Model Predictive Control and Reinforcement Learning
– Lecture 10: Model Predictive Path Integral (MPPI) Control –

Hannes Homburger and Jasper Hoffmann

HTWG Konstanz and University Freiburg

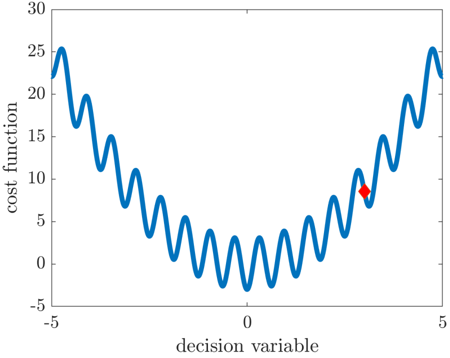
October 12, 2023

It could be so easy...

Scalar Nonlinear Non-Convex Problem

$$\min_{x \in \mathbb{R}} f(x)$$

with initial guess x_0



It could be so easy...

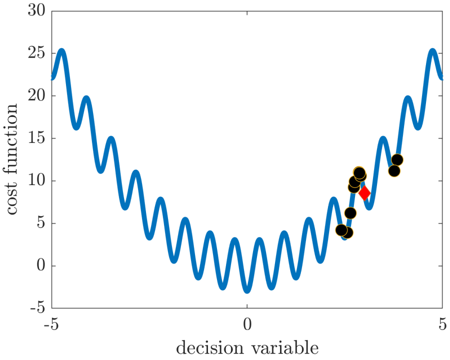
Scalar Nonlinear Non-Convex Problem

$$\min_{x \in \mathbb{R}} f(x)$$

with initial guess x_0

Idea of Random Search

1. Sample around initial guess
2.
3.
4.



It could be so easy...

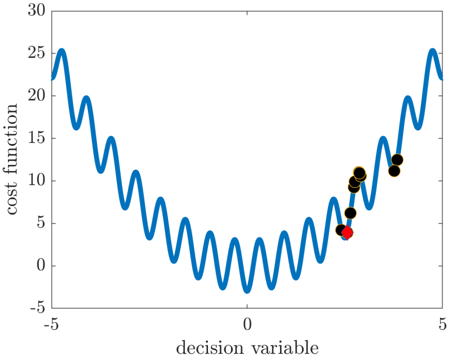
Scalar Nonlinear Non-Convex Problem

$$\min_{x \in \mathbb{R}} f(x)$$

with initial guess x_0

Idea of Random Search

1. Sample around initial guess
2. Evaluate costs and pick best
3.
4.



It could be so easy...

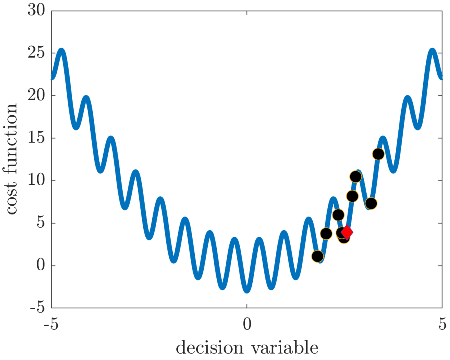
Scalar Nonlinear Non-Convex Problem

$$\min_{x \in \mathbb{R}} f(x)$$

with initial guess x_0

Idea of Random Search

1. Sample around initial guess
2. Evaluate costs and pick best
3. New initial guess $\leftarrow$ best sample
4. Repeat



It could be so easy...

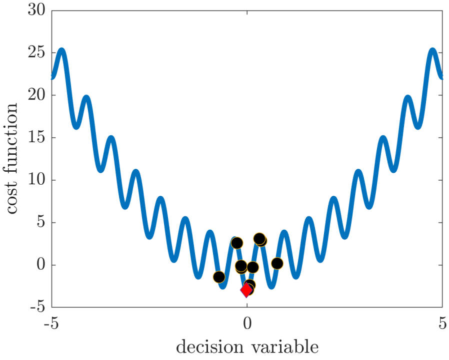
Scalar Nonlinear Non-Convex Problem

$$\min_{x \in \mathbb{R}} f(x)$$

with initial guess x_0

Idea of Random Search

1. Sample around initial guess
2. Evaluate costs and pick best
3. New initial guess $\leftarrow$ best sample
4. Repeat



... but there is no lunch for free!

Overview

Path Integral Control - Selected Literature

Part 1 - Path Integral Control:

- H. J. Kappen, Linear theory for control of nonlinear stochastic systems, Physical Review Letters, Vol. 95, No. 20, 2005, Paper 200201. doi:10.1103/PhysRevLett.95.200201

Part 2 - MPPI:

- G. Williams, A. Aldrich, and E. A. Theodorou, Model predictive path integral control: From theory to parallel computation. In Journal of Guidance, Control, and Dynamics, Vol. 40, No. 2, 2017. doi:10.2514/1.G001921
- G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, Information theoretic model predictive control: Theory and applications to autonomous driving. In IEEE Transactions on Robotics, Vol. 34, No. 6, 2018. doi:10.1109/TRO.2018.2865891

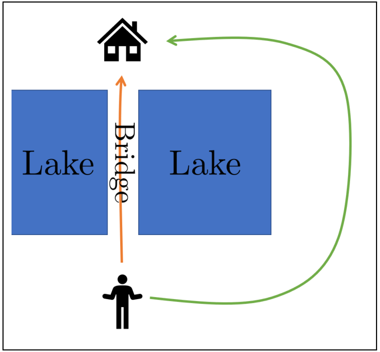
Part 3 - Classification, Applications, and Literature:

- Overview slide at the end

The optimal way to get home

Would you choose the bridge or go around the lake?

- The solution of some optimal control problems depends strongly on the influence of stochastic properties



How can we address a stochastic optimal control problem using path integral control?

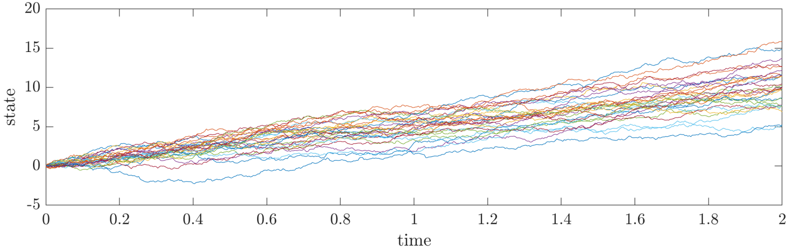
Path Integral Control - Assumptions

Dynamics

Let $s_t = s(t) \in \mathbb{R}^{n_s}$ denote the state, $a_t = a(t) \in \mathbb{R}^{n_a}$ denote the action, and considering a control-affine, stochastic differential equation (SDE) of the form

$$\mathrm{d}s_t = [f(s_t) + G(s_t)a_t]\mathrm{d}t + B(s_t)\mathrm{d}w,$$

where $\mathrm{d}w$ denotes an n_w dimensional Wiener process.



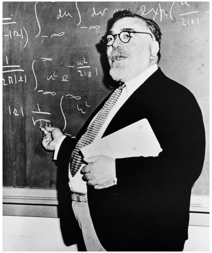
Definition Wiener Process

Durret (1996), Stochastic Calculus - A Practical Introduction, CRC Press

Definition Wiener process

The Wiener process w_t is is characterized by the following properties:

1. $w_0 = 0$ (almost surely)
2. w has independent increments
3. w has Gaussian increments: $w_{t+\Delta t} - w_t \sim \mathcal{N}(0, \Delta t)$
4. w has continuous path in t (almost surely)



Norbert Wiener[1]
(1894-1964)

[1] The references for all images can be found on the last slide.

Another Perspective on the Wiener Process (Donsker's theorem)

Wiener process as a limit of random walk

Let $\xi_1, \xi_2, ..., \xi_{\lfloor nt \rfloor}$ be i.i.d.[1] random variables with mean 0 and variance 1. The random step function

$$w_t^n = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \xi_k,$$

has the property $\lim_{n\to\infty} w_{t+\Delta t}^n - w_t^n \sim \mathcal{N}(0, \Delta t)$, and thereby approaches a Wiener process.

[1] i.i.d. means independent and identically distributed

Path Integral Control - Assumptions

Dynamics

Let $s_t = s(t) \in \mathbb{R}^{n_s}$ denote the state, $a_t = a(t) \in \mathbb{R}^{n_a}$ denote the action, and considering a control-affine, stochastic differential equation (SDE) of the form

$$\mathrm{d}s_t = [f(s_t) + G(s_t)a_t]\mathrm{d}t + B(s_t)\mathrm{d}w, \tag{1}$$

where $\mathrm{d}w$ denotes an n_w dimensional Wiener process.

Costs

Considering a quadratic action cost and an arbitrary state-dependent cost in form of

$$L(s_t, a_t) = q(s_t) + \frac{1}{2} a_t^\top R(s_t) a_t, \tag{2}$$

where $R(s_t) \succ 0$. The terminal cost function is denoted by $E(s_T)$.

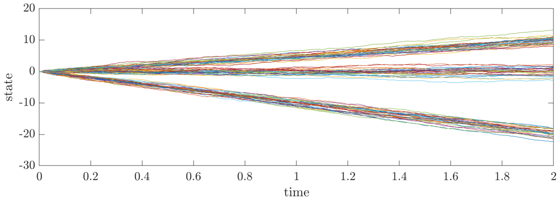
Value Function

Value Function

The value function corresponding to the dynamics (1) and the costs (2) is defined as

$$V(s_t, t) = \min_{a_{(\cdot)}(s)} \mathbb{E}_{s(\cdot)\sim\mathbb{Q}_{a(\cdot)}} \left[E(s_T) + \int_t^T \left(q(s_\tau) + \frac{1}{2} a_\tau(s_\tau)^\top R(s_\tau) a_\tau(s_\tau) \right) \mathrm{d}\tau \right],$$

where $\mathbb{E}_{s(\cdot)\sim\mathbb{Q}_{a(\cdot)}}[\cdot]$ denotes the expectation over trajectories taken with respect to dynamics (1) applying the policy $a_{(\cdot)}(s)$.



Which policy is optimal?

Path Integral Control - Stochastic Hamilton-Jacobi-Bellman Equation

Recursive Formulation of Value Function

Using the principle of optimality, the value function can be expressed as

$$V(s_t, t) = \min_a \mathbb{E}_{\mathrm{d}s_t \sim \mathbb{Q}_a} \left[L(s_t, a)\mathrm{d}t + V(s_t + \mathrm{d}s_t, t + \mathrm{d}t) \right]$$

with boundary condition $V(s_T, T) = E(s_T)$

Stochastic HJB Equation

For given dynamics and costs the value function is given by the solution of the PDE

$$-\frac{\partial V}{\partial t}(s_t, t) = \min_a \left(q(s_t) + \frac{1}{2} a^\top R(s_t) a + [f(s_t) + G(s_t)a]^\top V_s + \frac{1}{2} \mathrm{tr}(B(s_t)B(s_t)^\top V_{ss}) \right)$$

with boundary condition $V(s_T, T) = E(s_T)$, gradient $V_s = \frac{\partial V}{\partial s}(s_t, t)^\top$, and hessian $V_{ss} = \frac{\partial V_s}{\partial s}$

Path Integral Control - Optimal Controls

- Analyzing the Stochastic HJB equation

$$-\frac{\partial V}{\partial t}(s_t, t) = \min_a \left(q(s_t) + \frac{1}{2} a^\top R(s_t) a + [f(s_t) + G(s_t) a]^\top V_s + \frac{1}{2} \text{tr}(B(s_t) B(s_t)^\top V_{ss}) \right)$$

$\rightarrow$ Only the red parts depend on a

Path Integral Control - Optimal Controls

- Analyzing the Stochastic HJB equation

$$-\frac{\partial V}{\partial t}(s_t, t) = \min_a \left(q(s_t) + \frac{1}{2} a^\top R(s_t) a + [f(s_t) + G(s_t) a]^\top V_s + \frac{1}{2} \mathrm{tr}(B(s_t) B(s_t)^\top V_{ss}) \right)$$

$\rightarrow$ Only the red parts depend on a

- An unconstrained quadratic problem with optimal action

$$a^*(s_t, t) = -R(s_t)^{-1} G(s_t)^\top V_s(s_t, t)$$

Path Integral Control - Optimal Controls

- Analyzing the Stochastic HJB equation

$$-\frac{\partial V}{\partial t}(s_t, t) = \min_a \left(q(s_t) + \frac{1}{2} a^\top R(s_t) a + [f(s_t) + G(s_t) a]^\top V_s + \frac{1}{2} \text{tr}(B(s_t) B(s_t)^\top V_{ss}) \right)$$

$\rightarrow$ Only the red parts depend on a

- An unconstrained quadratic problem with optimal action

$$a^*(s_t, t) = -R(s_t)^{-1} G(s_t)^\top V_s(s_t, t)$$

- Reinstate the solution yields nonlinear PDE

$$-\frac{\partial V}{\partial t}(s_t, t) = q(s_t) + f(s_t)^\top V_s - \frac{1}{2} V_s^\top G(s_t) R(s_t)^{-1} G(s_t)^\top V_s + \frac{1}{2} \text{tr}(B(s_t) B(s_t)^\top V_{ss})$$

with boundary condition $V(s_T, T) = E(s_T)$.

Path Integral Control - Recap so far

- Value function can be calculated by solving a special nonlinear backward-in-time PDE
- Classical methods to solve this PDE suffer from curse of dimensionality

Path Integral Control - Recap so far

- Value function can be calculated by solving a special nonlinear backward-in-time PDE
- Classical methods to solve this PDE suffer from curse of dimensionality

Question

Is there a better possibility to solve this PDE by exploiting its structure?

Path Integral Control - Desirability Function

- In controls, transformations are often used to simplify problems - let's do this!

Path Integral Control - Desirability Function

- In controls, transformations are often used to simplify problems - let's do this!
- Using the exponential transform of the value function $\rightarrow \Psi$ a.k.a. *desirability function*.

$$V(s_t, t) = -\lambda \log(\Psi(s_t, t)),$$

where $\lambda > 0$ is a parameter

Path Integral Control - Desirability Function

- In controls, transformations are often used to simplify problems - let's do this!
- Using the exponential transform of the value function $\rightarrow \Psi$ a.k.a. *desirability function*.

$$V(s_t, t) = -\lambda \log(\Psi(s_t, t)),$$

where $\lambda > 0$ is a parameter, the partial derivatives are given by

$$\partial_t V = -\frac{\lambda}{\Psi}\partial_t \Psi \quad \text{and} \quad V_s = -\frac{\lambda}{\Psi}\Psi_s$$

and the hessian

$$V_{ss} = -\frac{\lambda}{\Psi}\Psi_{ss} + \frac{\lambda}{\Psi^2}\Psi_s\Psi_s^\top$$

Path Integral Control - Transformation in Linear PDE

Substituting $\boxed{\partial_t V = -\frac{\lambda}{\Psi}\partial_t \Psi}$, $\boxed{V_s = -\frac{\lambda}{\Psi}\Psi_s}$, and $\boxed{V_{ss} = -\frac{\lambda}{\Psi}\Psi_{ss} + \frac{\lambda}{\Psi^2}\Psi_s\Psi_s^\top}$ in

$$-\partial_t V = q(s_t) + f(s_t)^\top V_s - \frac{1}{2}{V_s}^\top G(s_t)R(s_t)^{-1}G(s_t)^\top V_s + \frac{1}{2}\text{tr}(B(s_t)B(s_t)^\top V_{ss})$$

Path Integral Control - Transformation in Linear PDE

Substituting $\boxed{\partial_t V = -\frac{\lambda}{\Psi}\partial_t \Psi}$, $\boxed{V_s = -\frac{\lambda}{\Psi}\Psi_s}$, and $\boxed{V_{ss} = -\frac{\lambda}{\Psi}\Psi_{ss} + \frac{\lambda}{\Psi^2}\Psi_s\Psi_s^\top}$ in

$$-\partial_t V = q(s_t) + f(s_t)^\top V_s - \frac{1}{2} V_s{}^\top G(s_t) R(s_t)^{-1} G(s_t)^\top V_s + \frac{1}{2}\text{tr}(B(s_t)B(s_t)^\top V_{ss})$$

and omitting now all arguments yields

$$-\left(-\frac{\lambda}{\Psi}\partial_t \Psi\right) = q + f^\top\left(-\frac{\lambda}{\Psi}\Psi_s\right) - \frac{1}{2}\left(-\frac{\lambda}{\Psi}\Psi_s\right)^\top G R^{-1} G^\top \left(-\frac{\lambda}{\Psi}\Psi_s\right)$$
$$+ \frac{1}{2}\text{tr}\left(BB^\top\left(-\frac{\lambda}{\Psi}\Psi_{ss} + \frac{\lambda}{\Psi^2}\Psi_s\Psi_s^\top\right)\right)$$

Path Integral Control - Transformation in Linear PDE

Substituting $\partial_t V = -\frac{\lambda}{\Psi}\partial_t \Psi$, $V_s = -\frac{\lambda}{\Psi}\Psi_s$, and $V_{ss} = -\frac{\lambda}{\Psi}\Psi_{ss} + \frac{\lambda}{\Psi^2}\Psi_s\Psi_s^\top$ in

$$-\partial_t V = q(s_t) + f(s_t)^\top V_s - \frac{1}{2}V_s{}^\top G(s_t)R(s_t)^{-1}G(s_t)^\top V_s + \frac{1}{2}\text{tr}(B(s_t)B(s_t)^\top V_{ss})$$

and omitting now all arguments yields

$$-\left(-\frac{\lambda}{\Psi}\partial_t\Psi\right) = q + f^\top\left(-\frac{\lambda}{\Psi}\Psi_s\right) - \frac{1}{2}\left(-\frac{\lambda}{\Psi}\Psi_s\right)^\top GR^{-1}G^\top\left(-\frac{\lambda}{\Psi}\Psi_s\right) + \frac{1}{2}\text{tr}\left(BB^\top\left(-\frac{\lambda}{\Psi}\Psi_{ss} + \frac{\lambda}{\Psi^2}\Psi_s\Psi_s^\top\right)\right)$$

multiplying with $\frac{\Psi}{\lambda}$ yields - a still nonlinear PDE :(

$$\partial_t\Psi = \frac{\Psi}{\lambda}q - f^\top\Psi_s - \frac{\lambda}{2\Psi}\Psi_s^\top GR^{-1}G^\top\Psi_s - \frac{1}{2}\text{tr}\left(BB^\top\left(\Psi_{ss}\right)\right) + \frac{1}{2\Psi}\text{tr}\left(BB^\top\Psi_s\Psi_s^\top\right)$$

Path Integral Control - Transformation in Linear PDE

- Using a basic property of the trace:

$$\mathrm{tr}\left(BB^\top \Psi_s \Psi_s^\top\right) = \mathrm{tr}\left(\left(BB^\top \Psi_s\right)^\top \Psi_s\right) = \mathrm{tr}\left(\Psi_s^\top BB^\top \Psi_s\right) = \Psi_s^\top BB^\top \Psi_s$$

Path Integral Control - Transformation in Linear PDE

- Using a basic property of the trace:

$$\mathrm{tr}\left(BB^\top \Psi_s \Psi_s^\top\right) = \mathrm{tr}\left(\left(BB^\top \Psi_s\right)^\top \Psi_s\right) = \mathrm{tr}\left(\Psi_s^\top BB^\top \Psi_s\right) = \Psi_s^\top BB^\top \Psi_s$$

- Replace the trace yields:

$$\partial_t \Psi = \frac{\Psi}{\lambda} q - f^\top \Psi_s - \frac{\lambda}{2\Psi} \Psi_s^\top G R^{-1} G^\top \Psi_s - \frac{1}{2}\mathrm{tr}\left(BB^\top \Psi_{ss}\right) + \frac{1}{2\Psi} \Psi_s^\top BB^\top \Psi_s$$

If the assumption $\boxed{\lambda G R^{-1} G^\top = BB^\top}$ holds, the nonlinear terms cancel out and the **linear** PDE

$$\partial_t \Psi = \frac{\Psi}{\lambda} q - f^\top \Psi_s - \frac{1}{2}\mathrm{tr}\left(BB^\top \Psi_{ss}\right)$$

remains.

Path Integral Control - Feynman-Kac Lemma

Application of Feynman-Kac Lemma[1]

The solution of the parabolic PDE

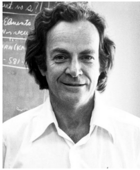
$$\partial_t \Psi = \frac{\Psi}{\lambda} q - f^\top \Psi_s - \frac{1}{2} \mathrm{tr}\left(BB^\top \Psi_{ss}\right)$$

with boundary condition $\Psi(s_T, T) = \exp\left(-\frac{1}{\lambda} E(s_T)\right)$ is given by the expectation

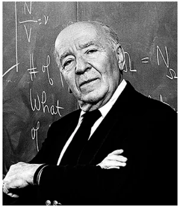
$$\Psi(s_t, t) = \mathbb{E}_{s(\cdot)\sim\mathbb{P}}\left[\exp\left(-\frac{1}{\lambda}\int_t^T q(s_\tau)\mathrm{d}\tau\right)\Psi(s_T, T)\right],$$

where $\mathbb{E}_{s(\cdot)\sim\mathbb{P}}[\cdot]$ denotes the expectation over trajectories taken with respect to the **uncontrolled** system dynamics $\mathrm{d}s_t = f(s_t)\mathrm{d}t + B(s_t)\mathrm{d}w$.

[1] Øksendal (2000), Stochastic Differential Equation, Springer.



Richard Feynman (1918-1988)



Mark Kac (1914-1984)

Path Integral Control - Value Function as Expectation

- Defining the state-dependent portion of the costs of a trajectory $s(\cdot)$ as

$$S(s(\cdot)) := \int_t^T q(s_\tau)\mathrm{d}\tau + E(s_T)$$

- Then we can express the desirability function as a *Path Integral*

$$\Psi(s_t, t) = \mathbb{E}_{s(\cdot)\sim\mathbb{P}}\left[\exp\left(-\frac{1}{\lambda}S(s(\cdot))\right)\right]$$

- Transform it back yields the value function as expectation

$$V(s_t, t) = -\lambda \log \mathbb{E}_{s(\cdot)\sim\mathbb{P}}\left[\exp\left(-\frac{1}{\lambda}S(s(\cdot))\right)\right]$$

Path Integral Control - Controls as Expectation

Remember: The optimal action $\boxed{a^*(s_t, t) = -R(s_t)^{-1} G(s_t)^\top V_s(s_t, t)}$ is dependent on V_s

Analytically computation of V_s is lengthy but straightforward[1] and results in

$$a^* \mathrm{d}t = R^{-1} G^\top (G R^{-1} G^\top)^{-1} \frac{\mathbb{E}_{s(\cdot) \sim \mathbb{P}} \left[\exp\left(-\frac{1}{\lambda} S(s(\cdot))\right) B \mathrm{d}w \right]}{\mathbb{E}_{s(\cdot) \sim \mathbb{P}} \left[\exp\left(-\frac{1}{\lambda} S(s(\cdot))\right) \right]}$$

[1] Theodorou, Buchli, Schaal (2010), A Generalized Path Integral Approach to Reinforcement Learning.

Path Integral Control - Time Discrete Approximation

The time discrete approximation with $\Delta t = T/N$ yields

$$a^* = \underbrace{R^{-1}G^\top (GR^{-1}G^\top)^{-1}}_{\text{projection operator}} \frac{\mathbb{E}_{\hat{s}\sim\hat{\mathbb{P}}}\left[\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s})\right) B \frac{\epsilon_k}{\sqrt{\Delta t}}\right]}{\mathbb{E}_{\hat{s}\sim\hat{\mathbb{P}}}\left[\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s})\right)\right]},$$

where

- $\mathrm{d}w \approx \epsilon_k \sqrt{\Delta t}$ with random variable $\epsilon_k \sim \mathcal{N}(0, I_{n_a \times n_a})$,
- state sequence $\hat{s} = (\hat{s}_0, \hat{s}_1, ..., \hat{s}_N)$,
- path costs $\hat{S}(\hat{s}) = E(s_N) + \sum_{k=0}^{N-1} q(\hat{s}_k)\Delta t$,

and $\mathbb{E}_{\hat{s}\sim\hat{\mathbb{P}}}[\cdot]$ denotes the expectation over trajectories taken with respect to the dynamics

$$\hat{s}_{k+1} = \hat{s}_k + f(\hat{s}_k)\Delta t + B(\hat{s}_k)\epsilon_k\sqrt{\Delta t} \quad \text{with} \quad \hat{s}_0 = s_t$$

- set $\Delta t = 1$ because the choice of unit of time is arbitrary (without loss of generality)

Path Integral Control - Monte Carlo Estimation

In special case $B = G\sqrt{\Sigma}$

$$a_k^* = \frac{\mathbb{E}_{\hat{s}\sim\hat{\mathbb{P}}}\left[\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s})\right) v_k\right]}{\mathbb{E}_{\hat{s}\sim\hat{\mathbb{P}}}\left[\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s})\right)\right]}$$

the projection operator vanishes, where Σ is the covariance of $v_k \sim \mathcal{N}(0, \Sigma)$.

Note, in this special case, assumption $\lambda G R^{-1} G^\top = B B^\top$ reduces to $R = \lambda \Sigma^{-1}$.

Path Integral Control - Monte Carlo Estimation

The **optimal action**

$$a_k^* = \frac{\mathbb{E}_{\hat{s}\sim\hat{\mathbb{P}}}\left[\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s})\right)v_k\right]}{\mathbb{E}_{\hat{s}\sim\hat{\mathbb{P}}}\left[\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s})\right)\right]}$$

can be approximated using **Monte Carlo estimation**:

$$a_k^* \approx \frac{1}{I}\sum_{i=0}^{I-1}\frac{\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s}^i)\right)v_{k,i}}{\frac{1}{I}\sum_{n=0}^{I-1}\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s}^n)\right)} = \sum_{i=0}^{I-1} w_i v_{k,i}, \quad \text{with } w_i = \frac{\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s}^i)\right)}{\sum_{n=0}^{I-1}\exp\left(-\frac{1}{\lambda}\hat{S}(\hat{s}^n)\right)}$$

Recap so far...

- Closed-form solution of the HJB and the corresponding optimal control by a transformation in an expectation over all possible trajectories - a so called *Path Integral*
- This expectation can then be approximated via a Monte Carlo approximation using forward sampling of the uncontrolled stochastic system dynamics

Question

Which control methods can be derived using the path integral framework?

Path Integral Control - Methods

Open-Loop Planning with Path Integrals[1]

Sampling takes place from the initial state of the optimal control problem. Most straightforward approach. However, no reaction to noise.

$$a_k^*(s) \approx A^*(k)$$

Policy Improvement with Path Integrals[2]

Sampling in policy parameter space. More effective approach by employing the path integral control framework to find the optimal parameters of a feedback control policy.

$$a^*(t, s) \approx \pi_{\theta^*}(t, s)$$

Model Predictive Path Integral Control[3]

MPC setting: Open-loop control sequence is constantly optimized while processing in real time. High calculation effort.

$$a_k^*(s) \approx A_{s_0=s}^*(0)$$

[1] e.g. Theodorou, Todorov (2012), Relative Entropy and Free Energy Dualities
[2] e.g. Theodorou, Buchli, Schaal (2010), Learning Policy Improvements with Path Integrals
[3] e.g. Gómez, ..., Kappen (2016), Real-Time Stochastic Optimal Control for Multi-Agent Quadrotor Systems

Major Issue

- Major issue: Expectation is taken with respect to the uncontrolled dynamics of the system. This is problematic because the probability of sampling low-cost trajectories is very low.

How can we circumvent this issue?

- Importance sampling!

Importance Sampling

The expected value is defined as

$$\mathbb{E}_{v\sim\mathbb{Q}}[v] = \int vq(v)dv,$$

where $q(v)$ denotes the probability density function (pdf) of probability distribution $\mathbb{Q}$. We can extend this expression to

$$\int vq(v)dv = \int vq(v)\frac{p(v)}{p(v)}dv,$$

where $p(v)$ is the pdf of probability distribution $\mathbb{P}$. If for $\mathbb{Q}$ and $\mathbb{P}$ hold $(p(v) = 0) \leftrightarrow (q(v) = 0)$ (absolute continuity), then

$$\mathbb{E}_{v\sim\mathbb{Q}}[v] = \mathbb{E}_{v\sim\mathbb{P}}[w(v)v] \text{ with } w(v) = \frac{q(v)}{p(v)}$$

holds and we can sample from another distribution.

MPPI - Algorithm

Algorithm

1. Sample input trajectories around initial guess
2. Simulate and compute path costs for each sampled trajectory
3. Compute weights corresponding to sampled trajectories
4. Approximate optimal control sequence via weighted mean
5. Apply first element of optimal control sequence, shift, and go to 1

MPPI - Algorithm in Detail

Wiliams et al. (2017), Information Theoretic MPC for Model-Based Reinforcement Learning

Algorithm 1 MPPI

Input: F: Discrete-time system dynamics;
I: Number of sampled trajectories;
N: Number of time steps;
$(a_0, a_1, ..., a_{N-1})$: Initial control sequence;
$q(\cdot), E(\cdot)$: State dependent stage costs and terminal costs;
Σ, λ: Covariance and temperature; ← covariance $\Sigma \in \mathbb{R}^{n_a \times n_a}$ and temperature $\lambda \in \mathbb{R}$

while *task not completed* **do**
  $s_0 \leftarrow$ GetRecentStateEstimate();
  for $i \in \{0, 1, ..., I-1\}$ **do**
    $\hat{s}_0^i \leftarrow s_0$;
    Sample $\{\varepsilon_0^i, \varepsilon_1^i, ..., \varepsilon_{N-1}^i\}$; ← with $\epsilon_k^i \sim \mathcal{N}(0, \Sigma)$ for $k = 0, 1, ..., T-1$
    $\hat{S}_i \leftarrow 0$;
    for $k \in \{0, 1, ..., N-1\}$ **do**
      $\hat{s}_{k+1}^i \leftarrow F(\hat{s}_k^i, a_k + \varepsilon_k^i)$;
      $\hat{S}_i \leftarrow \hat{S}_i + q(\hat{s}_k^i) + \lambda a_k^\top \Sigma^{-1} \varepsilon_k^i$; ← the second term is the correction term for importance sampling
    end for
    $\hat{S}_i \leftarrow \hat{S}_i + E(\hat{s}_N^i)$;
  end for
  $\beta \leftarrow \min_i [\hat{S}_i]$
  $\eta \leftarrow \sum_{i=0}^{I-1} \exp(-\frac{1}{\lambda}(\hat{S}_i - \beta))$; ← using β is technical trick for numerical stability
  for $i \in \{0, 1, ..., I-1\}$ **do**
    $w_i \leftarrow \frac{1}{\eta} \exp(-\frac{1}{\lambda}(\hat{S}_i - \beta))$;
  end for
  for $k \in \{0, 1, ..., N-1\}$ **do**
    $a_k \leftarrow a_k + \sum_{i=0}^{I-1} w_i \varepsilon_k^i$;
  end for
  SendToActuators(a_0);
  for $k \in \{0, 1, ..., N-2\}$ **do**
    $a_k \leftarrow a_{k+1}$; ← shift for warmstart
  end for
end while

Classification

Strengths of MPPI:

- Can handle nonsmooth costs and dynamics
- Easy to implement
- Computation in parallel tasks (GPU)
- Considering stochastic dynamics

Weaknesses of MPPI:

- No sufficient condition
- Performance highly depends on initial guess
- Relation $R = \lambda \Sigma^{-1}$ is necessary
- A lot of samples are necessary to explore high dimensional space

Full Scale Applications

Check out our website

Further readings

Basics

- Kappen (2005), Path integrals and symmetry breaking for optimal control theory
- Gómez et al. (2014), Policy Search for Path Integral Control
- Wiliams et al. (2018), Information Theoretic Model Predictive Control: Theory and Applications to Autonomous Driving

Recent research

- Lefebvre et al. (2019), Path Integral Policy Improvement with DDP
- Kusumoto et al. (2019), Informed Information Theoretic Model Predictive Control
- Balci et al. (2022), Constrained Covariance Steering Based Tube-MPPI
- Wang et al. (2022), Sampling-Based Optimization for Multi-Agent MPC
- Kim et al. (2022), Smooth Model Predictive Path Integral Control Without Smoothing
- Streichenberg et al. (2023), Multi-Agent Path Integral Control for Interaction-Aware Motion Planning in Urban Canals

Attachment: Stochastic HJB Equation

Sketch for one dimensional case

$$V(s,t) = \min_a \mathbb{E}_{\mathrm{d}s_t \sim \mathbb{Q}(a)} \left[L(s,a)\mathrm{d}t + V(s + \mathrm{d}s_t, t + \mathrm{d}t) \right]$$

Assuming V is differentiable in x and t - we can use a Taylor series

$$V(s + \mathrm{d}s_t, t + \mathrm{d}t) = V(s,t) + V_t \mathrm{d}t + V_s \mathrm{d}s_t + \frac{1}{2} V_{ss} (\mathrm{d}s_t)^2 + V_{st} \mathrm{d}s_t \mathrm{d}_t + \text{"Higher Order Terms"}$$

With general dynamics

$$\mathrm{d}s_t = \overline{f}(t, s_t, a_t)\mathrm{d}t + B(t, s_t, a_t)\mathrm{d}w$$

we can formally write

$$(\mathrm{d}s_t)^2 = \overline{f}^2 (\mathrm{d}t)^2 + B^2 (\mathrm{d}w)^2 + 2\overline{f} B \mathrm{d}w \mathrm{d}t$$

$$\mathrm{d}s_t \mathrm{d}t = \overline{f} (\mathrm{d}t)^2 + B \mathrm{d}w \mathrm{d}t$$

Attachement: Stochastic HJB Equation

Applying rules of stochastic calculus with $(\mathrm{d}w)^2 = \mathrm{d}t$, $\mathrm{d}w\mathrm{d}t = 0$, $\mathrm{d}t^2 = 0$, and $\mathbb{E}[\mathrm{d}w] = 0$ yield

$$V = \min_a \left[L\mathrm{d}t + V + V_t\mathrm{d}t + V_s\overline{f}\mathrm{d}t + \frac{1}{2}V_{ss}B^2\mathrm{d}t + \text{"Higher Order Terms"} \right]$$

Subtracting V and $V_t\mathrm{d}t$ on both sides and dividing by $\mathrm{d}t$ results in **stochastic HJB**:

$$-V_t = \min_a \left[L + \overline{f}V_s + \frac{1}{2}B^2V_{ss} \right]$$

In multidimensional case[1]:

$$-V_t = \min_u \left[L + \overline{f}^\top V_s + \frac{1}{2}\mathrm{tr}(BB^\top V_{ss}) \right]$$

[1] Fleming and Rishel (1975), Deterministic and stochastic optimal control. Springer.

References Pictures

- Picture Norbert Wiener, Britannica, https://www.britannica.com/biography/Norbert-Wiener, retrieved 12.09.2023
- Picture Richard Feynman, The Guardian, https://www.theguardian.com/science/2011/may/15/quantum-man-richard-feynman-review, retrieved 14.09.2023
- Picture Mark Kac, Portret z historia Mark Kac, https://www.czczaplinski.com/post/portret-z-historia-mark-kac, retrieved 14.09.2023