ECE276B: Planning & Learning in Robotics
Lecture 14: Continuous-time Optimal Control

Instructor:

Nikolay Atanasov: natanasov@ucsd.edu

Teaching Assistants:

Zhichao Li: zhl355@eng.ucsd.edu

Ehsan Zobeidi: ezobeidi@eng.ucsd.edu

Ibrahim Akbar: iakbar@eng.ucsd.edu

UC San Diego

JACOBS SCHOOL OF ENGINEERING

Electrical and Computer Engineering

Continuous-time System Dynamics

- **time**: $t \in [0, T]$
- **state**: $x(t) \in \mathcal{X} \subseteq \mathbb{R}^n$, $\forall t \in [0, T]$
- **control**: $u(t) \in \mathcal{U} \subseteq \mathbb{R}^m$, $\forall t \in [0, T]$
- **motion model**: a stochastic differential equation (SDE):

$$dx = f(x(t), u(t))dt + C(x(t), u(t))d\omega, \qquad x(0) = x_0$$

- **noise**: Brownian motion $\omega(t)$ (integral of white noise)
  - Robert Brown made microscopic observations in 1827 that small particles in plant pollen, when immersed in liquid, exhibit highly irregular motion
  - **Definition**: a continuous stationary stochastic process $\omega(t)$ satisfying:
    - $\omega(0) = 0$
    - $\omega(t)$ is almost surely continuous (but nowhere differentiable)
    - $\omega(t)$ has independent increments, i.e., $(\omega(t_2) - \omega(t_1))$ and $(\omega(t_4) - \omega(t_3))$ are independent for $0 \leq t_1 < t_2 \leq t_3 < t_4$
    - $\omega(t) - \omega(s) \sim \mathcal{N}(0, t - s)$ for $0 \leq s \leq t$

Continuous-time System Dynamics

- 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:

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

Continuous-time Optimal Control

- Infinite-dimensional dynamic constrained optimization:

$$\min_{\pi \in PC^0([0,T],\mathcal{U})} V^{\pi}(0,x_0) := \mathbb{E}\left\{ \underbrace{\int_0^T \ell(x(t),\pi(t,x(t)))dt}_{\text{stage cost}} + \underbrace{\mathfrak{q}(x(T))}_{\text{terminal cost}} \;\middle|\; x(0) = x_0 \right\}$$

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

- **Admissible policies**: $u(t) := \pi(t,x(t)) \in \Pi := PC^0([0,T],\mathcal{U})$ 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 and control constraints can be imposed via the sets $\mathcal{X}$ and $\mathcal{U}$

Assumptions

- f is continuously differentiable wrt to x and continuous wrt u
- **Existence and Uniqueness**: for any admissible policy $\pi \in \Pi$ and initial $x(\tau) \in \mathcal{X}$, $\tau \in [0, T]$, the **noise-free** system, $\dot{x}(t) = f(x(t), \pi(t, x(t)))$, has a **unique state trajectory** $x(t)$, $t \in [\tau, T]$.
- The stage cost $\ell(x, u)$ is continuously differentiable wrt x and continuous wrt u
- The terminal cost $\mathfrak{q}(x)$ is continuously differentiable wrt x

Examples: Existence and Uniqueness

- **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

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

$$x(t) = 0,\ \forall t$$

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

Special case: Calculus of Variations

- **Calculus of Variations**: an infinite-dimensional static constrained optimization:

$$\begin{aligned}\min_{y\in C^1([a,b],\mathbb{R}^m)} \quad & \int_a^b \ell(y(x),\dot{y}(x))dx + \mathfrak{q}(y(b)) \\ \text{s.t.} \quad & y(a) = y_0,\ y(b) = y_f\end{aligned}$$

- It is a special case of the deterministic continuous-time optimal control problem for a **fully-actuated** system ($\dot{x} = u$) with $t \leftarrow x$, $x(t) \leftarrow y(x)$, $u(t) = \dot{x}(t) \leftarrow \dot{y}(x)$

Optimal Value Function

- The closed-loop cost $V^*(t,x)$ associated with an optimal control policy $u^*(t) := \pi^*(t, x(t))$ at state x and time t:

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

HJB PDE

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

$$V^*(T,x) = \mathfrak{q}(x), \quad \forall x \in \mathcal{X}$$

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

for all $t \in [0, T]$ 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 the Bellman Equation

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
- **Motion model**: $dx = f(x(t), u(t))dt + C(x(t), u(t))d\omega$ with $x(0) = x_0$
- **Euler Discretization** of the SDE with time step τ:
  - Discretize $[0, T]$ into N pieces of width $\tau := \frac{T}{N}$
  - 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**: $\tau\ell(x, u)$
- **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 $t := k\tau$

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

- Note that $x' = x + d$ where $d \sim \mathcal{N}(\tau f(x,u), \tau \Sigma(x,u))$
- Taylor-series expansion of $V(t+\tau, x')$ around (t,x):

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

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:

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

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

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

Infinite-Horizon Stochastic Optimal Control

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

HJB PDEs for the Optimal Value Function

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

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

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

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

Existence and Uniqueness of HJB PDE 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 value function
- The HJB PDE may not have a classical solution, in which case the optimal value function is not smooth (e.g. bang-bang control)
- The HJB PDE always has a unique viscosity solution which is the optimal value function
- 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, i.e., noise tends to smooth the optimal value function

Example 1: Guessing a Solution for the HJB PDE

- System: $\dot{x}(t) = u(t)$, $|u(t)| \leq 1$, $0 \leq t \leq 1$
- Costs: $\ell(x, u) = 0$ and $\mathfrak{q}(x) = \frac{1}{2}x^2$ 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:

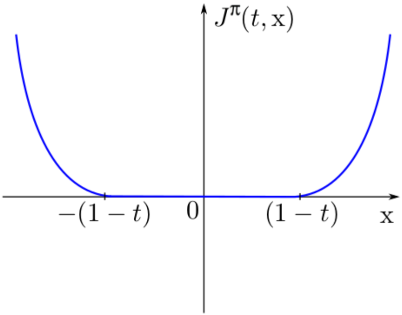
$$\pi(t, x) = -sgn(x) := \begin{cases} -1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x < 0 \end{cases}$$

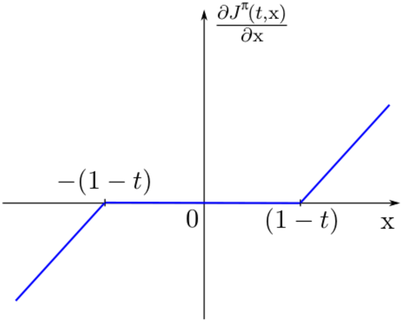
- The value in not smooth: $V^{\pi}(t, x) = \frac{1}{2}\left(\max\{0, |x| - (1-t)\}\right)^2$
- We will verify that this function satisfies the HJB and is therefore indeed the optimal value function

Example 1: Partial Derivative wrt x

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

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

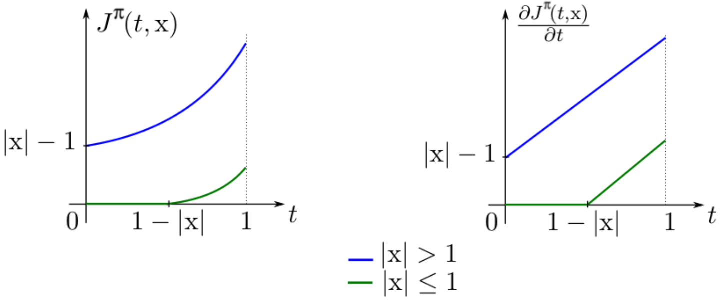




Example 1: Partial Derivative wrt t

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

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



Example 1: Guessing a Solution for the HJB PDE

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

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

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

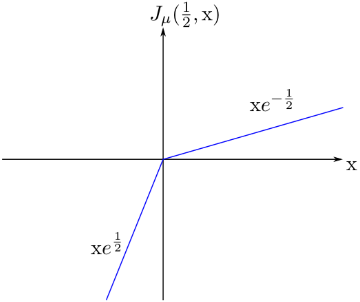
Example 2: HJB PDE without a Classical Solution

- System: $\dot{x}(t) = x(t)u(t)$, $|u(t)| \leq 1$, $0 \leq t \leq 1$
- Costs: $\ell(x, u) = 0$ and $\mathfrak{q}(x) = x$ 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 value function:

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



- The value function is not differentiable wrt x at $x = 0$ and hence does not satisfy the HJB PDE in the classical sense

Optimality Conditions

- The HJB PDE is not a necessary condition for optimality of the continuous-time optimal control problem but it is sufficient

Theorem: HJB PDE Sufficiency

Suppose that $V(t, x)$ is continuously differentiable in t and x and solves the HJB PDE:

$$V(T, x) = \mathfrak{q}(x), \quad \forall x \in \mathcal{X}$$

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

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 PDE for all t and x and is piecewise-continuous in t. Then, under the assumptions on Slide 5, $V(t, x)$ is the unique solution of the HJB PDE and is equal to the optimal value function $V^*(t, x)$, while $\pi^*(t, x)$ is an optimal policy.

More Tractable Problems

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

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

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

$$\begin{aligned}\pi^* &= \arg\min_u \left\{ q + \frac{1}{2}u^T Ru + (a + Bu)^T V_x + \frac{1}{2}\operatorname{tr}(CC^T V_{xx}) \right\} \\ &= -R^{-1}B^T V_x\end{aligned}$$

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

$$H[x, \pi^*, V_x] = q + a^T V_x + \frac{1}{2}\operatorname{tr}(CC^T V_{xx}) - \frac{1}{2}V_x^T BR^{-1}B^T V_x$$

More Tractable Problems (Generalizations)

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

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

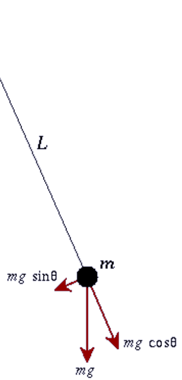
- **Convex-in-control Costs**: $\ell(x,u) = q(x) + \sum_j r(u_j)$ with convex $r(\cdot)$:

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

- **Example**:

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

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**: $\ell(x, u) = q(x) + \frac{r}{2}u^2$
- Optimal value and policy (discounted problem):

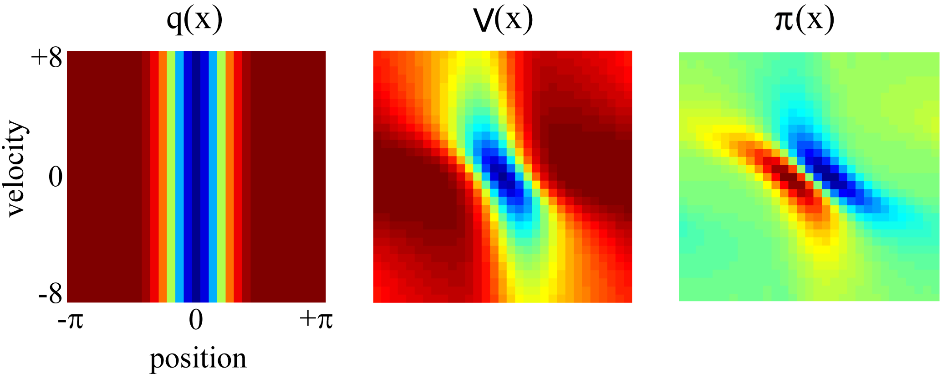
$$\pi^*(x) = -\frac{1}{r}V^*_{x_2}(x)$$

$$\frac{1}{\gamma}V^*(x) = q(x) + x_2 V^*_{x_1}(x) + k\sin(x_1)V^*_{x_2}(x) + \frac{\sigma^2}{2}V^*_{x_2x_2}(x) - \frac{1}{2r}(V^*_{x_2}(x))^2$$

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:

$$V^{(n+1)}(x) = V^{(n)}(x) - \alpha \left(V^{(n)}(x) - \gamma \min_u H[x, u, \nabla_x V^{(n)}(\cdot)] \right), \quad \alpha = 0.01$$

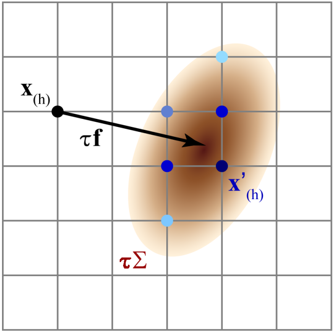


MDP Discretization

- 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 $V^*_{(h)}$ converges to the solution V^* of the continuous problem (even for non-smooth V^*)

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

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