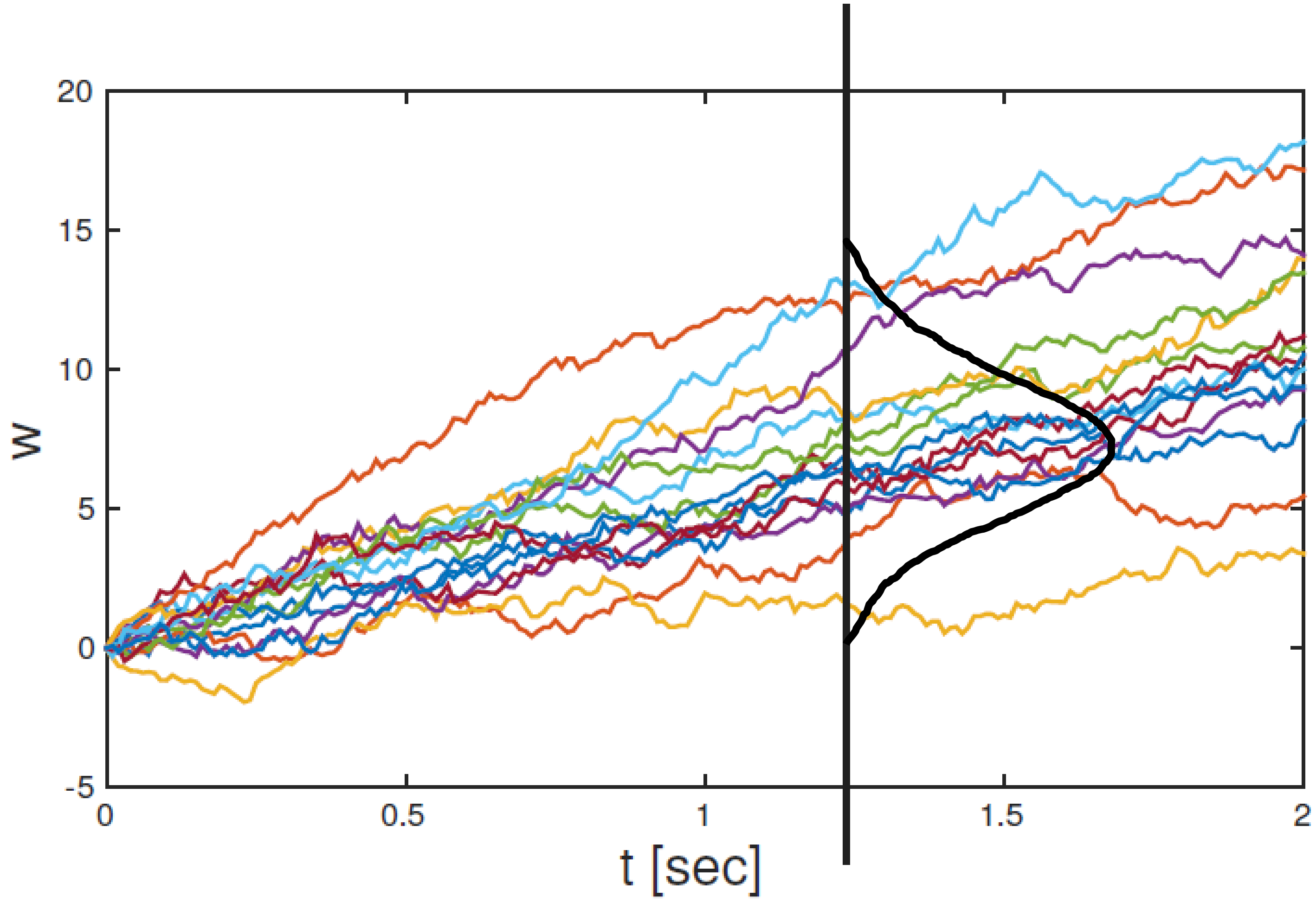
Fokker Planck Eq.

$$\mathbb{P}_{\mathbf{x(t)}}(t = s, \mathbf{x} \mid s, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$$

Initial Condition

The effective covariance

Fokker Planck Equation (cnt)



Linear Markov Decision Process

Three conditions on the optimal control problem:

1) Quadratic control cost

$$J = E\left\{\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x}) + \frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u}\, dt\right\}$$

2) Control affine system

$$d\mathbf{x} = \mathbf{f}(t,\mathbf{x})dt + \mathbf{g}(t,\mathbf{x})\left(\mathbf{u}dt + d\mathbf{w}\right), \qquad d\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}dt)$$

$$\dot{\mathbf{x}} = \mathbf{f}(t,\mathbf{x}) + \mathbf{g}(t,\mathbf{x})\left(\mathbf{u} + \varepsilon\right), \qquad \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$$

3) $\mathbf{R}\mathbf{\Sigma} = \lambda\mathbf{I}$

Linear Markov Decision Process (cnt)

nonlinear PDE

$$-\partial_t V^* = q - \frac{1}{2}\nabla_x^T V^* \, \Xi \nabla_x V^* + \nabla_x^T V^* \, \mathbf{f} + \frac{\lambda}{2}\mathrm{Tr}[\nabla_{xx} V^* \, \Xi]$$

$$V^*(t, \mathbf{x}) = -\lambda \log \Psi(t, \mathbf{x})$$

$$-\partial_t \Psi = -\frac{1}{\lambda} q \Psi + \mathbf{f}^T \nabla_x \Psi + \frac{\lambda}{2}\mathrm{Tr}[\Xi \nabla_{xx} \Psi]$$

$$-\partial_t \Psi = \mathrm{H}[\Psi] \qquad \mathrm{H} = -\frac{1}{\lambda} q + \mathbf{f}^T \nabla_x + \frac{\lambda}{2}\mathrm{Tr}[\Xi \nabla_{xx}]$$

$$\Psi(t_f, \mathbf{x}) = \exp\left(-\frac{1}{\lambda}\Phi(\mathbf{x})\right)$$

Final Value problem

$$\mathbf{g} \boldsymbol{\Sigma} \mathbf{g}^T = \lambda \Xi$$

The effective Covariance

Integral by Parts

$$d(f\,g) = df\,g + f\,dg$$

$$\int_{-\infty}^{+\infty} f(x)\,g^1(x)dx = fg(+\infty) - fg(-\infty) - \int_{-\infty}^{+\infty} f^1(x)\,g(x)dx$$

$$\lim_{|x|\to\infty} g(x) = 0$$

$$= -\int_{-\infty}^{+\infty} f^1(x)\,g(x)dx$$

In general case:

$$\int_{-\infty}^{+\infty} f(x)\,g^i(x)dx = (-1)^i \int_{-\infty}^{+\infty} f^i(x)\,g(x)dx$$

ADRL

ETH Zürich

Path Integral Optimal Control

Function Inner Product

- Inner product for two vectors

$$< \mathbf{u} \mid \mathbf{v} >= \mathbf{u}^T \mathbf{v} = \sum_i u_i v_i$$

- Inner product for two functions

$$< f \mid g >= \int_{-\infty}^{\infty} f(x) g(x) \, dx$$

Function Inner Product (cnt.)

- Hermitian Conjugate operator ($\mathbf{H}^\dagger$) of a linear operator $\mathbf{H}$

$$< \mathbf{u} \mid \mathbf{H}\mathbf{v} > = < \mathbf{H}^\dagger \mathbf{u} \mid \mathbf{v} >$$

$$\mathbf{u}^T(\mathbf{H}\mathbf{v}) = (\mathbf{H}^\dagger \mathbf{u})^T \mathbf{v}$$

$$\boxed{\mathbf{H}^\dagger = \mathbf{H}^T}$$

- In the function space

$$< f \mid \mathbf{H}g > = < \mathbf{H}^\dagger f \mid g >$$

$$\int_{-\infty}^{\infty} f(x)\,\mathbf{H}g(x)\,dx = \int_{-\infty}^{\infty} \mathbf{H}^\dagger f(x)\,g(x)\,dx$$

Path Integral: Inner Product

- Assume the following inner product

$$< \rho \mid \Psi >= \int \rho(t, \mathbf{x}) \Psi(t, \mathbf{x}) d\mathbf{x}$$

where Ψ is the Desirability function,

and ρ is an arbitrary function which satisfies:

$$\lim_{\|x\| \to \infty} \rho(t, \mathbf{x}) = 0$$

Path Integral: Inner Product (cnt)

- Assume the linear operator introduced by the Fokker Planck equation

$$\mathtt{H} = -\frac{1}{\lambda}q + \mathbf{f}^T\nabla_x + \frac{\lambda}{2}\mathtt{Tr}[\Xi\nabla_{xx}]$$
$$= -\frac{1}{\lambda}q + \sum_i \mathbf{f}_i\frac{\partial}{\partial_{x_i}} + \frac{\lambda}{2}\sum_{i,j}\Xi_{ij}\frac{\partial^2}{\partial_{x_i}\partial_{x_j}}$$

What is Hermitian Conjugate of “H” in the function space?

Path Integral: Inner Product (cnt)

According to the Hermitian Conjugate definition:

$$< \rho \mid \mathrm{H}[\Psi] > = < \mathrm{H}^{\dagger}[\rho] \mid \Psi >$$

By using integral by parts:

$$\mathrm{H}^{\dagger} = -\frac{1}{\lambda} q - \sum_{i} \frac{\partial \mathbf{f}_i}{\partial_{x_i}} + \frac{\lambda}{2} \sum_{i,j} \frac{\partial^2 \mathbf{\Xi}_{ij}}{\partial_{x_i} \partial_{x_j}}$$

$$= -\frac{1}{\lambda} q - \nabla_x^T \mathbf{f} + \frac{\lambda}{2} \mathrm{Tr}[\nabla_{xx} \mathbf{\Xi}]$$

Path Integral: Inner Product (cnt)

Summary:

$$< \rho \mid \mathtt{H}[\Psi] > = < \mathtt{H}^{\dagger}[\rho] \mid \Psi >$$

$$\mathtt{H} = -\frac{1}{\lambda} q + \mathbf{f}^T \nabla_x + \frac{\lambda}{2} \mathtt{Tr}[\mathbf{\Xi} \nabla_{xx}]$$

$$\mathtt{H}^{\dagger} = -\frac{1}{\lambda} q - \nabla_x^T \mathbf{f} + \frac{\lambda}{2} \mathtt{Tr}[\nabla_{xx} \mathbf{\Xi}]$$

A D R L

ETH Zürich

ρ Function

- General idea: if ρ satisfies the following

$$\frac{d}{dt} < \rho \mid \Psi >= 0$$

1) ρ can be a solution to an initial value problem $\rho(t = s, \mathbf{x})$

2) The following equality holds

$$< \rho \mid \Psi > (t = s) =< \rho \mid \Psi > (t = t_f)$$

ρ Function (cnt)

Starting with: $\frac{d}{dt} < \rho \mid \Psi >= 0$

$$
\begin{aligned}
0 &= \frac{d}{dt} < \rho \mid \Psi > \\
&= \int \partial_t \Big(\rho(t, \mathbf{x}) \Psi(t, \mathbf{x}) \Big) d\mathbf{x} \\
&= \int \partial_t \rho(t, \mathbf{x}) \Psi(t, \mathbf{x}) + \rho(t, \mathbf{x}) \partial_t \Psi(t, \mathbf{x}) d\mathbf{x} \\
&= < \partial_t \rho \mid \Psi > + < \rho \mid \partial_t \Psi >
\end{aligned}
$$

It satisfies the LMDP

$-\partial_t \Psi = \mathtt{H}[\Psi]$

ρ Function (cnt)

$$0 =< \partial_t \rho \mid \Psi > - < \rho \mid \mathtt{H}[\Psi] >$$

Using the Hermitian Conjugate operator

$$0 =< \partial_t \rho \mid \Psi > - < \mathtt{H}^\dagger[\rho] \mid \Psi >$$

$$< \partial_t \rho - \mathtt{H}^\dagger[\rho] \mid \Psi >= 0$$

A trivial solution is:

$$\partial_t \rho = \mathtt{H}^\dagger[\rho]$$

$$= -\frac{1}{\lambda} q \rho - \nabla_x^T (\mathbf{f} \rho) + \frac{\lambda}{2} \mathtt{Tr}[\nabla_{xx}(\Xi \rho)]$$

Comparison with Fokker Planck

$$\partial_t \mathbb{P} = -\nabla_x^T (\mathbf{f}\mathbb{P}) + \frac{1}{2}\mathrm{Tr}\left[\nabla_{xx}(\mathbf{g}\mathbf{g}^T\mathbb{P})\right]$$

$$\partial_t \rho = -\frac{1}{\lambda} q\rho - \nabla_x^T (\mathbf{f}\rho) + \frac{\lambda}{2}\mathtt{Tr}[\nabla_{xx}(\boldsymbol{\Xi}\rho)]$$

It attenuates the probability distribution over time.

Comparison with Fokker Planck (cnt)

$$\partial_t \mathbb{P} = -\nabla_x^T (\mathbf{f}\mathbb{P}) + \frac{1}{2}\mathrm{Tr}\left[\nabla_{xx}(\mathbf{g}\mathbf{g}^T\mathbb{P})\right]$$

$$\mathbb{P}_{\mathbf{x(t)}}(t = s, \mathbf{x} \mid s, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$$

The initial condition

$$d\mathbf{x} = \mathbf{f}(t, \mathbf{x})dt + \mathbf{g}(t, \mathbf{x})d\mathbf{w}, \qquad \mathcal{N}(\mathbf{0}, \mathbf{I}dt)$$

$$\mathbf{x}(s) = \mathbf{y}$$

This can be used to extract samples

Comparison with Fokker Planck (cnt)

- An initial condition:

$$\partial_t \rho = -\frac{1}{\lambda} q\rho - \nabla_x^T(\mathbf{f}\rho) + \frac{\lambda}{2}\mathtt{Tr}[\nabla_{xx}(\Xi\rho)]$$

$$\rho(t = s, \mathbf{x}) = \delta(\mathbf{x} - \mathbf{y})$$

- A method to numerically simulate the solution:

$$d\mathbf{x}(t_i) = \mathbf{f}(t_i, \mathbf{x}(t_i))dt + \mathbf{g}(t_i, \mathbf{x}(t_i))d\mathbf{w}, \qquad \mathbf{x}(t_0 = s) = \mathbf{y} \qquad d\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}dt)$$

$$\begin{cases} \mathbf{x}(t_{i+1}) = \mathbf{x}(t_i) + d\mathbf{x}(t_i) & \text{with probability } \exp\left(-\frac{1}{\lambda}qdt\right) \\ \mathbf{x}(t_{i+1}) : \text{annihilation} & \text{with probability } 1 - \exp\left(-\frac{1}{\lambda}qdt\right) \end{cases}$$

ρ Function : Features

- It is a MDP: $\tau = \{\mathbf{x}(t_0), \mathbf{x}(t_1), \ldots \mathbf{x}(t_N)\}$

$$\rho(\tau \mid s, \mathbf{y}) = \prod_{i=0}^{N-1} \rho(t_{i+1}, \mathbf{x}(t_{i+1}) \mid t_i, \mathbf{x}(t_i)), \qquad \mathbf{x}(t_0 = s) = \mathbf{y}$$

- The conditioned probability(!!) is

$$\rho(t_{i+1}, \mathbf{x}(t_{i+1}) \mid t_i, \mathbf{x}(t_i)) = \mathrm{e}^{-\frac{1}{\lambda} q(t_i, \mathbf{x}(t_i)) dt} \mathcal{N}\Big(\mathbf{x}(t_i) + \mathbf{f}(t_i, \mathbf{x}(t_i)) dt, \boldsymbol{\Xi}(t_i, \mathbf{x}(t_i)) dt\Big)$$

The probability of keeping the sample

Trajectory PDF

- Trajectory joint probability distribution

$$\rho(\tau \mid s, \mathbf{y}) = \prod_{i=0}^{N-1} \mathrm{e}^{-\frac{1}{\lambda} q(t_i, \mathbf{x}(t_i)) dt} \mathcal{N}\Big(\mathbf{x}(t_i) + \mathbf{f}(t_i, \mathbf{x}(t_i))dt, \boldsymbol{\Xi}(t_i, \mathbf{x}(t_i))dt\Big)$$

$$= \prod_{i=0}^{N-1} \mathcal{N}\Big(\mathbf{x}(t_i) + \mathbf{f}(t_i, \mathbf{x}(t_i))dt, \boldsymbol{\Xi}(t_i, \mathbf{x}(t_i))dt\Big) \; \mathrm{e}^{\sum_{i=0}^{N-1} -\frac{1}{\lambda} q(t_i, \mathbf{x}(t_i)) dt}$$

$$= \mathbb{P}_{uc}(\tau \mid s, \mathbf{y}) \; \mathrm{e}^{\sum_{i=0}^{N-1} -\frac{1}{\lambda} q(t_i, \mathbf{x}(t_i)) dt}$$

where $\mathbb{P}_{uc}$ is the uncontrolled system trajectory PDF.

$$d\mathbf{x}(t_i) = \mathbf{f}(t_i, \mathbf{x}(t_i))dt + \mathbf{g}(t_i, \mathbf{x}(t_i))d\mathbf{w}, \qquad \mathbf{x}(t_0 = s) = \mathbf{y} \quad d\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma} dt)$$

A Single State PDF

Marginalize the trajectory joint PDF

$$\tau = \{\mathbf{x}(t_0), \mathbf{x}(t_1) \ldots, \mathbf{x}(t_n)\}$$

Sub-trajectory

$$\rho(\tau \mid s, \mathbf{y}) = \mathbb{P}_{uc}(\tau \mid s, \mathbf{y}) \ \mathrm{e}^{\sum_{i=0}^{n-1} -\frac{1}{\lambda} q(t_i, \mathbf{x}(t_i)) dt}$$

Sub-trajectory PDF

$$\rho(\mathbf{x}(t_n) \mid s, \mathbf{y}) = \int \mathbb{P}_{uc}(\tau \mid s, \mathbf{y}) \ \mathrm{e}^{\sum_{i=0}^{n-1} -\frac{1}{\lambda} q(t_i, \mathbf{x}(t_i)) dt} d\mathbf{x}(t_1) \ldots d\mathbf{x}(t_{n-1})$$

ADRL

ETH Zürich

ρ Function

1) ρ should be a solution to an initial value problem

$$\rho(\mathbf{x}(t_n) \mid s, \mathbf{y})$$

$$\rho(t = s, \mathbf{x}) = \delta(\mathbf{x} - \mathbf{y})$$

2) The following equality holds

$$< \rho \mid \Psi > (t = s) =< \rho \mid \Psi > (t = t_f)$$

Time Invariant Inner Product

- Equating the inner product at time s and t_f

$$< \rho \mid \Psi > (t = s) =< \rho \mid \Psi > (t = t_f)$$

$$\int \rho(s, \mathbf{x}_0)\Psi(s, \mathbf{x}_0)d\mathbf{x}_0 = \int \rho(t_f, \mathbf{x}_N)\Psi(t_f, \mathbf{x}_N)d\mathbf{x}_N$$

using the initial condition for ρ

$$\int \delta(\mathbf{x}_0 - \mathbf{y})\Psi(s, \mathbf{x}_0)d\mathbf{x}_0 = \int \rho(t_f, \mathbf{x}_N)\Psi(t_f, \mathbf{x}_N)d\mathbf{x}_N$$

$$\Psi(s, \mathbf{y}) = \int \rho(t_f, \mathbf{x}_N)\Psi(t_f, \mathbf{x}_N)d\mathbf{x}_N$$

ADRL

ETH Zürich

Time Invariant Inner Product (cnt)

using the terminal condition for Ψ

$$\Psi(s, \mathbf{y}) = \int \rho(t_f, \mathbf{x}_N) \Psi(t_f, \mathbf{x}_N) d\mathbf{x}_N$$

$$\Psi(s, \mathbf{y}) = \int \rho(t_f, \mathbf{x}_N) \, \mathrm{e}^{-\frac{1}{\lambda}\Phi(\mathbf{x}_N)} \, d\mathbf{x}_N$$

We know the PDF of a single state

$$\rho(t_f, \mathbf{x}_N) = \int \rho(\tau \mid s, \mathbf{y}) \, d\mathbf{x}(t_1) \ldots \mathbf{x}(t_{N-1})$$

$$= \int \mathbb{P}_{uc}(\tau \mid s, \mathbf{y}) \;\; \mathrm{e}^{\sum_{i=0}^{N-1} -\frac{1}{\lambda} q(t_i, \mathbf{x}(t_i)) dt} \;\; d\mathbf{x}(t_1) \ldots d\mathbf{x}(t_{N-1})$$

ADRL

ETH Zürich

Path Integral

$$\Psi(s,\mathbf{y}) = \int \mathbb{P}_{uc}(\tau \mid s,\mathbf{y}) \; \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}_N)+\sum_{i=0}^{N-1} q(t_i,\mathbf{x}(t_i))dt\right)} \; d\mathbf{x}(t_1)\dots d\mathbf{x}(t_{N-1}) d\mathbf{x}_N$$

Equivalently

$$\Psi(s,\mathbf{y}) = \mathrm{E}_{\tau_{uc}}\left\{ \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_N))+\sum_{i=0}^{N-1} q(t_i,\mathbf{x}(t_i))dt\right)} \right\}$$

$$= \mathrm{E}_{\tau_{uc}}\left\{ \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f))+\int_{t_0}^{t_f} q(t,\mathbf{x}) \; dt\right)} \right\}$$

Samples can be generated by

$$d\mathbf{x} = \mathbf{f}(t,\mathbf{x})dt + \mathbf{g}(t,\mathbf{x})d\mathbf{w}, \qquad d\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}dt), \quad \mathbf{x}(t=s) = \mathbf{y}$$

ADRL

ETH Zürich

Closer look at Path Integral formula

- For calculating the Desirability function at each point

$$\Psi(s, \mathbf{y}) = \mathrm{E}_{\tau_{uc}} \left\{ \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x})\ dt\right)} \right\}$$

$$d\mathbf{x} = \mathbf{f}(t, \mathbf{x})dt + \mathbf{g}(t, \mathbf{x})d\mathbf{w}, \qquad d\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}dt), \quad \mathbf{x}(t = s) = \mathbf{y}$$

1) Forward simulate the uncontrolled system from $(s, \mathbf{y})$ up to t_f
2) Integrate the cost over the generated path

Path Integral: Optimal Control

- Directly calculating the optimal control

$$\mathbf{u}^*(s,\mathbf{y}) = -\mathbf{R}^{-1}\mathbf{g}^T(s,\mathbf{y})\nabla_y V^*(s,\mathbf{y})$$
$$= \lambda\mathbf{R}^{-1}\mathbf{g}^T(s,\mathbf{y})\frac{\nabla_y \Psi(s,\mathbf{y})}{\Psi(s,\mathbf{y})}$$

After a tedious calculation

$$\mathbf{u}^*(s,\mathbf{y}) = \lim_{\Delta s \to 0} \frac{\mathrm{E}_{\tau_{uc}}\left\{\int_s^{s+\Delta s} d\mathbf{w}\ \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x})\ dt\right)}\right\}}{\Delta s\, \mathrm{E}_{\tau_{uc}}\left\{\mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x})\ dt\right)}\right\}}$$

$$d\mathbf{x} = \mathbf{f}(t,\mathbf{x})dt + \mathbf{g}(t,\mathbf{x})d\mathbf{w}, \qquad d\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}dt), \quad \mathbf{x}(t=s) = \mathbf{y}$$

Path Integral: Optimal Control

- Using the white noise formulation $\varepsilon = \frac{d\mathbf{w}}{dt}$

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}) + \mathbf{g}(t, \mathbf{x})\varepsilon, \qquad \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{x}(t = s) = \mathbf{y}$$

$$\mathbf{u}^*(s, \mathbf{y}) = \frac{\mathrm{E}_{\tau_{uc}}\left\{\varepsilon \;\; \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x}) \;\, dt\right)}\right\}}{\mathrm{E}_{\tau_{uc}}\left\{\mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x}) \;\, dt\right)}\right\}}$$

Path Integral: issues (1)

- Inefficient sampling

$$\mathbf{u}^*(s,\mathbf{y}) = \mathrm{E}_{\tau_{uc}}\left\{ \varepsilon \frac{e^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x})\ dt\right)}}{\mathrm{E}_{\tau_{uc}}\left\{ e^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x})\ dt\right)}\right\}}\right\}$$

Soft Max

It just has significant value for near optimal solution

What are the chances to hit the optimal solution by a random walk?

Importance Sampling

Path Integral: issues (2)

- Point-wise estimation of the optimal controls

$$\mathbf{u}^*(s,\mathbf{y}) = \mathrm{E}_{\tau_{uc}}\left\{ \varepsilon \frac{\mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x})\ dt\right)}}{\mathrm{E}_{\tau_{uc}}\left\{\mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x})\ dt\right)}\right\}} \right\}$$

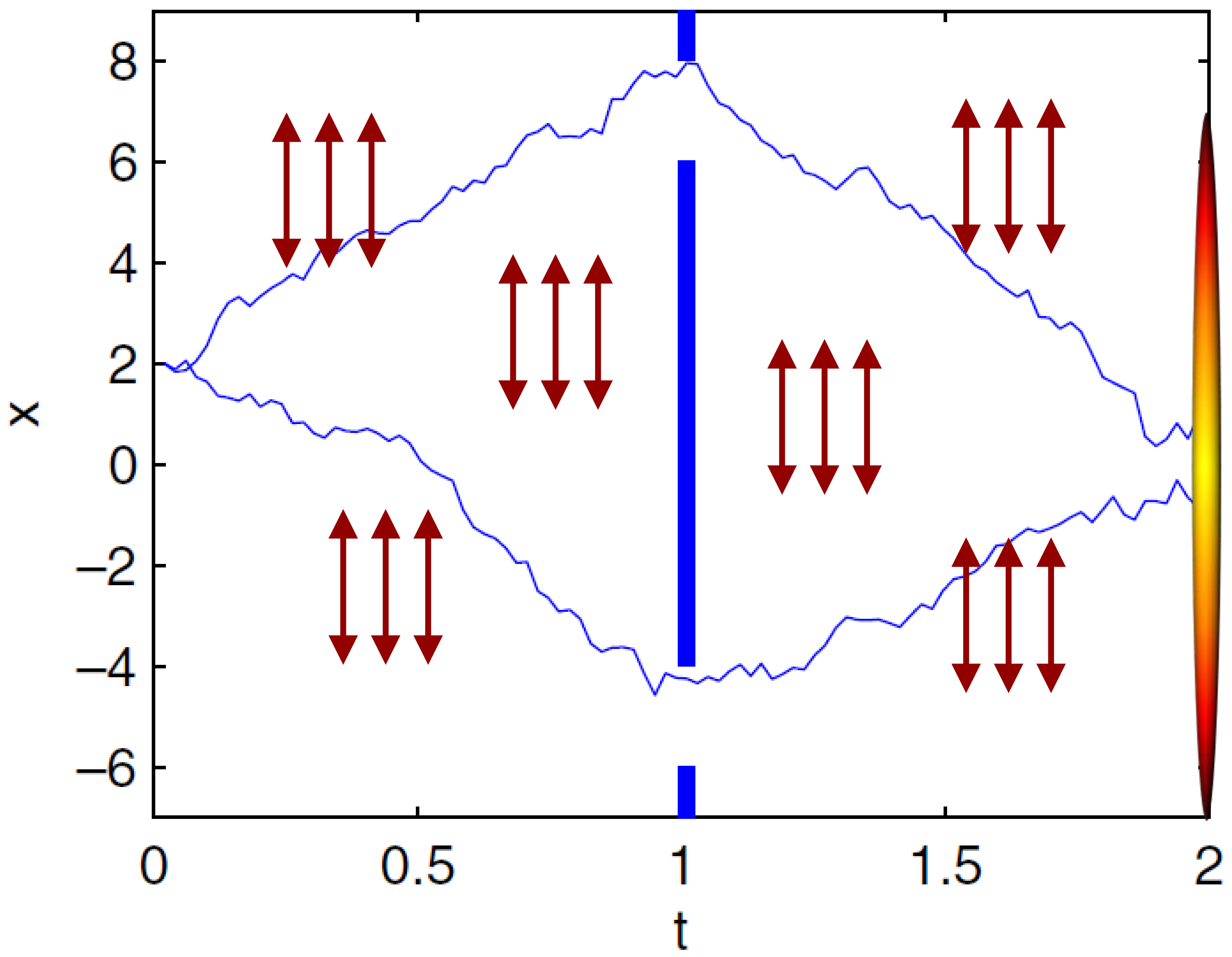
The optimal control is estimated independently for each point

Does the optimal control change drastically from one point to the other?

Function Approximation

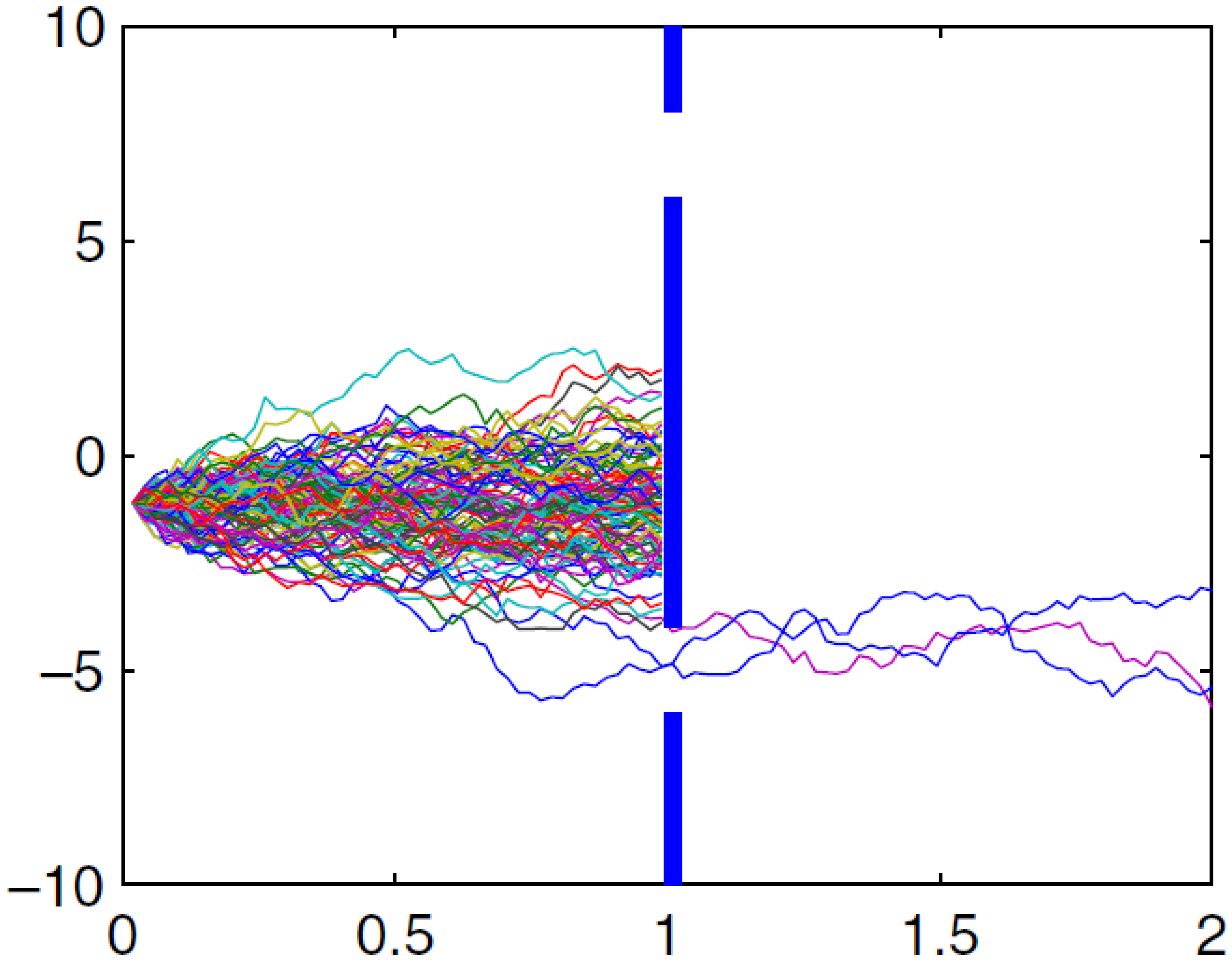
Importance Sampling: Example

Double-slit problem



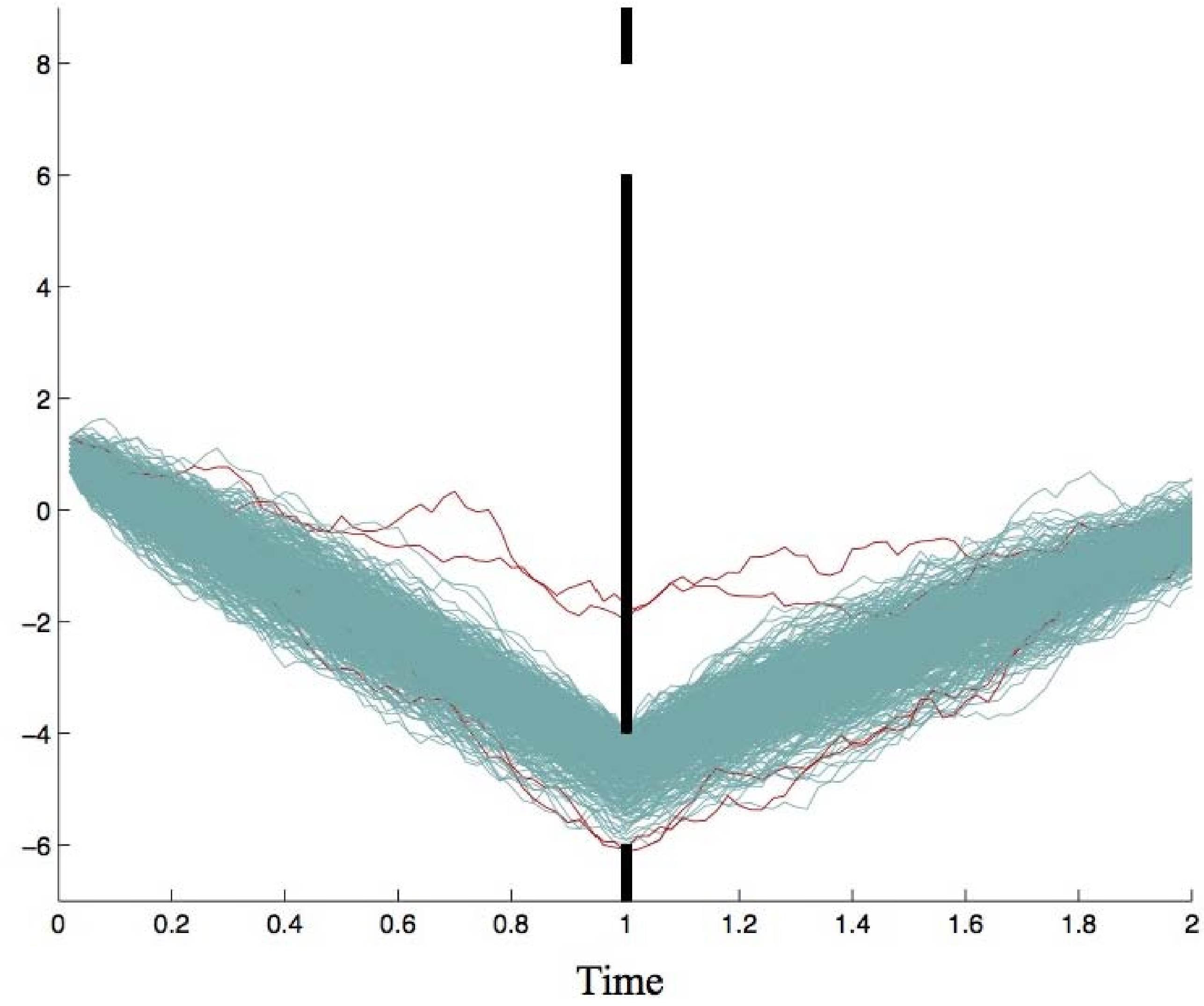
Importance Sampling: Example (cnt)

The original Path Integral sampling approach:



Importance Sampling: Example (cnt)

We would have a better sampling efficiency, if we could have biased the sampling towards each of the slits!

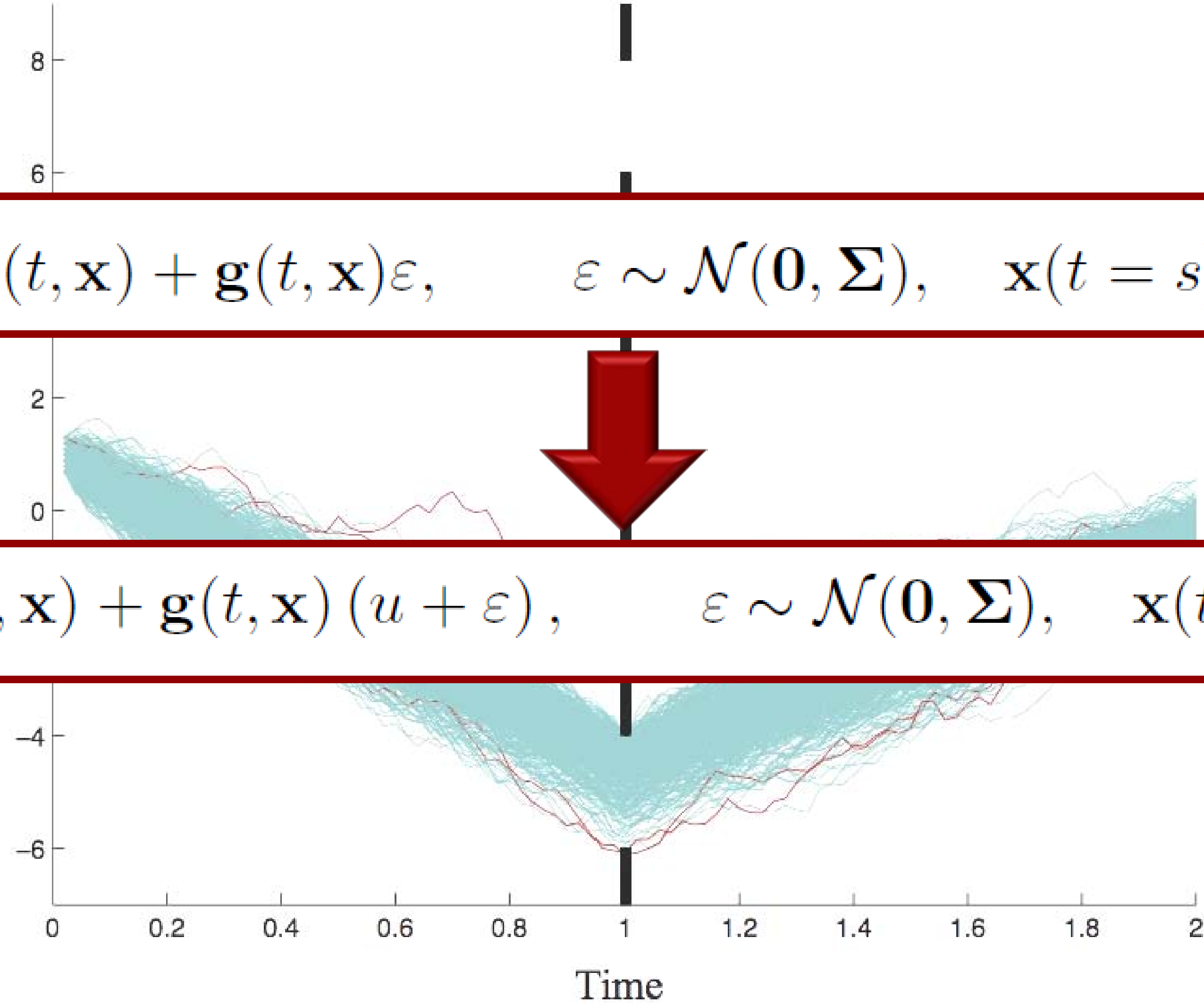


Importance Sampling: Example (cnt)

We would have a better sampling efficiency, if we could have biased the sampling towards each of the slits!

$$\dot{\mathbf{x}} = \mathbf{f}(t,\mathbf{x}) + \mathbf{g}(t,\mathbf{x})\varepsilon, \qquad \varepsilon \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma}), \quad \mathbf{x}(t=s)=\mathbf{y}$$

$$\dot{\mathbf{x}} = \mathbf{f}(t,\mathbf{x}) + \mathbf{g}(t,\mathbf{x})\left(u+\varepsilon\right), \qquad \varepsilon \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma}), \quad \mathbf{x}(t=s)=\mathbf{y}$$



Importance Sampling: Motivations

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}) + \mathbf{g}(t, \mathbf{x})\,(u + \varepsilon)\,, \qquad \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{x}(t = s) = \mathbf{y}$$

1) We have an initial guess about the optimal solution

2) We want to improve the controller incrementally

Importance Sampling: Introduction

Assume the following expectation problem where x is a random variable with probability distribution $p(x)$ and $f(x)$ is an arbitrary deterministic function.

$$E_p[f(x)] = \int_{-\infty}^{\infty} f(x)\ p(x)dx$$

We will assume that we have another random variable named y with the probability distribution $q(y)$ and Lets assume that calculating the expectation of an arbitrary function for this random variable is less costly than the previous one.

$$E_p[f(x)] = E_q[w(y)f(y)], \qquad w(y) = \frac{p(y)}{q(y)}$$

$$\begin{aligned} E_q[w(y)f(y)] &= \int_{-\infty}^{\infty} w(y)f(y)q(y)\ dy \\ &= \int_{-\infty}^{\infty} \frac{p(y)}{\cancel{q(y)}} f(y)\cancel{q(y)}\ dy \\ &= \int_{-\infty}^{\infty} p(y)f(y)\ dy = E_p[f(x)] \end{aligned}$$

The key is to multiply by the importance weight!

Path Integral: Importance Sampling

$$\mathbf{u}^*(s,\mathbf{y}) = \frac{\mathrm{E}_{\tau_{uc}}\left\{ \varepsilon \;\; \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x})\; dt\right)} \right\}}{\mathrm{E}_{\tau_{uc}}\left\{ \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x})\; dt\right)} \right\}}$$

~~$\dot{\mathbf{x}} = \mathbf{f}(t,\mathbf{x}) + \mathbf{g}(t,\mathbf{x})\varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}), \quad \mathbf{x}(t=s) = \mathbf{y}$~~

$$\dot{\mathbf{x}} = \mathbf{f}(t,\mathbf{x}) + \mathbf{g}(t,\mathbf{x})\left(u + \varepsilon\right), \qquad \varepsilon \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}), \quad \mathbf{x}(t=s) = \mathbf{y}$$

The importance weight: $\dfrac{\mathbb{P}_{uc}(\tau \mid s,\mathbf{y})}{\mathbb{P}_{c}(\tau \mid s,\mathbf{y})}$

Path Integral: Importance Sampling

$$\frac{\mathbb{P}_{uc}(\tau \mid s, \mathbf{y})}{\mathbb{P}_c(\tau \mid s, \mathbf{y})} = \mathrm{e}^{-\frac{1}{\lambda}\int_{t_0}^{t_f} \frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u}dt + \mathbf{u}^T\mathbf{R}d\mathbf{w}}$$

$$\mathbf{u}^*(s, \mathbf{y}) = \frac{\mathrm{E}_{\tau_{uc}}\left\{\varepsilon \;\; \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x}) \; dt\right)}\right\}}{\mathrm{E}_{\tau_{uc}}\left\{\mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x}) \; dt\right)}\right\}}$$

$$\mathrm{E}_{\tau_c}\left\{\mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x}) + \frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u} \; dt + \mathbf{u}^T\mathbf{R}d\mathbf{w}\right)}\right\}$$

Path Integral: Importance Sampling

$$\frac{\mathbb{P}_{uc}(\tau \mid s, \mathbf{y})}{\mathbb{P}_{c}(\tau \mid s, \mathbf{y})} = \mathrm{e}^{-\frac{1}{\lambda} \int_{t_0}^{t_f} \frac{1}{2}\mathbf{u}^T \mathbf{R} \mathbf{u} dt + \mathbf{u}^T \mathbf{R} d\mathbf{w}}$$

$$\mathbf{u}^*(s, \mathbf{y}) = \frac{\mathrm{E}_{\tau_{uc}} \left\{ \varepsilon \;\; \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x}) \; dt\right)} \right\}}{\mathrm{E}_{\tau_{uc}} \left\{ \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x}) \; dt\right)} \right\}}$$

$$\mathrm{E}_{\tau_c} \left\{ (\mathbf{u} + \varepsilon) \;\; \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} q(t,\mathbf{x}) + \frac{1}{2}\mathbf{u}^T \mathbf{R} \mathbf{u} \; dt + \mathbf{u}^T \mathbf{R} d\mathbf{w}\right)} \right\}$$

Path Integral: Importance Sampling

$$\mathbf{u}^*(s,\mathbf{y}) = \frac{\mathrm{E}_{\tau_c}\left\{(\mathbf{u}+\varepsilon)\ \mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f))+\int_{t_0}^{t_f} q(t,\mathbf{x})+\frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u}\ dt+\mathbf{u}^T\mathbf{R}d\mathbf{w}\right)}\right\}}{\mathrm{E}_{\tau_c}\left\{\mathrm{e}^{-\frac{1}{\lambda}\left(\Phi(\mathbf{x}(t_f))+\int_{t_0}^{t_f} q(t,\mathbf{x})+\frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u}\ dt+\mathbf{u}^T\mathbf{R}d\mathbf{w}\right)}\right\}}$$

$$R(\tau;s,\mathbf{y}) = \Phi(\mathbf{x}(t_f)) + \int_s^{t_f}\left(q(t,\mathbf{x})+\frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u}\right)dt + \int_s^{t_f}\mathbf{u}^T\mathbf{R}d\mathbf{w}$$

It is actually the Return, we have previously used in the RL section!

$$J = E[R(\tau;t_0,\mathbf{x}_0)]$$

$$\boxed{\mathbf{u}^*(s,\mathbf{y}) = \mathbf{u}(s,\mathbf{y}) + \frac{\mathrm{E}_{\tau_c}\left\{\varepsilon\ \mathrm{e}^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}{\mathrm{E}_{\tau_c}\left\{\mathrm{e}^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}}$$

ADRL

ETH Zürich

Path Integral: IS proof

The means are different!

$$\frac{\mathbb{P}_{uc}(\tau \mid s, \mathbf{y})}{\mathbb{P}_c(\tau \mid s, \mathbf{y})} = \frac{\prod_{i=0}^{N-1} \mathcal{N}\Big(\mathbf{x}(t_i) + \mathbf{f}(t_i, \mathbf{x}(t_i))dt, \boldsymbol{\Xi}(t_i, \mathbf{x}(t_i))dt\Big)}{\prod_{i=0}^{N-1} \mathcal{N}\Big(\mathbf{x}(t_i) + \mathbf{f}(t_i, \mathbf{x}(t_i))dt + \mathbf{g}(t_i, \mathbf{x}(t_i))\mathbf{u}(t_i)dt, \boldsymbol{\Xi}(t_i, \mathbf{x}(t_i))dt\Big)}$$

$$\frac{\mathbb{P}_{uc}(\tau \mid s, \mathbf{y})}{\mathbb{P}_c(\tau \mid s, \mathbf{y})} = \prod_{i=0}^{N-1} \frac{\exp\Big(-\frac{1}{2}\|\mathbf{x}_{i+1} - \mathbf{x}_i - \mathbf{f}_i dt\|^2_{\boldsymbol{\Xi}_i dt}\Big)}{\exp\Big(-\frac{1}{2}\|\mathbf{x}_{i+1} - \mathbf{x}_i - \mathbf{f}_i dt - \mathbf{g}_i\mathbf{u}_i dt\|^2_{\boldsymbol{\Xi}_i dt}\Big)}$$

$$\frac{\mathbb{P}_{uc}(\tau \mid s, \mathbf{y})}{\mathbb{P}_c(\tau \mid s, \mathbf{y})} = e^{-\frac{1}{\lambda}\int_{t_0}^{t_f} \frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u}dt + \mathbf{u}^T\mathbf{R}d\mathbf{w}}$$

Path Integral: IS summery

$$\mathbf{u}^*(s,\mathbf{y}) = \mathbf{u}(s,\mathbf{y}) + \frac{\mathrm{E}_{\tau_c}\left\{\varepsilon \;\; \mathrm{e}^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}{\mathrm{E}_{\tau_c}\left\{\mathrm{e}^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}$$

Optimal Control

$$\dot{\mathbf{x}} = \mathbf{f}(t,\mathbf{x}) + \mathbf{g}(t,\mathbf{x})\,(u+\varepsilon)\,, \qquad \varepsilon \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma}), \quad \mathbf{x}(t=s) = \mathbf{y}$$

Sampling System

$$R(\tau;s,\mathbf{y}) = \Phi(\mathbf{x}(t_f)) + \int_s^{t_f}\left(q(t,\mathbf{x}) + \frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u}\right)dt + \int_s^{t_f}\mathbf{u}^T\mathbf{R}d\mathbf{w}$$

Return: integral of the cost over path

ADRL

ETH Zürich

Function Approximation

Motivation

Function Approximation (cnt)

- Approximating the optimal control with a **Linear Model** (linear w.r.t. to the parameters)

$$u_i^*(s,\mathbf{y}) = \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i^* + error$$

Basis Function: nonlinear function of time and state

Parameter Vector: approximation parameter

Error: approximation error

A D R L

ETH Zürich

Function Approximation (cnt)

- Approximation needs to have a criterion.

$$\boldsymbol{\theta}_i^* = \underset{\theta_i}{\operatorname{argmax}}\, L(\boldsymbol{\theta}_i)$$
$$= \underset{\boldsymbol{\theta}_i}{\operatorname{argmax}} \int_{t_0}^{t_f} \int_{\Omega} \frac{1}{2} \| u_i^*(s, \mathbf{y}) - \boldsymbol{\Upsilon}_i^T(s, \mathbf{y}) \boldsymbol{\theta}_i \|_2^2 \, p(s, \mathbf{y}) d\mathbf{y} ds$$

  - Mean Square Error (MSE) criterion
  - $\int_{t_0}^{t_f} \int_{\Omega} p(s, \mathbf{y}) d\mathbf{y} ds = 1$

Path Integral: Function Approximation

We have two optimization problems:

1) The Optimal Control problem with the solution

$$\mathbf{u}^*(s,\mathbf{y}) = \mathbf{u}(s,\mathbf{y}) + \frac{\mathrm{E}_{\tau_c}\left\{\varepsilon \;\; \mathrm{e}^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}{\mathrm{E}_{\tau_c}\left\{\mathrm{e}^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}$$

2) The Function Approximation problem

$$\boldsymbol{\theta}_i^* = \underset{\boldsymbol{\theta}_i}{\operatorname{argmax}} \int_{t_0}^{t_f} \int_{\Omega} \frac{1}{2} \| u_i^*(s,\mathbf{y}) - \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i \|_2^2 \; p(s,\mathbf{y}) d\mathbf{y} ds$$

Path Integral: Function Approximation (cnt)

We can define these optimization problems as a single optimization problem.

$$u_i^*(s,\mathbf{y}) \approx \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i^*$$

Approximated Optimal Control

$$\boldsymbol{\theta}_i^* = \boldsymbol{\theta}_{i,c} + \underset{\Delta\boldsymbol{\theta}_i}{\operatorname{argmin}} \int \frac{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}}{E_{\tau_c}\left\{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}} \|\boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\Delta\boldsymbol{\theta}_i - \varepsilon\|_2^2 \, \mathbb{P}_{\tau_c}(\tau \mid s,\mathbf{y}) p(s,\mathbf{y}) d\tau d\mathbf{y} ds$$

Linear Regression

$$\dot{\mathbf{x}} = \mathbf{f}(t,\mathbf{x}) + \mathbf{g}(t,\mathbf{x})(u+\varepsilon), \qquad \varepsilon \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}), \quad \mathbf{x}(t=s) = \mathbf{y}$$

$$u_i(s,\mathbf{y}) \approx \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_{i,c}$$

Sampling System

Path Integral: FA proof

$$\boldsymbol{\theta}_i^* = \underset{\boldsymbol{\theta}_i}{\operatorname{argmax}} \int_{t_0}^{t_f} \int_\Omega \frac{1}{2} \| u_i^*(s,\mathbf{y}) - \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i \|_2^2 \; p(s,\mathbf{y}) d\mathbf{y} ds$$

Function Approximation problem

$$\frac{\partial L(\boldsymbol{\theta}_i^*)}{\partial \boldsymbol{\theta}_i} = \int_{t_0}^{t_f} \int_\Omega \left(u_i^*(s,\mathbf{y}) - \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i^* \right) \boldsymbol{\Upsilon}_i(s,\mathbf{y}) \; p(s,\mathbf{y}) d\mathbf{y} ds = 0$$

$$\int_{t_0}^{t_f} \int_\Omega \frac{\mathrm{E}_{\tau_c}\left\{ \mathrm{e}^{-\frac{1}{\lambda} R(\tau;s,\mathbf{y})} \right\}}{\mathrm{E}_{\tau_c}\left\{ \mathrm{e}^{-\frac{1}{\lambda} R(\tau;s,\mathbf{y})} \right\}} \left(u_i^*(s,\mathbf{y}) - \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i^* \right) \boldsymbol{\Upsilon}_i(s,\mathbf{y}) \; p(s,\mathbf{y}) d\mathbf{y} ds = 0$$

$$\int_{t_0}^{t_f} \int_\Omega \mathrm{E}_{\tau_c} \left\{ \frac{\mathrm{e}^{-\frac{1}{\lambda} R(\tau;s,\mathbf{y})}}{\mathrm{E}_{\tau_c}\left\{ \mathrm{e}^{-\frac{1}{\lambda} R(\tau;s,\mathbf{y})} \right\}} \left(u_i^*(s,\mathbf{y}) - \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i^* \right) \boldsymbol{\Upsilon}_i(s,\mathbf{y}) \; p(s,\mathbf{y}) \right\} d\mathbf{y} ds = 0$$

A D R L

ETH Zürich

Path Integral: FA proof (cnt)

$$\int \frac{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}}{E_{\tau_c}\left\{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}\Big(u_i^*(s,\mathbf{y}) - \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i^*\Big)\boldsymbol{\Upsilon}_i(s,\mathbf{y})\ \mathbb{P}_{\tau_c}(\tau \mid s,\mathbf{y})p(s,\mathbf{y})d\tau d\mathbf{y}ds = 0$$

$$\int \frac{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}}{E_{\tau_c}\left\{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}\Big(\underbrace{u_i^* - u_i - \varepsilon}\, + u_i + \varepsilon - \boldsymbol{\Upsilon}_i^T\boldsymbol{\theta}_i^*\Big)\boldsymbol{\Upsilon}_i(s,\mathbf{y})\ \mathbb{P}_{\tau_c}(\tau \mid s,\mathbf{y})p(s,\mathbf{y})d\tau d\mathbf{y}ds = 0$$

For the first three terms.

Getting the integral w.r.t. trajectory

$$\int \frac{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}}{E_{\tau_c}\left\{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}\Big(u_i^*(s,\mathbf{y}) - u_i(s,\mathbf{y}) - \varepsilon\Big)\boldsymbol{\Upsilon}_i(s,\mathbf{y})\ \mathbb{P}_{\tau_c}(\tau \mid s,\mathbf{y})p(s,\mathbf{y})d\tau d\mathbf{y}ds =$$

$$\int \left(u_i^*(s,\mathbf{y}) - u_i(s,\mathbf{y}) - \frac{E_{\tau_c}\left\{\varepsilon\, e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}{E_{\tau_c}\left\{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}\right)\boldsymbol{\Upsilon}_i(s,\mathbf{y})\ (\tau \mid s,\mathbf{y})p(s,\mathbf{y})d\mathbf{y}ds = 0$$

Path Integral: FA proof (cnt)

$$\int \frac{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}}{E_{\tau_c}\left\{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}\left(u_i(s,\mathbf{y}) + \varepsilon - \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i^*\right)\boldsymbol{\Upsilon}_i(s,\mathbf{y})\ \mathbb{P}_{\tau_c}(\tau \mid s,\mathbf{y})p(s,\mathbf{y})d\tau d\mathbf{y}ds = 0$$

It is equivalent to the following optimization

$$\boldsymbol{\theta}_i^* = \underset{\boldsymbol{\theta}_i}{\operatorname{argmin}} \int \frac{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}}{E_{\tau_c}\left\{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}\|\boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_i - u_i(s,\mathbf{y}) - \varepsilon\|_2^2\ \mathbb{P}_{\tau_c}(\tau \mid s,\mathbf{y})p(s,\mathbf{y})d\tau d\mathbf{y}ds$$

If we use the same function approximation for $u_i(s,\mathbf{y}) \approx \boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\boldsymbol{\theta}_{i,c}$

$$\boldsymbol{\theta}_i^* = \boldsymbol{\theta}_{i,c} + \underset{\Delta\boldsymbol{\theta}_i}{\operatorname{argmin}} \int \frac{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}}{E_{\tau_c}\left\{e^{-\frac{1}{\lambda}R(\tau;s,\mathbf{y})}\right\}}\|\boldsymbol{\Upsilon}_i^T(s,\mathbf{y})\Delta\boldsymbol{\theta}_i - \varepsilon\|_2^2\ \mathbb{P}_{\tau_c}(\tau \mid s,\mathbf{y})p(s,\mathbf{y})d\tau d\mathbf{y}ds$$

Thanks!