Controlled Diffusions and Hamilton-Jacobi Bellman Equations

Emo Todorov

Applied Mathematics and Computer Science & Engineering

University of Washington

Winter 2014

Continuous-time formulation

Notation and terminology:

$\mathbf{x}(t) \in \mathbb{R}^n$	state vector
$\mathbf{u}(t) \in \mathbb{R}^m$	control vector
$\boldsymbol{\omega}(t) \in \mathbb{R}^k$	Brownian motion (integral of white noise)
$d\mathbf{x} = \mathbf{f}(\mathbf{x}, \mathbf{u})\, dt + G(\mathbf{x}, \mathbf{u})\, d\boldsymbol{\omega}$	continuous-time dynamics
$\Sigma(\mathbf{x}, \mathbf{u}) = G(\mathbf{x}, \mathbf{u})\, G(\mathbf{x}, \mathbf{u})^{\mathsf{T}}$	noise covariance
$\ell(\mathbf{x}, \mathbf{u}) \geq 0$	cost for choosing control $\mathbf{u}$ in state $\mathbf{x}$
$q_{\mathcal{T}}(\mathbf{x}) \geq 0$	(optional) scalar cost at terminal states $\mathbf{x} \in \mathcal{T}$
$\boldsymbol{\pi}(\mathbf{x}) \in \mathbb{R}^m$	control law
$v^{\boldsymbol{\pi}}(\mathbf{x}) \geq 0$	value/cost-to-go function
$\boldsymbol{\pi}^*(\mathbf{x}),\ \mathbf{v}^*(\mathbf{x})$	optimal control law and its value function

Stochastic differential equations and integrals

Ito diffusion / stochastic differential equation (SDE):

$$dx = f(x)\,dt + g(x)\,d\omega$$

This cannot be written as $\dot{x} = f(x) + g(x)\,\dot{\omega}$ because $\dot{\omega}$ does not exist.
The SDE means that the time-integrals of the two sides are equal:

$$x(T) - x(0) = \int_0^T f(x(t))\,dt + \int_0^T g(x(t))\,d\omega(t)$$

The last term is an Ito integral. For an Ito process $y(t)$ *adapted to* $\omega(t)$, i.e. depending on the sample path only up to time t, this integral is

Definition (Ito integral)

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

Replacing $y(t_i)$ with $y((t_{i+1}+t_i)/2)$ yields the Stratonovich integral.

Stochastic chain rule and integration by parts

A twice-differentiable function $a(x)$ of an Ito diffusion $dx = f(x)\,dt + g(x)\,d\omega$ is an Ito process (not necessarily a diffusion) which satisfies:

Lemma (Ito)

$$da(x(t)) = a'(x(t))\,dx(t) + \frac{1}{2}a''(x(t))\,g(x(t))^2\,dt$$

This is the stochastic version of the chain rule.

Stochastic chain rule and integration by parts

A twice-differentiable function $a(x)$ of an Ito diffusion $dx = f(x)\,dt + g(x)\,d\omega$ is an Ito process (not necessarily a diffusion) which satisfies:

Lemma (Ito)

$$da(x(t)) = a'(x(t))\,dx(t) + \frac{1}{2}a''(x(t))\,g(x(t))^2\,dt$$

This is the stochastic version of the chain rule.
There is also a stochastic version of integration by parts:

$$x(T)\,y(T) - x(0)\,y(0) = \int_0^T x(t)\,dy(t) + \int_0^T y(t)\,dx(t) + [x,y]_T$$

The last term (which would be 0 if $x(t)$ or $y(t)$ were differentiable) is

Definition (quadratic covariation)

$$[x,y]_T \triangleq \lim_{\substack{N\to\infty \\ 0=t_0<t_1<\cdots<t_N=T}} \sum_{i=0}^{N-1} \left(x(t_{i+1}) - x(t_i)\right)\left(y(t_{i+1}) - y(t_i)\right)$$

For a diffusion with constant noise amplitude we have $[x,x]_T = g^2 T$.

Forward and backward equations, generator

Let $p(y, s|x, t)$, $s \geq t$ denote the transition probability density under the Ito diffusion $dx = f(x)\,dt + g(x)\,d\omega$. Then p satisfies the following PDEs:

Theorem (Kolmogorov equations)

forward (FP) equation $$\frac{\partial}{\partial s}p = -\frac{\partial}{\partial y}(fp) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(g^2 p\right)$$

backward equation $$-\frac{\partial}{\partial t}p = f\frac{\partial}{\partial x}(p) + \frac{1}{2}g^2\frac{\partial^2}{\partial x^2}(p) = \mathcal{L}\left[p\left(y, s|\cdot, t\right)\right]$$

Forward and backward equations, generator

Let $p(y, s|x, t)$, $s \geq t$ denote the transition probability density under the Ito diffusion $dx = f(x)\,dt + g(x)\,d\omega$. Then p satisfies the following PDEs:

Theorem (Kolmogorov equations)

forward (FP) equation $$\frac{\partial}{\partial s} p = -\frac{\partial}{\partial y}(fp) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(g^2 p\right)$$

backward equation $$-\frac{\partial}{\partial t} p = f\frac{\partial}{\partial x}(p) + \frac{1}{2} g^2 \frac{\partial^2}{\partial x^2}(p) = \mathcal{L}\left[p(y, s|\cdot, t)\right]$$

The operator $\mathcal{L}$ which computes expected directional derivatives is called the *generator* of the stochastic process. It satisfies (in the vector case):

Theorem (generator)

$$\mathcal{L}[v(\cdot)](\mathbf{x}) \triangleq \lim_{\Delta \to 0} \frac{E^{\mathbf{x}(0)=\mathbf{x}}\left[v(\mathbf{x}(\Delta))\right] - v(\mathbf{x})}{\Delta} = \mathbf{f}(\mathbf{x})^T v_{\mathbf{x}}(\mathbf{x}) + \frac{1}{2}\operatorname{tr}\left(\Sigma(\mathbf{x})\, v_{\mathbf{xx}}(\mathbf{x})\right)$$

Discretizing the time axis

Consider the explicit Euler discretization with time step Δ:

$$\mathbf{x}(t+\Delta) = \mathbf{x}(t) + \Delta \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)) + \sqrt{\Delta} G(\mathbf{x}(t), \mathbf{u}(t)) \boldsymbol{\varepsilon}(t)$$

where $\boldsymbol{\varepsilon}(t) \sim N(0, I)$. The term $\sqrt{\Delta}$ appears because the variance grows linearly with time.

Thus the transition probability $p(\mathbf{x}'|\mathbf{x}, \mathbf{u})$ is Gaussian, with mean $\mathbf{x} + \Delta \mathbf{f}(\mathbf{x}, \mathbf{u})$ and covariance matrix $\Delta \Sigma(\mathbf{x}, \mathbf{u})$. The one-step cost is $\Delta \ell(\mathbf{x}, \mathbf{u})$.

Now we can apply the Bellman equation (in the finite horizon setting):

$$\begin{aligned} v(\mathbf{x}, t) &= \min_{\mathbf{u}} \left\{ \Delta \ell(\mathbf{x}, \mathbf{u}) + E_{\mathbf{x}' \sim p(\cdot|\mathbf{x}, \mathbf{u})} \left[v\left(\mathbf{x}', t+\Delta\right) \right] \right\} = \\ & \min_{\mathbf{u}} \left\{ \Delta \ell(\mathbf{x}, \mathbf{u}) + E_{\mathbf{d} \sim N(\Delta \mathbf{f}(\mathbf{x}, \mathbf{u}),\, \Delta \Sigma(\mathbf{x}, \mathbf{u}))} \left[v(\mathbf{x} + \mathbf{d}, t+\Delta) \right] \right\} \end{aligned}$$

Next we use the Taylor-series expansion of v ...

Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned} v(\mathbf{x}+\mathbf{d}, t+\Delta) &= v(\mathbf{x}, t) + \Delta v_t(\mathbf{x}, t) + o\left(\Delta^2\right) + \\ &+ \mathbf{d}^\mathsf{T} v_\mathbf{x}(\mathbf{x}, t) + \frac{1}{2}\mathbf{d}^\mathsf{T} v_{\mathbf{xx}}(\mathbf{x}, t)\,\mathbf{d} + o\left(\mathbf{d}^3\right) \end{aligned}$$

Using the fact that $E\left[\mathbf{d}^\mathsf{T} M \mathbf{d}\right] = \mathrm{tr}\left(\mathrm{cov}\left[\mathbf{d}\right] M\right) + o\left(\Delta^2\right)$, the expectation is

$$\begin{aligned} E_\mathbf{d}\left[v(\mathbf{x}+\mathbf{d}, t+\Delta)\right] &= v(\mathbf{x}, t) + \Delta v_t(\mathbf{x}, t) + o\left(\Delta^2\right) + \\ &+ \Delta \mathbf{f}(\mathbf{x}, \mathbf{u})^\mathsf{T} v_\mathbf{x}(\mathbf{x}, t) + \frac{\Delta}{2}\mathrm{tr}\left(\Sigma(\mathbf{x}, \mathbf{u})\, v_{\mathbf{xx}}(\mathbf{x}, t)\right) \end{aligned}$$

Hamilton-Jacobi-Bellman (HJB) equation

$$
\begin{aligned}
v(\mathbf{x}+\mathbf{d}, t+\Delta) &= v(\mathbf{x}, t) + \Delta v_t(\mathbf{x}, t) + o\left(\Delta^2\right) + \\
&+ \mathbf{d}^\mathsf{T} v_\mathbf{x}(\mathbf{x}, t) + \frac{1}{2}\mathbf{d}^\mathsf{T} v_{\mathbf{xx}}(\mathbf{x}, t)\,\mathbf{d} + o\left(\mathbf{d}^3\right)
\end{aligned}
$$

Using the fact that $E\left[\mathbf{d}^\mathsf{T} M \mathbf{d}\right] = \operatorname{tr}(\operatorname{cov}[\mathbf{d}]\,M) + o(\Delta^2)$, the expectation is

$$
\begin{aligned}
E_\mathbf{d}\left[v(\mathbf{x}+\mathbf{d}, t+\Delta)\right] &= v(\mathbf{x}, t) + \Delta v_t(\mathbf{x}, t) + o\left(\Delta^2\right) + \\
&+ \Delta \mathbf{f}(\mathbf{x}, \mathbf{u})^\mathsf{T} v_\mathbf{x}(\mathbf{x}, t) + \frac{\Delta}{2} \operatorname{tr}\left(\Sigma(\mathbf{x}, \mathbf{u})\, v_{\mathbf{xx}}(\mathbf{x}, t)\right)
\end{aligned}
$$

Substituting in the Bellman equation,

$$
v(\mathbf{x}, t) = \min_\mathbf{u} \left\{ \begin{array}{l} \Delta \ell(\mathbf{x}, \mathbf{u}) + v(\mathbf{x}, t) + \Delta v_t(\mathbf{x}, t) + o(\Delta^2) + \\ + \Delta \mathbf{f}(\mathbf{x}, \mathbf{u})^\mathsf{T} v_\mathbf{x}(\mathbf{x}, t) + \frac{\Delta}{2} \operatorname{tr}\left(\Sigma(\mathbf{x}, \mathbf{u})\, v_{\mathbf{xx}}(\mathbf{x}, t)\right) \end{array} \right\}
$$

Simplifying, dividing by Δ and taking $\Delta \to 0$ yields the HJB equation

$$
-v_t(\mathbf{x}, t) = \min_\mathbf{u} \left\{ \ell(\mathbf{x}, \mathbf{u}) + \mathbf{f}(\mathbf{x}, \mathbf{u})^\mathsf{T} v_\mathbf{x}(\mathbf{x}) + \frac{1}{2} \operatorname{tr}\left(\Sigma(\mathbf{x}, \mathbf{u})\, v_{\mathbf{xx}}(\mathbf{x})\right) \right\}
$$

HJB equations for different problem formulations

Definition (Hamiltonian)

$$H\left[\mathbf{x},\mathbf{u},v\left(\cdot\right)\right] \triangleq \ell\left(\mathbf{x},\mathbf{u}\right) + \mathbf{f}\left(\mathbf{x},\mathbf{u}\right)^{\mathsf{T}} v_{\mathbf{x}}\left(\mathbf{x}\right) + \frac{1}{2}\operatorname{tr}\left(\Sigma\left(\mathbf{x},\mathbf{u}\right) v_{\mathbf{xx}}\left(\mathbf{x}\right)\right) = \ell + \mathcal{L}\left[v\right]$$

The HJB equations for the optimal cost-to-go v^* are

Theorem (HJB equations)

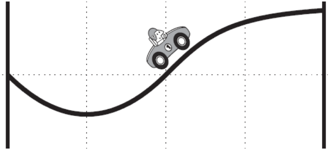
first exit	$0 = \min_{\mathbf{u}} H\left[\mathbf{x},\mathbf{u},v^*\left(\cdot\right)\right]$	$v^*\left(\mathbf{x} \in \mathcal{T}\right) = q_{\mathcal{T}}\left(\mathbf{x}\right)$
finite horizon	$-v_t^*\left(\mathbf{x},t\right) = \min_{\mathbf{u}} H\left[\mathbf{x},\mathbf{u},v^*\left(\cdot,t\right)\right]$	$v^*\left(\mathbf{x},T\right) = q_{\mathcal{T}}\left(\mathbf{x}\right)$
discounted	$\frac{1}{\tau} v^*\left(\mathbf{x}\right) = \min_{\mathbf{u}} H\left[\mathbf{x},\mathbf{u},v^*\left(\cdot\right)\right]$	
average	$c = \min_{\mathbf{u}} H\left[\mathbf{x},\mathbf{u},\widetilde{v}^*\left(\cdot\right)\right]$	

Discounted cost-to-go: $v^{\pi}\left(\mathbf{x}\right) = E\int_0^{\infty} \exp\left(-t/\tau\right)\ell\left(\mathbf{x}\left(t\right),\mathbf{u}\left(t\right)\right)dt.$

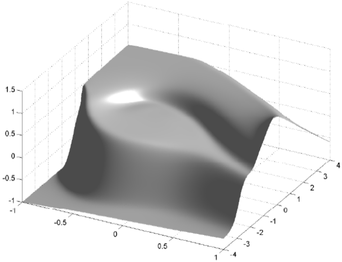
Existence and uniqueness of solutions

- The HJB equation has at most one classic solution (i.e. a function which satisfies the PDE everywhere.)
- If a classic solution exists then it is the optimal cost-to-go function.
- The HJB equation may not have a classic solution; in that case the optimal cost-to-go function is non-smooth (e.g. bang-bang control.)
- The HJB equation always has a unique viscosity solution which is the optimal cost-to-go function.
- Approximation schemes based on MDP discretization (see below) are guaranteed to converge to the unique viscosity solution / optimal cost-to-go function.
- 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.

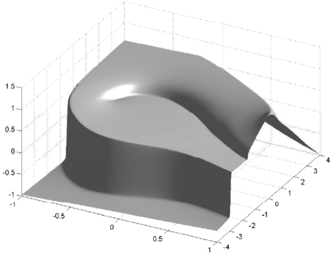
Example of noise smoothing



Noisy dynamics



Deterministic dynamics



Tassa and Erez (2007)

More tractable problems

Consider a restricted family of problems with dynamics and cost

$$\begin{aligned} d\mathbf{x} &= \left(\mathbf{a}\left(\mathbf{x}\right) + B\left(\mathbf{x}\right)\mathbf{u}\right)dt + C\left(\mathbf{x}\right)d\boldsymbol{\omega} \\ \ell\left(\mathbf{x},\mathbf{u}\right) &= q\left(\mathbf{x}\right) + \frac{1}{2}\mathbf{u}^{\mathsf{T}}R\left(\mathbf{x}\right)\mathbf{u} \end{aligned}$$

For such problems the Hamiltonian can be minimized analytically w.r.t. $\mathbf{u}$. Suppressing the dependence on $\mathbf{x}$ for clarity, we have

$$\min_{\mathbf{u}} H = \min_{\mathbf{u}} \left\{ q + \frac{1}{2}\mathbf{u}^{\mathsf{T}}R\mathbf{u} + \left(\mathbf{a} + B\mathbf{u}\right)^{\mathsf{T}} v_{\mathbf{x}} + \frac{1}{2}\operatorname{tr}\left(CC^{\mathsf{T}}v_{\mathbf{xx}}\right) \right\}$$

The minimum is achieved at $\mathbf{u}^* = -R^{-1}B^{\mathsf{T}}v_{\mathbf{x}}$ and the result is

$$\min_{\mathbf{u}} H = q + \mathbf{a}^{\mathsf{T}}v_{\mathbf{x}} + \frac{1}{2}\operatorname{tr}\left(CC^{\mathsf{T}}v_{\mathbf{xx}}\right) - \frac{1}{2}v_{\mathbf{x}}^{\mathsf{T}}BR^{-1}B^{\mathsf{T}}v_{\mathbf{x}}$$

Thus the HJB equations become 2nd-order quadratic PDEs, no longer involving the min operator.

More tractable problems (generalizations)

- Allowing control-multiplicative noise:

$$\Sigma(\mathbf{x}, \mathbf{u}) = C_0(\mathbf{x})\, C_0(\mathbf{x})^\mathsf{T} + \sum_{k=1}^{K} C_k(\mathbf{x})\, \mathbf{u}\mathbf{u}^\mathsf{T} C_k(\mathbf{x})^\mathsf{T}$$

  The optimal control law becomes:

$$\mathbf{u}^* = -\left(R + \sum_{k=1}^{K} C_k^\mathsf{T} v_{\mathbf{x}\mathbf{x}} C_k\right)^{-1} B^\mathsf{T} v_{\mathbf{x}}$$

More tractable problems (generalizations)

- Allowing control-multiplicative noise:

$$\Sigma(\mathbf{x},\mathbf{u}) = C_0(\mathbf{x})\,C_0(\mathbf{x})^\mathsf{T} + \sum_{k=1}^{K} C_k(\mathbf{x})\,\mathbf{u}\mathbf{u}^\mathsf{T} C_k(\mathbf{x})^\mathsf{T}$$

The optimal control law becomes:

$$\mathbf{u}^* = -\left(R + \sum_{k=1}^{K} C_k^\mathsf{T} v_{\mathbf{xx}} C_k\right)^{-1} B^\mathsf{T} v_{\mathbf{x}}$$

- Allowing more general control costs:

$$\ell(\mathbf{x},\mathbf{u}) = q(\mathbf{x}) + \sum_i r(u_i)\,, \quad r : \text{convex}$$

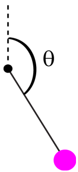
The optimal control law becomes:

$$\mathbf{u}^* = \arg\min_{\mathbf{u}} \left\{\sum_i r(u_i) + \mathbf{u}^\mathsf{T} B^\mathsf{T} v_{\mathbf{x}}\right\} = (r')^{-1}\left(-B^\mathsf{T} v_{\mathbf{x}}\right)$$

$$r(u) = s\int_0^{|u|} \operatorname{atanh}\left(\frac{w}{u_{\max}}\right) dw \implies \mathbf{u}^* = u_{\max}\tanh\left(-s^{-1} B^\mathsf{T} v_{\mathbf{x}}\right)$$

Pendulum example

$$\ddot{\theta} = k \sin(\theta) + u$$



First-order form:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix}$$

$$\mathbf{a}(\mathbf{x}) = \begin{bmatrix} x_2 \\ k \sin(x_1) \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Stochastic dynamics:

$$d\mathbf{x} = \mathbf{a}(\mathbf{x})\,dt + B\,(u\,dt + \sigma d\omega)$$

Cost and optimal control:

$$\begin{aligned} \ell(\mathbf{x}, u) &= q(\mathbf{x}) + \frac{r}{2} u^2 \\ u^*(\mathbf{x}) &= -r^{-1} v_{x_2}(\mathbf{x}) \end{aligned}$$

HJB equation (discounted):

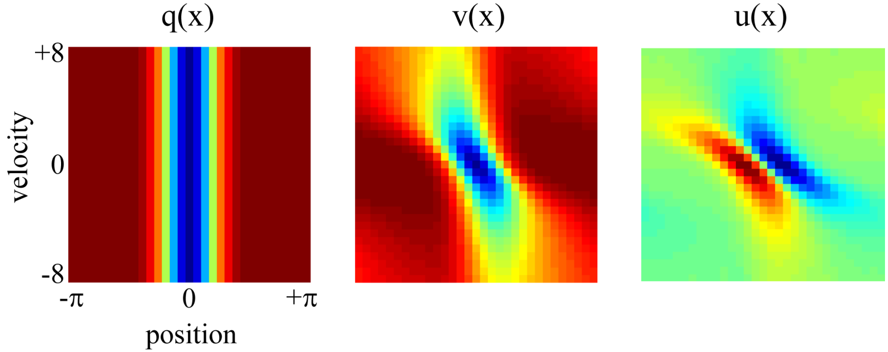
$$\begin{aligned} \frac{1}{\tau} v &= q + x_2 v_{x_1} + k \sin(x_1)\, v_{x_2} \\ &\quad + \frac{\sigma^2}{2} v_{x_2 x_2} - \frac{1}{2r} v_{x_2}^2 \end{aligned}$$

Pendulum example continued

Parameters: $k = \sigma = r = 1, \quad \tau = 0.3, \quad q = 1 - \exp\left(-2\theta^2\right), \quad \beta = 0.99$

Dicretize state space, approximate derivatives via finite differences, iterate:

$$v^{(n+1)} = \beta v^{(n)} + (1-\beta)\,\tau \min_u H^{(n)}$$

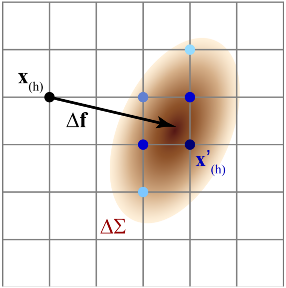


MDP discretization

Define discrete state and control spaces $\mathcal{X}_{(h)} \subset \mathbb{R}^n$, $\mathcal{U}_{(h)} \subset \mathbb{R}^m$ and discrete time step $\Delta_{(h)}$, where h is a "coarseness" parameter and $h \to 0$ corresponds to infinitely dense discretization. Construct $p_{(h)}\left(\mathbf{x}'_{(h)} | \mathbf{x}_{(h)}, \mathbf{u}_{(h)}\right)$ s.t.

Definition (local consistency)

$$\begin{aligned} \mathbf{d} &\triangleq \mathbf{x}'_{(h)} - \mathbf{x}_{(h)} \\ E[\mathbf{d}] &= \Delta_{(h)} \mathbf{f}(\mathbf{x}_{(h)}, \mathbf{u}_{(h)}) + o(\Delta_{(h)}) \\ \mathrm{cov}[\mathbf{d}] &= \Delta_{(h)} \Sigma(\mathbf{x}_{(h)}, \mathbf{u}_{(h)}) + o(\Delta_{(h)}) \end{aligned}$$



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

Constructing the MDP

For each $\mathbf{x}_{(h)}, \mathbf{u}_{(h)}$ choose vectors $\{\mathbf{v}_i\}_{i=1\cdots K}$ such that all possible next states are $\mathbf{x}'_{(h)} = \mathbf{x}_{(h)} + h\mathbf{v}_i$. Then compute $p^i_{(h)} = p_{(h)}\left(\mathbf{x}_{(h)} + h\mathbf{v}_i | \mathbf{x}_{(h)}, \mathbf{u}_{(h)}\right)$ as:

- Find w_i, y_i s.t.

$$
\begin{aligned}
\textstyle\sum_i w_i \mathbf{v}_i &= \mathbf{f} \\
\textstyle\sum_i y_i \mathbf{v}_i \mathbf{v}_i^{\mathsf{T}} &= \Sigma \\
\textstyle\sum_i y_i \mathbf{v}_i &= 0 \\
\textstyle\sum_i w_i &= 1, \quad w_i \geq 0 \\
\textstyle\sum_i y_i &= 1, \quad y_i \geq 0
\end{aligned}
$$

and set

$$
\begin{aligned}
p^i_{(h)} &= \frac{h w_i + y_i}{h+1} \\
\Delta_{(h)} &= \frac{h^2}{h+1}
\end{aligned}
$$

Constructing the MDP

For each $\mathbf{x}_{(h)}, \mathbf{u}_{(h)}$ choose vectors $\{\mathbf{v}_i\}_{i=1\cdots K}$ such that all possible next states are $\mathbf{x}'_{(h)} = \mathbf{x}_{(h)} + h\mathbf{v}_i$. Then compute $p^i_{(h)} = p_{(h)}\left(\mathbf{x}_{(h)} + h\mathbf{v}_i | \mathbf{x}_{(h)}, \mathbf{u}_{(h)}\right)$ as:

- Find w_i, y_i s.t.

$$\begin{aligned} \textstyle\sum_i w_i \mathbf{v}_i &= \mathbf{f} \\ \textstyle\sum_i y_i \mathbf{v}_i \mathbf{v}_i^{\mathsf{T}} &= \Sigma \\ \textstyle\sum_i y_i \mathbf{v}_i &= 0 \\ \textstyle\sum_i w_i &= 1, \quad w_i \geq 0 \\ \textstyle\sum_i y_i &= 1, \quad y_i \geq 0 \end{aligned}$$

  and set

$$\begin{aligned} p^i_{(h)} &= \frac{h w_i + y_i}{h+1} \\ \Delta_{(h)} &= \frac{h^2}{h+1} \end{aligned}$$

- Set $\Delta_{(h)} = h$ and minimize

$$\left\| \Sigma - h \textstyle\sum_i p^i_{(h)} \left(\mathbf{v}_i - \mathbf{f}\right)\left(\mathbf{v}_i - \mathbf{f}\right)^{\mathsf{T}} \right\|^2$$

  s.t.

$$\begin{aligned} \textstyle\sum_i p^i_{(h)} \mathbf{v}_i &= \mathbf{f} \\ \textstyle\sum_i p^i_{(h)} &= 1, \quad p^i_{(h)} \geq 0 \end{aligned}$$

Constructing the MDP

For each $\mathbf{x}_{(h)}, \mathbf{u}_{(h)}$ choose vectors $\{\mathbf{v}_i\}_{i=1\cdots K}$ such that all possible next states are $\mathbf{x}'_{(h)} = \mathbf{x}_{(h)} + h\mathbf{v}_i$. Then compute $p^i_{(h)} = p_{(h)}\left(\mathbf{x}_{(h)} + h\mathbf{v}_i | \mathbf{x}_{(h)}, \mathbf{u}_{(h)}\right)$ as:

- Find w_i, y_i s.t.

$$\begin{aligned} \textstyle\sum_i w_i \mathbf{v}_i &= \mathbf{f} \\ \textstyle\sum_i y_i \mathbf{v}_i \mathbf{v}_i^\mathsf{T} &= \Sigma \\ \textstyle\sum_i y_i \mathbf{v}_i &= 0 \\ \textstyle\sum_i w_i &= 1, \quad w_i \geq 0 \\ \textstyle\sum_i y_i &= 1, \quad y_i \geq 0 \end{aligned}$$

  and set

$$\begin{aligned} p^i_{(h)} &= \frac{h w_i + y_i}{h+1} \\ \Delta_{(h)} &= \frac{h^2}{h+1} \end{aligned}$$

- Set $\Delta_{(h)} = h$ and minimize

$$\left\| \Sigma - h \textstyle\sum_i p^i_{(h)} \left(\mathbf{v}_i - \mathbf{f}\right) \left(\mathbf{v}_i - \mathbf{f}\right)^\mathsf{T} \right\|^2$$

  s.t.

$$\begin{aligned} \textstyle\sum_i p^i_{(h)} \mathbf{v}_i &= \mathbf{f} \\ \textstyle\sum_i p^i_{(h)} &= 1, \quad p^i_{(h)} \geq 0 \end{aligned}$$

- Set $\Delta_{(h)} = h$ and

$$p^i_{(h)} \propto N\left(\mathbf{x}_{(h)} + h\mathbf{v}_i;\ h\mathbf{f},\ h\Sigma\right)$$