ECE276B: Planning & Learning in Robotics
Lecture 12: Continuous-time Optimal Control

Lecturer:

Nikolay Atanasov: natanasov@ucsd.edu

Teaching Assistants:

Tianyu Wang: tiw161@eng.ucsd.edu

Yongxi Lu: yol070@eng.ucsd.edu

UC San Diego

JACOBS SCHOOL OF ENGINEERING

Electrical and Computer Engineering

Continuous-time Optimal Control

- **System Dynamics**:
  - time $t \in [t_0, T]$
  - state $x(t) \in \mathcal{X} \subseteq \mathbb{R}^n$, $\forall t \in [t_0, T]$
  - control $u(t) \in \mathcal{U} \subseteq \mathbb{R}^m$, $\forall t \in [t_0, T]$
  - stochastic differential equation (Ito diffusion):

$$dx = f(x(t), u(t))dt + C(x(t), u(t))d\omega, \qquad x(t_0) = x_0$$

- **Noise**: Brownian motion $\omega(t)$ (integral of white noise)
- Infinite-dimensional dynamic constrained optimization:

$$\min_{\pi \in PC^0([t_0,T],\mathcal{U})} J^\pi(t_0, x_0) := \mathbb{E}\left\{ \underbrace{\int_{t_0}^{T} g(x(t), \pi(t, x(t)))dt}_{\text{running cost}} + \underbrace{g_T(x(T))}_{\text{terminal cost}} \,\middle|\, x(t_0) = x_0 \right\}$$

$$\begin{aligned} \text{s.t.} \quad & dx = f(x(t), \pi(t, x(t)))dt + C(x(t), \pi(t, x(t)))d\omega. \\ & x(t) \in \mathcal{X},\ \pi(t, x(t)) \in \mathcal{U} \end{aligned}$$

  - **Admissible policies**: $u(t) := \pi(t, x) \in \Pi := PC^0([t_0, T], \mathcal{U})$ are piecewise cont. functions that map a state x at time t to a control input
  - T can be free or fixed; $x(T)$ can be free or in a target set $\mathcal{T}$
  - Additional state constraints can be imposed via the set $\mathcal{X}$

Assumptions and Technical Details

- **Assumptions**
  1. f is cont-diffable wrt to x and cont wrt u
  2. **Existence and Uniqueness**: for any admissible policy π and initial $x(\tau) \in \mathcal{X}$, $\tau \in [t_0, T]$, the **noise-free** system has a unique **state trajectory** $x(t)$
  3. The running cost $g(x, u)$ is cont-diffable wrt x and cont wrt u
  4. The terminal cost $g_T(x)$ is cont-diffable wrt x

- The SDE means that the time-integrals of the two sides are equal:

$$x(T) - x(0) = \int_0^T f(x(t), u(t))dt + \underbrace{\int_0^T C(x(t), u(t))d\omega(t)}_{\text{Ito intergral}}$$

- Cannot be written as $\dot{x} = f(x, u) + C(x, u)\dot{\omega}$ because $\dot{\omega}$ does not exist

- The **Ito integral** of a random process $y(t)$ adapted to $\omega(t)$, i.e., $y(t)$ depends on the sample path of $\omega(t)$ up to time t, is:

$$\int_0^T y(t)d\omega(t) := \lim_{\substack{N\to\infty \\ 0=t_0<t_1<\cdots<t_N=T}} \sum_{i=0}^{N-1} y(t_i)(\omega(t_{i+1}) - \omega(t_i))$$

Existence and Uniqueness

- **Existence and Uniqueness**: for any admissible policy π and initial $x(\tau) \in \mathcal{X}$, $\tau \in [t_0, T]$, the **noise-free** system has a unique **state trajectory** $x(t)$

- **Example**: Existence in not guaranteed in general

$$\dot{x}(t) = x(t)^2, \; x(0) = 1$$

$$\text{Solution does not exist for} \, T \geq 1 : x(t) = \frac{1}{1-t}$$

- **Example**: Uniqueness in not guaranteed in general

$$\dot{x}(t) = x(t)^{\frac{1}{3}}, \; x(0) = 0$$

$$\text{Infinite number of solutions :} \quad \begin{aligned} & x(t) = 0, \; \forall t \\ & x(t) = \begin{cases} 0 & \text{for } 0 \leq t \leq \tau \\ \left(\frac{2}{3}(t-\tau)\right)^{3/2} & \text{for } t > \tau \end{cases} \end{aligned}$$

Calculus of Variations

- An infinite-dimensional <u>static</u> constrained optimization
- It is a special case of deterministic continuous-time optimal control for a <u>fully-actuated</u> system ($\dot{x} = u$) with $t \leftarrow x$, $x(t) \leftarrow y(x)$, $u(t) = \dot{x}(t) \leftarrow \dot{y}(x)$, $g_T(x(T)) \leftarrow h(y(b))$:

$$\min_{y \in C^1([a,b],\mathbb{R}^m)} \quad \int_a^b g(y(x), \dot{y}(x))dx + h(y(b))$$
$$\text{s.t.} \quad y(a) = y_0, \; y(b) = y_f$$

Optimal Cost-to-Go

- **Optimal Value/Cost-to-go Function**: the closed loop cost $J^*(t, x)$ associated with an optimal feedback control law $u^*(t) := \pi^*(t, x)$ at state x and time t:

$$J^*(t, x) \leq J^{\pi}(t, x), \quad \forall \pi \in \Pi, x \in \mathcal{X}$$

HJB PDE

The Hamilton-Jacobi-Bellman (HJB) partial differential equation (PDE) is satisfied for all time-state pairs (t, x) by the optimal cost-to-go $J^*(t, x)$:

$$J^*(T, x) = g_T(x), \quad \forall x \in \mathcal{X}$$

$$-\frac{\partial}{\partial t} J^*(t, x) = \min_{u \in \mathcal{U}} \left\{ g(x, u) + \nabla_x J^*(t, x)^T f(x, u) + \frac{1}{2} \operatorname{tr}\left(\Sigma(x, u)\left[\nabla_x^2 J^*(t, x)\right]\right) \right\}$$

for all $t \in [t_0, T]$ and $x \in \mathcal{X}$ and where $\Sigma(x, u) := C(x, u) C^T(x, u)$.

- The HJB PDE is the continuous-time analog of Dynamic Programming

HJB PDE Derivation

- A discrete-time approximation of the cont.-time optimal control problem can be used to derive the HJB PDE from the DP algorithm
- **Euler Discretization** of the SDE with time step τ:
  - Discretize $[t_0, T]$ into N pieces of width $\tau := \frac{T - t_0}{N}$
  - Define $x_k := x(k\tau)$ and $u_k := u(k\tau)$ for $k = 0, \ldots, N$
  - **Discretized system dynamics**:

$$x_{k+1} = x_k + \tau f(x_k, u_k) + \epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0, \tau\Sigma(x_k, u_k))$$

    so that the motion model is specified by a Gaussian pdf:

$$p_f(x' \mid x, u) = \phi(x'; x + \tau f(x, u), \tau\Sigma(x, u))$$

  - **Discretized stage cost**: $\tau g(x, u)$
- **Idea**: apply the Bellman Equation to the now discrete-time problem and take the limit as $\tau \to 0$ to obtain a "continuous-time Bellman Equation"

HJB PDE Derivation

- **Bellman Equation**: finite-horizon problem with $t := k\tau$

$$V(t,x) = \min_{u \in \mathcal{U}(x)} \left\{ \tau g(x,u) + \mathbb{E}_{x' \sim p_f(\cdot|x,u)} \left[V(t+\tau, x') \right] \right\}$$

- Taylor-series expansion of $V(t+\tau, x')$ around (t,x):

$$\begin{aligned} V(t+\tau, x+d) = & V(t,x) + \tau \frac{\partial V}{\partial t}(t,x) + o(\tau^2) \\ & + \left[\nabla_x V(t,x)\right]^T d + \frac{1}{2} d^T \left[\nabla_x^2 V(t,x)\right] d + o(d^3) \end{aligned}$$

where $d \sim \mathcal{N}(\tau f(x,u), \tau \Sigma(x,u))$

- Note that $\mathbb{E}\left[d^T M d\right] = \text{tr}(\Sigma M) + \text{tr}(\mu \mu^T M)$ for $d \sim \mathcal{N}(\mu, \Sigma)$

HJB PDE Derivation

- Note that $\mathbb{E}\left[d^T M d\right] = \text{tr}(\Sigma M) + \text{tr}(\mu\mu^T M)$ for $d \sim \mathcal{N}(\mu, \Sigma)$ so that:

$$\mathbb{E}_{x' \sim p_f(\cdot|x,u)}\left[V(t+\tau, x')\right] = V(t,x) + \tau \frac{\partial V}{\partial t}(t,x) + o(\tau^2)$$
$$+ \tau \left[\nabla_x V(t,x)\right]^T f(x,u) + \frac{\tau}{2} \text{tr}\left(\Sigma(x,u)\left[\nabla_x^2 V(t,x)\right]\right)$$

- Substituting in the Bellman Equation and simplifying, we get:

$$0 = \min_{u \in \mathcal{U}(x)} \left\{ g(x,u) + \frac{\partial V}{\partial t}(t,x) + \left[\nabla_x V(t,x)\right]^T f(x,u) + \frac{1}{2} \text{tr}\left(\Sigma(x,u)\left[\nabla_x^2 V(t,x)\right]\right) + \frac{o(\tau^2)}{\tau} \right\}$$

- Taking the limit as $\tau \to 0$ (assuming it can be exchanged with $\min_{u \in \mathcal{U}}$) leads to the HJB PDE:

$$-\frac{\partial V}{\partial t}(t,x) = \min_{u \in \mathcal{U}(x)} \left\{ g(x,u) + \left[\nabla_x V(t,x)\right]^T f(x,u) + \frac{1}{2} \text{tr}\left(\Sigma(x,u)\left[\nabla_x^2 V(t,x)\right]\right) \right\}$$

HJB PDEs for Different Problem Formulations

- **Discounted Problem**: $J^{\pi}(x) := \mathbb{E}\left[\int_0^{\infty} \underbrace{e^{-\frac{t}{\gamma}}}_{\text{discount}} g(x(t), \pi(t, x(t)))dt\right]$ with $\gamma \in [0, \infty)$

HJB PDEs for the Optimal Cost-to-go

Hamiltonian: $$H[x, u, p] = g(x, u) + p^T f(x, u) + \frac{1}{2}\operatorname{tr}\left(\Sigma(x, u)[\nabla_x p(x)]\right)$$

Finite Horizon: $$-\frac{\partial J^*}{\partial t}(t, x) = \min_{u \in \mathcal{U}(x)} H[x, u, \nabla_x J^*(t, \cdot)], \qquad J^*(T, x) = g_T(x)$$

First Exit: $$0 = \min_{u \in \mathcal{U}(x)} H[x, u, \nabla_x J^*(\cdot)], \qquad J^*(x \in \mathcal{T}) = g_{\mathcal{T}}(x)$$

Discounted: $$\frac{1}{\gamma} J^*(x) = \min_{u \in \mathcal{U}(x)} H[x, u, \nabla_x J^*(\cdot)]$$

Existence and Uniqueness of Solutions

- The HJB PDE has at most one classical solution (i.e., a function which satisfies the PDE everywhere)
- If a classical solution exists then it is the optimal cost-to-go
- The HJB PDE may not have a classical solution, in which case the optimal cost-to-go is not smooth (e.g. bang-bang control)
- The HJB PDE always has a unique viscosity solution which is the optimal cost-to-go
- Approximation schemes based on MDP discretization are guaranteed to converge to the unique viscosity solution
- Most continuous function approximation schemes (which scale better) are unable to represent non-smooth solutions
- All examples of non-smoothness seem to be deterministic; noise tends to smooth the optimal cost-to-go function

HJB Example 1

- System: $\dot{x}(t) = u(t)$, $|u(t)| \leq 1$, $0 \leq t \leq 1$
- Costs: $g(x, u) = 0$ and $g_T(x) = \frac{1}{2}x^2$ for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$
- Since we only care about the square of the terminal state, we can construct a candidate optimal policy that drives the state towards 0 as quickly as possible and maintains it there:

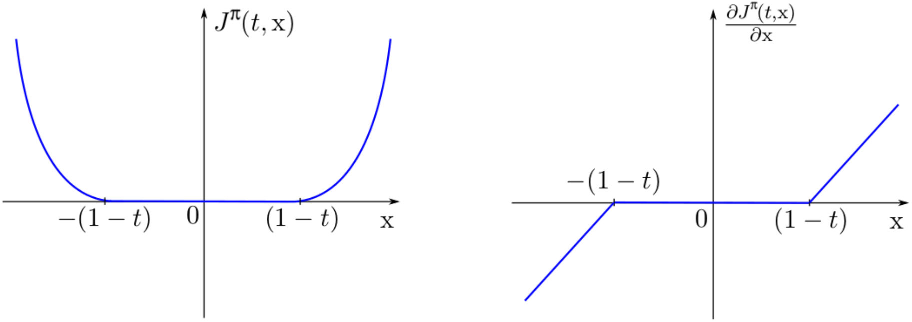
$$\pi(t, x) = -sgn(x) := \begin{cases} -1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x < 0 \end{cases}$$

- The cost-to-go in not smooth: $J^{\pi}(t, x) = \frac{1}{2}\left(\max\left\{0, |x| - (1 - t)\right\}\right)^2$
- We will verify that this function satisfies the HJB and is therefore indeed the optimal cost-to-go function

HJB Example 1: Partial Derivative wrt x

- Cost function and its partial derivative wrt x for fixed t:

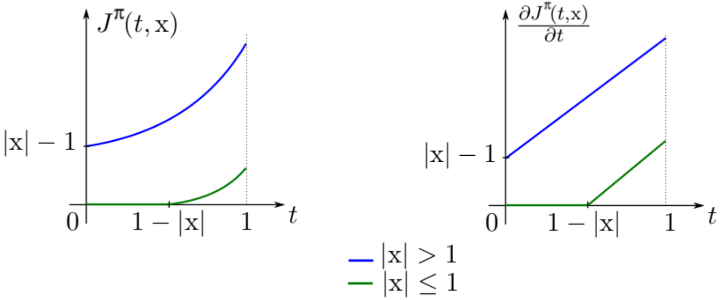
$$J^{\pi}(t,x) = \frac{1}{2}\left(\max\{0, |x| - (1-t)\}\right)^2 \qquad \frac{\partial J^{\pi}(t,x)}{\partial x} = sgn(x)\max\{0, |x| - (1-t)\}$$



HJB Example 1: Partial Derivative wrt t

- Cost function and its partial derivative wrt t for fixed x:

$$J^\pi(t,x) = \frac{1}{2}\left(\max\{0, |x| - (1-t)\}\right)^2 \qquad \frac{\partial J^\pi(t,x)}{\partial t} = \max\{0, |x| - (1-t)\}$$



HJB Example 1

- Boundary condition: $J^{\pi}(1,x) = \frac{1}{2}x^2 = g_T(x)$
- The minimum in the HJB PDE is obtained by $u = -sgn(x)$:

$$\min_{|u|\leq 1}\left(\frac{\partial J^{\pi}(t,x)}{\partial t} + \frac{\partial J^{\pi}(t,x)}{\partial x}u\right) = \min_{|u|\leq 1}\left((1+sgn(x)u)\left(\max\{0,|x|-(1-t)\}\right)\right) = 0$$

- Conclusion: $J^{\pi}(t,x) = J^{*}(t,x)$ and $\pi^{*}(t,x) = -sgn(x)$ is an optimal policy
- Note that this is a simple example. In general, solving the HJB is nontrivial.

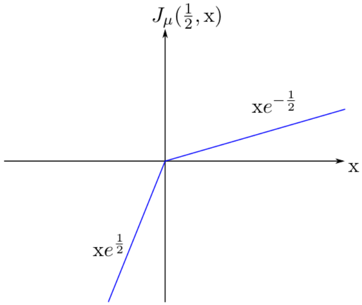
HJB Example 2

- System: $\dot{x}(t) = x(t)u(t)$, $|u(t)| \leq 1$, $0 \leq t \leq 1$
- Costs: $g(x, u) = 0$ and $g_T(x) = x$ for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$
- Optimal policy:

$$\pi(t, x) = \begin{cases} -1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x < 0 \end{cases}$$

- Optimal cost-to-go:

$$J^\pi(t, x) = \begin{cases} e^{t-1}x & x > 0 \\ 0 & x = 0 \\ e^{1-t}x & x < 0 \end{cases}$$



- The cost-to-go function is not diffable wrt x at $x = 0$ and hence does not satisfy the HJB in the classical sense

Theorem: HJB PDE Sufficiency

Suppose that $V(t,x)$ is <u>cont-diffable</u> in t and x and solves the HJB equation:

$$V(T,x) = g_T(x), \quad \forall x \in \mathcal{X}$$
$$-\frac{\partial V(t,x)}{\partial t} = \min_{u \in \mathcal{U}} \left[g(x,u) + \nabla_x V(t,x)^T f(x,u) + \frac{1}{2} \operatorname{tr}\left(\Sigma(x,u) \left[\nabla_x^2 V(t,x) \right] \right) \right]$$

for all $x \in \mathcal{X}$ and $0 \le t \le T$. Suppose also that the policy $\pi^*(t,x)$ attains the minimum in the HJB for all t and x and is <u>piecewise-cont</u> in t. Then under the assumptions on Slide 3, $V(t,x)$ is the unique solution of the HJB equation and is equal to the optimal cost-to-go $J^*(t,x)$, while $\pi^*(t,x)$ is an optimal policy.

More Tractable Problems

- Consider a restricted class of system dynamics and cost functions:

$$dx = (a(x) + B(x)u)dt + C(x)d\omega$$
$$g(x, u) = q(x) + \frac{1}{2}u^T R(x)u$$

- The Hamiltonian can be minimized analytically wrt u for such problems (suppressing the dependence on x for clarity):

$$\pi^* = \arg\min_u \left\{ q + \frac{1}{2}u^T Ru + (a + Bu)^T V_x + \frac{1}{2}\operatorname{tr}(CC^T V_{xx}) \right\}$$
$$= -R^{-1}B^T V_x$$
$$H[x, \pi^*, V_x] = q + a^T V_x + \frac{1}{2}\operatorname{tr}(CC^T V_{xx}) - \frac{1}{2}V_x^T BR^{-1}B^T V_x$$

- The HJB PDE becomes second-order quadratic, no longer involving the min operator!

More Tractable Problems (Generalizations)

- **Control-multiplicative Noise**: $\Sigma(x,u) = C_0(x)C_0(x)^T + \sum_j C_j(x)uu^T C_j(x)^T$

$$\pi^* = -\left(R + \sum_j C_j^T V_{xx} C_j\right)^{-1} B^T V_x$$

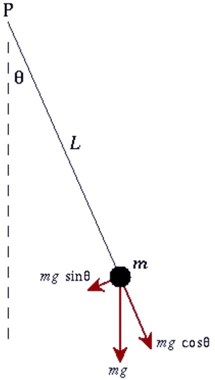
- **Convex-in-control Costs**: $g(x,u) = q(x) + \sum_j r(u_j)$ with convex r:

$$\pi^* = \arg\min_u \left\{ \sum_j r(u_j) + u^T B^T V_x \right\} = (r')^{-1}\left(-B^T V_x\right)$$

- **Example**:

$$r(u) = s\int_0^{|u|} \mathbf{atanh}\left(\frac{\omega}{u_{max}}\right) d\omega \quad \Rightarrow \quad \pi^* = u_{max}\tanh\left(-s^{-1}B^T V_x\right)$$

Pendulum Example



- **Pendulum dynamics** (Newton's second law for rotational systems):

$$mL^2\ddot{\theta} = u - mgL\sin\theta + noise$$

- State-space form with $x = (x_1, x_2) = (\theta, \dot{\theta})$:

$$dx = \begin{bmatrix} x_2 \\ k\sin(x_1) \end{bmatrix} dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} (udt + \sigma d\omega)$$

- **Stage cost**: $g(x, u) = q(x) + \frac{r}{2}u^2$
- Optimal cost-to-go and policy (discounted formulation):

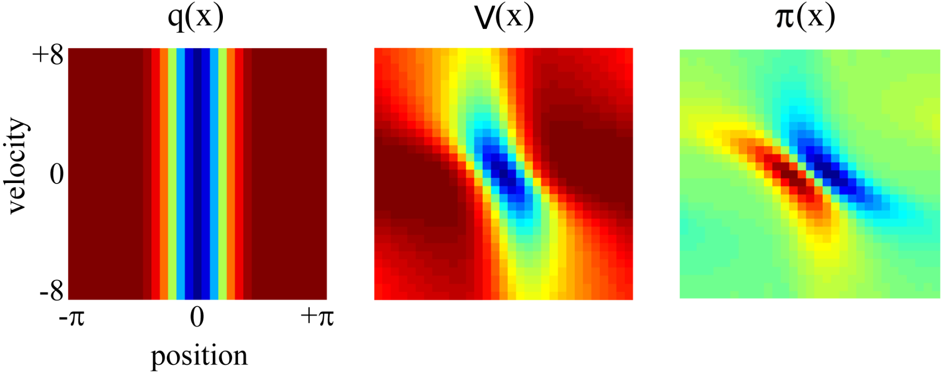
$$\pi^*(x) = -\frac{1}{r} J^*_{x_2}(x)$$

$$\frac{1}{\gamma} J^*(x) = q(x) + x_2 J^*_{x_1}(x) + k\sin(x_1) J^*_{x_2}(x) + \frac{\sigma^2}{2} J^*_{x_2 x_2}(x) - \frac{1}{2r}(J^*_{x_2}(x))^2$$

Pendulum Example

- Parameters: $k = \sigma = r = 1$, $\gamma = 0.3$, $q(\theta, \dot{\theta}) = 1 - \exp(-2\theta^2)$
- Discretize the state space, approximate derivatives via finite differences, and iterate:

$$V^{(n+1)}(x) = \beta V^{(n)}(x) + (1-\beta)\gamma \min_u H[x, u, V^{(n)}(\cdot)], \qquad \beta = 0.99$$

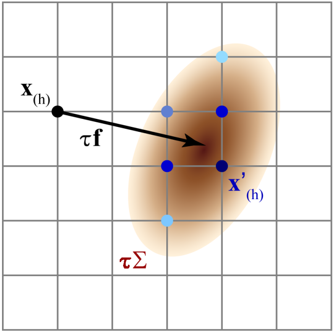


MDP Discretization

- Define discrete state space $\mathcal{X}_{(h)} \subset \mathbb{R}^n$, control space $\mathcal{U}_{(h)} \subset \mathbb{R}^m$, and time step $\tau_{(h)}$, where h is a coarseness parameter such that $h \to 0$ corresponds to infinitely dense discretization

- **Local Consistency**: construct a motion model $x'_{(h)} = x_{(h)} + d$ with:

$$\mathbb{E}[d] = \tau_{(h)} f(x_{(h)}, u_{(h)}) + o(\tau_{(h)})$$
$$\textbf{cov}[d] = \tau_{(h)} \Sigma(x_{(h)}, u_{(h)}) + o(\tau_{(h)})$$



- **Kushner and Dupois**: In the limit $h \to 0$, the MDP solution $J^*_{(h)}$ converges to the solution J^* of the continuous problem (even for non-smooth J^*)

MDP Discretization

- For each $x_{(h)}, u_{(h)}$ choose vectors $\{d_j\}_{j=1}^{K}$ such that all possible next states are $x'_{(h)} = x_{(h)} + hd_j$
- Specify $\tau_{(h)}$ and $p^j_{(h)} := p_f(x_{(h)} + hd_j \mid x_{(h)}, u_{(h)})$ according to one of the strategies:

1. $\tau_{(h)} = \frac{h^2}{h+1}$ and $p^j_{(h)} = \frac{h\alpha_j + \beta_j}{h+1}$ for α_j, β_j such that:

$$\begin{aligned}
\textstyle\sum_j \alpha_j d_j &= f(x_{(h)}, u_{(h)}) \\
\textstyle\sum_j \beta_j d_j &= 0 \\
\textstyle\sum_j \beta_j d_j d_j^T &= \Sigma(x_{(h)}, u_{(h)}) \\
\textstyle\sum_j \alpha_j &= 1,\ \alpha_j \geq 0 \\
\textstyle\sum_j \beta_j &= 1,\ \beta_j \geq 0
\end{aligned}$$

2. $\tau_{(h)} = h$ and

$$\begin{aligned}
\min_{\{p^j_{(h)}\}} \ & \|\Sigma - h \sum_j p^j_{(h)} (d_j - f)(d_j - f)^T\|^2 \\
\text{s.t} \ & \sum_j p^j_{(h)} d_j = f(x_{(h)}, u_{(h)}) \\
& \sum_j p^j_{(h)} = 1,\ p^j_{(h)} \geq 0
\end{aligned}$$

3. $\tau_{(h)} = h$ and

$$p^j_{(h)} \propto \phi(x_{(h)} + hd_j; hf(x_{(h)}, u_{(h)}), h\Sigma(x_{(h)}, u_{(h)}))$$