Optimal and Learning Control for Autonomous Robots

Lecture 3

ADRL

Jonas Buchli
Agile & Dexterous Robotics Lab

FNSNF

ETH Zürich

Class logistics

Office hours: Thu, 18-19 Room: ML J37.1

First office hour March 5

Erratum Script

p14

$$\frac{dV^*}{dt} = V_t^* + V_{\mathbf{x}}^{*T}\mathbf{f} + \frac{1}{2}\mathrm{Tr}\left[V_{\mathbf{xx}}^* E[(\mathbf{f}+\mathbf{Bw})(\mathbf{f}+\mathbf{Bw})^T]\Delta t\right]. \qquad (1.55)$$

A D R L

ETH Zürich

Lecture 3 Goals

- ★Continuous time optimal control problem
- ★Value function and optimal value function
- ★Hamilton Jacobi Bellman Equation

L2 Recap

Discrete optimal control problem

finite time, deterministic

Find control $u_k^* = \mu^*(k, x_k)$ minimizing

control (input) policy

$$J = \alpha^N \Phi(x_N) + \sum_{k=0}^{N-1} \alpha^k L_k(x_k, u_k)$$

Given constraints

$$x_{n+1} = f_n(x_n, u_n)$$

Goal: Optimal policy

$$\mu^* = \arg\min_u J$$

A D R L

ETH Zürich

The backwards nature of the value function

Bellman equation

$$V^\mu(n+1, x_{n+1})$$

$$V^\mu(n, \mathbf{x}) = L_n(\mathbf{x}, \mathbf{u}_n) + \alpha V^\mu\left(n+1, f_n\left(\mathbf{x}, \mathbf{u}_n\right)\right)$$

final condition $V^\mu(N, x) = \Phi(x)$

If I want to know V at given time n, need to start with final value and compute backwards

Optimal policy

optimal value function

$$V^*(n,x) \leq V^\mu(n,x) \qquad \forall n,x$$

Remember:V is based on cost ⇒ minimize

equivalent notation

$$V^*(n,x) = \min_\mu V^\mu(n,x) \qquad \forall n,x$$

Optimal policy is the one that minimizes RHS

$$\mu^* = \{\mathbf{u}_n^*, \ldots, \mathbf{u}_{N-1}^*\} = \arg\min_\mu V^\mu(n,\mathbf{x}) \qquad \forall n: 0,\ldots,N-1$$

substitute Bellman Equation into V^μ

$$V^\mu(n,\mathbf{x}) = L_n(\mathbf{x},\mathbf{u}_n) + \alpha V^\mu\left(n+1, f_n\left(\mathbf{x},\mathbf{u}_n\right)\right)$$

$$V^*(n,\mathbf{x}) = \min_{\mathbf{u}_n}\left[L_n(\mathbf{x},\mathbf{u}_n) + \alpha V^*\left(n+1,\mathbf{f}_n\left(\mathbf{x},\mathbf{u}_n\right)\right)\right]$$

Optimal Bellman Equation

ADRL

ETH Zürich

Optimal Bellman Equation

$$V^*(n, \mathbf{x}) = \min_{\mathbf{u}_n} \left[L_n(\mathbf{x}, \mathbf{u}_n) + \alpha V^*\left(n+1, \mathbf{f}_n\left(\mathbf{x}, \mathbf{u}_n\right)\right)\right]$$

★ Optimal Bellman Eq. computes
Optimal Value function

if u continuous: $$\frac{\partial}{\partial \mathbf{u}_n}\left[L_n(\mathbf{x}, \mathbf{u}_n) + \alpha V^*\left(n, \mathbf{f}_n\left(\mathbf{x}, \mathbf{u}_n\right)\right)\right] = 0$$

- Bellman Equation requires working 'backwards in time' / from end to start
- Bellman Equation allows to find optimal solution one step at a time
- ... whereas Value function requires optimization of the whole control sequence at once

$$V^\mu(n, x) = \alpha^{N-n}\Phi(x_N) + \sum_{k=n}^{N-1} \alpha^{k-n} L_k(x_k, u_k)$$

A D R L

ETH Zürich

Stochastic system

Additive:

$$x_{n+1} = f(x_n, u_n) + w_n$$

Additive noise

$$w_n \sim P_w(\cdot \mid x_n, u_n)$$

Conditional Probability Distribution
'function of state and control'

General:

$$x_{n+1} = x'$$

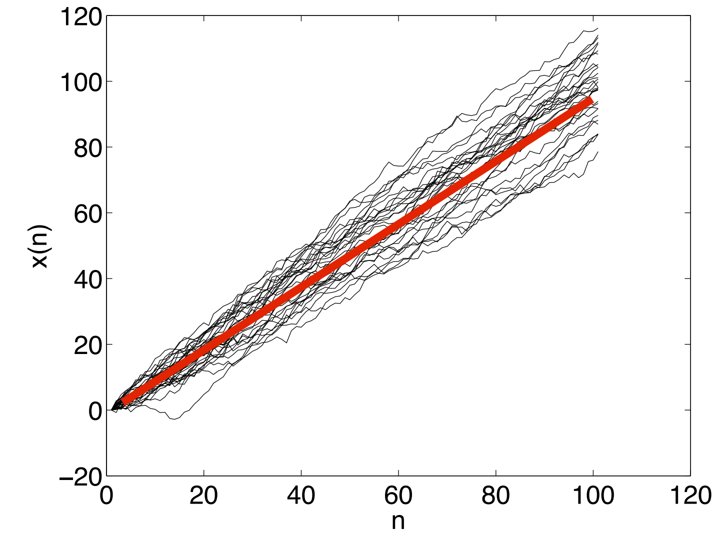
General stochastic dynamics

$$x' \sim P_f(\cdot \mid x_n, u_n)$$

A D R L

ETH Zürich

Expectation



Expected value of x:

Discrete states

$$E(x) = \sum_i P(x_i) x_i \approx \sum_s \frac{1}{N} x_s$$

$$x_s \sim P(x)$$

$$\sum_i P(x_i) = 1$$

$$P(x) \geq 0$$

'weighted average'

Continuous states

$$E(x) = \int p(x) x dx \approx \sum_s \frac{1}{N} x_s$$

$$\int p(x) dx = 1$$

Mathematical expectation itself is not a random variable!

Numerical approximation is a random variable.

Cost in stochastic system?

★ Even if we keep u fixed, path x(0...N) will be different each time

➡ thus so is cost

So how to minimize the cost???

✳ Idea: minimize 'in average', i.e. find best solution in average

- average = expected value

➡ minimize expected cost

Cost in stochastic problem

Expected cost:

$$J = E\left[\alpha^N \Phi(x_N) + \sum_{k=0}^{N-1} \alpha^k L_k(x_k, u_k)\right]$$

Cost is weighted average of all possible costs
Weight = probability of outcome

In stochastic optimal control: Can not optimize outcome, but only the average outcome (expected outcome). The actual cost in a 'rollout' will always be different from the expected cost.

A D R L

ETH Zürich

Value functions

Value function for policy

$$V^{\mu}(n,x) = E\left[\alpha^{N-n}\Phi(x_N) + \sum_{k=n}^{N-1} \alpha^{k-n} L_k(x_k,u_k)\right]$$

Optimal value function

$$V^{*}(n,x) = \min_{\mu} E\left[\alpha^{N}\Phi(x_N) + \sum_{k=n}^{N-1} \alpha^{k-n} L_k(x_k,u_k)\right]$$

Optimal policy

$$\mu^{*} = \arg\min_{\mu} E\left[\alpha^{N-n}\Phi(x_N) + \sum_{k=n}^{N-1} \alpha^{k-n} L_k(x_k,u_k)\right]$$

Value function and optimal policy are deterministic (but a function of probability distribution P)

Bellman equation

$$E(x) = \int p(x) x dx$$

sum over all x'

$$V^\mu(n,x) = L_n(x,u_n) + E_{x' \sim P_f(.|x,u_n)}\left[V^\mu(n+1,x')\right]$$

Optimal Bellman Equation

$$V^*(n,x) = \min_{u_n}\left[L_n(x,u_n) + E_{x' \sim P_f(\cdot|x,u_n)}\left[V^*(n+1,x')\right]\right]$$

Optimal Control

$$u^*(n) = \arg\min_{u_n}\left[L_n(x,u_n) + E_{x' \sim P_f(\cdot|x,u_n)}\left[V^*(n+1,f_n(x,u_n))\right]\right]$$

x' conditioned on x(n) and u(n)

optimal control is deterministic, not a random variable!

EOF Recap

L3

Calculus Notes (1)

function vs functional

function: $y = f(x) \quad x, y \in \mathbb{R}$

functional: $y = g(f) \quad f \in \mathcal{V}, y \in \mathbb{R}$ $\qquad \mathcal{V}$ vector space

functional: mapping from a vector (space) to a scalar

Remember: the ‘parametrization of a vector’ can be ‘continuous’: a continuous function is element of a vector space (cf. Fourier analysis)

$$J = f(x(t), u(t))$$

Calculus Notes (II)

total vs. partial derivative

$\frac{\partial}{\partial t}$ partial

$$f(y,t) = y + g(t)$$

total $\frac{d}{\mathrm{d}t}$

$$\frac{\partial}{\partial t} f(y,t) \qquad \frac{\partial}{\partial t} g(t) \qquad \frac{d}{\mathrm{d}t} g(t) \qquad \frac{d}{\mathrm{d}t} f(y,t)$$

$y = x(t)$

$$\frac{\partial}{\partial t} f(x(t),t) \qquad \frac{\partial}{\partial t} g(t) \qquad \frac{\partial}{\partial t} x(t) + \frac{\partial}{\partial t} g(t) \qquad \frac{d}{\mathrm{d}t} f(x(t),t)$$

$$\frac{d}{\mathrm{d}t} f(y,z,t) = \frac{\partial}{\partial y} f \frac{\partial}{\partial t} y + \frac{\partial}{\partial z} f \frac{\partial}{\partial t} z + \frac{\partial}{\partial q} f \frac{\partial}{\partial t} q(t)$$

time

states	discrete (time)	continuous (time)
discrete	gridworlds Markov chains	Petrinets Discrete Event Systems
continuous	$x_{n+1} = f(x_n, n)$ Map Difference equations $x_{n+1} = x_n + f_\Delta(x_n, n)$	Differential equations

L1 example - TSP

'Robots'
Newtonian Physics

Continuous time system

System dynamics

$$\dot{\mathbf{x}}(t) = \mathbf{f}_t(\mathbf{x}(t), \mathbf{u}(t))$$

Cost

$$J = e^{-\beta(t_f - t_0)}\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} e^{-\beta(t-t_0)} L(\mathbf{x}(t), \mathbf{u}(t))dt$$

$0 \leq \beta$ discount / decay rate

'exponential decay'

Continuous time optimal control problem

Find control $u^*(t) = \mu^*(t, x(t))$ minimizing

control (input) policy

$$J = e^{-\beta(t_f - t_0)}\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} e^{-\beta(t-t_0)} L(\mathbf{x}(t), \mathbf{u}(t))dt$$

Given constraints

$$\dot{\mathbf{x}}(t) = \mathbf{f}_t(\mathbf{x}(t), \mathbf{u}(t))$$

Goal: Optimal policy

$$\mu^* = \arg\min_u J$$

Value function

$$J = e^{-\beta(t_f - t_0)}\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} e^{-\beta(t - t_0)} L(\mathbf{x}(t), \mathbf{u}(t))dt$$

Value function

$$V^{\mu}(t, \mathbf{x}) = e^{-\beta(t_f - t)}\Phi(\mathbf{x}(t_f)) + \int_{t}^{t_f} e^{-\beta(t' - t)} L(\mathbf{x}(t'), \mathbf{u}^{\mu}(t'))dt'$$

Effect of final cost becomes more prominent, for later time (increasing t)

Cost equals Value function at time 0

$$J = V(0, x_0)$$

A D R L

ETH Zürich

Optimal Value & Policy

optimal value function

equivalent notation

$$V^*(t,x) \leq V^\mu(t,x) \qquad \forall t \in [t_0, t_f], x$$

Remember: V is based on cost ⇒ minimize

$$V^*(t,x) = \min_\mu V^\mu(t,x) \qquad \forall t \in [t, t_f], x$$

Optimal policy is the one that minimizes RHS

$$\mu^* = u(t) = \arg\min_\mu V^\mu(t,x) \qquad t \in [t_0, t_f]$$

- ★ Discrete system (L2) could use the Bellman equation to find V...
- ★ ... is there an equivalent for continuous time problems?
- ➡ Hamilton Jacobi Bellman Equation

Optimal Value Function

$$V^*(t, \mathbf{x}) = e^{-\beta(t_f - t)}\Phi(\mathbf{x}^*(t_f)) + \int_t^{t_f} e^{-\beta(t' - t)} L(\mathbf{x}^*(t'), \mathbf{u}^*(t'))dt'$$

Hamilton-Jacobi-Bellman Equation

Informal Derivation

Discretize

$$\delta t = \frac{t_f - t_0}{N}$$

$$\alpha = e^{-\beta \delta t} \approxeq 1 - \beta \delta t$$

$$t_n = t_0 + n\delta t$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \cdot \delta t$$

$$V^{\tilde{\mu}}(t, \mathbf{x}) = e^{-\beta(t_f - t)}\Phi(\mathbf{x}(t_f)) + \int_t^{t_f} e^{-\beta(t'-t)} L(\mathbf{x}(t'), \mathbf{u}^{\mu}(t'))\, dt'$$

$$\tilde{V}(t_n, \mathbf{x}) = \alpha^{N-n}\Phi(\mathbf{x}_N) + \sum_{k=n}^{N-1} \alpha^{k-n} L(\mathbf{x}_k, \mathbf{u}_k)\delta t$$

HJB Informal Derivation cont'd

$$\tilde{V}(t_n, \mathbf{x}) = \alpha^{N-n}\Phi(\mathbf{x}_N) + \sum_{k=n}^{N-1} \alpha^{k-n} L(\mathbf{x}_k, \mathbf{u}_k)\delta t$$

Use results from discrete optimal control (L2):

$$\tilde{V}^*(t_n, \mathbf{x}) = \min_{\mathbf{u}\in\mathbf{U}}\{L(\mathbf{x}, \mathbf{u})\delta t + \alpha\tilde{V}^*(t_{n+1}, \mathbf{x}_{n+1})\}$$

Taylor series of RHS:

For small δt

$$\tilde{V}^*(t_{n+1}, \mathbf{x}_{n+1}) = \tilde{V}^*(t_n + \delta t, \mathbf{x} + \mathbf{f}(\mathbf{x}, u)\delta t)$$

$$= \tilde{V}^*(t_n, \mathbf{x}) + \Delta\tilde{V}^*(t_n, \mathbf{x})$$

$$= \tilde{V}^*(t_n, \mathbf{x}) + \frac{\partial\tilde{V}^*(t_n, \mathbf{x})}{\partial t}\delta t + \left(\frac{\partial\tilde{V}^*(t_n, \mathbf{x})}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x}, \mathbf{u})\delta t$$

A D R L

ETH Zürich

$$\tilde{V}^*(t_{n+1}, \mathbf{x}_{n+1}) = \tilde{V}^*(t_n, \mathbf{x}) + \frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial t}\delta t + \left(\frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x}, \mathbf{u})\delta t$$

plug into

$$\tilde{V}^*(t_n, \mathbf{x}) = \min_{\mathbf{u}\in\mathbf{U}}\{L(\mathbf{x}, \mathbf{u})\delta t + \alpha \tilde{V}^*(t_{n+1}, \mathbf{x}_{n+1})\}$$

$$\tilde{V}^*(t_n, \mathbf{x}) = \min_{\mathbf{u}\in\mathbf{U}}\left\{L(\mathbf{x}, \mathbf{u})\delta t + \alpha\left[\tilde{V}^*(t_n, \mathbf{x}) + \frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial t}\delta t + \left(\frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x}, \mathbf{u})\delta t\right]\right\}$$

not dependent of u, take out of min op

$$(1-\alpha)\tilde{V}^*(t_n, \mathbf{x}) = \min_{\mathbf{u}\in\mathbf{U}}\left\{L(\mathbf{x}, \mathbf{u})\delta t + \alpha\left[\frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial t}\delta t + \left(\frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x}, \mathbf{u})\delta t\right]\right\}$$

$$\alpha \approx 1 - \beta\delta t$$

$$\beta\tilde{V}^*(t_n, \mathbf{x}) = \min_{\mathbf{u}\in\mathbf{U}}\left\{L(\mathbf{x}, \mathbf{u})\delta t + \alpha\left[\frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial t}\delta t + \left(\frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x}, \mathbf{u})\delta t\right]\right\}$$

not dependent of u, take out of min op

$$-\alpha\frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial t}\delta t + \beta\tilde{V}^*(t_n, \mathbf{x})\delta t = \min_{\mathbf{u}\in\mathbf{U}}\left\{L(\mathbf{x}, \mathbf{u})\delta t + \alpha\left(\frac{\partial \tilde{V}^*(t_n, \mathbf{x})}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x}, \mathbf{u})\delta t\right\}$$

$$\delta t \neq 0$$

$$\lim_{\delta t \to 0} \alpha = 1$$

$$\beta V^* - \frac{\partial V^*}{\partial t} = \min_{\mathbf{u}\in\mathbf{U}}\left\{L(\mathbf{x}, \mathbf{u}) + \left(\frac{\partial V^*}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x}, \mathbf{u})\right\}$$

Hamilton Jacobi Bellman Equation

Hamilton Jacobi Bellman Equation

Carl Gustav Jacob Jacobi (1804-1851)

William Rowan Hamilton (1805-1865)

Richard Bellman 1920-84

$$\frac{\partial V^*}{\partial t} = \beta V^* - \min_{\mathbf{u}\in\mathbf{U}} \left\{ L(\mathbf{x},\mathbf{u}) + \left(\frac{\partial V^*}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x},\mathbf{u}) \right\}$$

$$\beta V^* - \frac{\partial V^*}{\partial t} = \min_{\mathbf{u}\in\mathbf{U}} \left\{ L(\mathbf{x},\mathbf{u}) + \left(\frac{\partial V^*}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x},\mathbf{u}) \right\}$$

In general: Nonlinear, Partial Differential Equation

Has no analytical solution... : (

Backwards in time! $V^*(t_f, \mathbf{x}) = \Phi(\mathbf{x})$

$$\boxed{\frac{\partial V^{\mu}}{\partial t} + \left(\frac{\partial V^{\mu}}{\partial \mathbf{x}}\right)^{T} \mathbf{f}(\mathbf{x}, \mathbf{u})} = \boxed{\beta V^{\mu} - L(\mathbf{x}, \mathbf{u})}$$

total time derivative

immediate change of V

$$\frac{\partial V^{*}}{\partial t} = \beta V^{*} - \min_{\mathbf{u} \in \mathbf{U}} \left\{ L(\mathbf{x}, \mathbf{u}) + \left(\frac{\partial V^{*}}{\partial \mathbf{x}}\right)^{T} \mathbf{f}(\mathbf{x}, \mathbf{u}) \right\}$$

Conditions for optimal control

$$\beta = 0$$

no discount

HJB: $$\beta V^* - \frac{\partial V^*}{\partial t} = \min_{\mathbf{u} \in \mathbf{U}} \left\{ L(\mathbf{x}, \mathbf{u}) + \left(\frac{\partial V^*}{\partial \mathbf{x}} \right)^T \mathbf{f}(\mathbf{x}, \mathbf{u}) \right\}$$

min RHS: $$\frac{\delta L(\mathbf{x}, \mathbf{u})}{\delta \mathbf{u}} + \left(\frac{\partial V^*}{\partial \mathbf{x}} \right)^T \frac{\delta \mathbf{f}(\mathbf{x}, \mathbf{u})}{\delta \mathbf{u}} = \mathbf{0}$$

$$\frac{\delta}{\delta \mathbf{u}}$$

'small variation induced by small variation of u'

$$\frac{\delta L(\mathbf{x}, \mathbf{u})}{\delta \mathbf{u}} = - \left(\frac{\partial V^*}{\partial \mathbf{x}} \right)^T \frac{\delta \mathbf{f}(\mathbf{x}, \mathbf{u})}{\delta \mathbf{u}}$$

immediate cost 'paid' for small change in u

'gain' of value through change in state, induced by small change in u

'gain' of value induced by small change of u

Optimum: 'locally flat'!

$$\frac{dC}{dx} = 0$$

cost increases everywhere away from optimum

$$\frac{d^2C}{dx^2} < 0$$

A D R L

ETH Zürich

Infinite time

$$J = \int_{t_0}^{\infty} e^{-\beta(t-t_0)} L(\mathbf{x}(t), \mathbf{u}(t)) dt$$

Value function not function of time: $\frac{\partial V^*}{\partial t} = 0$

$$\beta V^* = \min_{\mathbf{u} \in \mathbf{U}} \{ L(\mathbf{x}, \mathbf{u}) + \left(\frac{\partial V^*}{\partial \mathbf{x}} \right)^T \mathbf{f}(\mathbf{x}, \mathbf{u}) \}$$

Stochastic system

$$\dot{\mathbf{x}}(t) = \mathbf{f}_t(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{B}(t)\mathbf{w}(t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

Mean: $E[\mathbf{w}(t)] = \bar{\mathbf{w}} = 0$ mean-free

Co-variance: $E[\mathbf{w}(t)\mathbf{w}(\tau)^T] = \mathbf{W}(t)\delta(t-\tau)$ uncorrelated over time

$$E[\mathbf{w}(t)\mathbf{w}(\tau)^T] = 0$$
$$t \neq \tau$$

Expected cost:

$$J = E\left\{e^{-\beta(t_f - t_0)}\Phi(\mathbf{x}(t_f)) + \int_{t_0}^{t_f} e^{-\beta(t'-t_0)} L(\mathbf{x}(t'), \mathbf{u}(t'))dt'\right\}$$

Derivation of stoch. HJB

Optimal value function (cost-to-go)

cost over optimal trajectory, using optimal control

$$V^*(t,\mathbf{x}) = E\left\{e^{-\beta(t_f-t)}\Phi(\mathbf{x}^*(t_f)) + \int_t^{t_f} e^{-\beta(t'-t)} L(\mathbf{x}^*(t'),\mathbf{u}^*(t'))dt'\right\}$$

Leibniz' rule (differentiation under integrals):

$$\frac{dV^*(t,x)}{dt} = E\left\{\beta e^{-\beta(t_f-t)}\Phi(\mathbf{x}^*(t_f)) + \beta\int_t^{t_f} e^{-\beta(t'-t)} L(\mathbf{x}^*(t'),\mathbf{u}^*(t'))dt' - L(\mathbf{x},\mathbf{u}^*(t))\right\}$$

$$\beta V^*(t,\mathbf{x}) - E\left\{L(\mathbf{x},\mathbf{u}^*(t))\right\}$$

known with certainty

$$\frac{dV^*(t,\mathbf{x})}{dt} = \beta V^*(t,\mathbf{x}) - L(\mathbf{x},\mathbf{u}^*(t))$$

intermediate result, to be used later

Intuition: immediate cost decreases cost-value, decay factor increases

A D R L

ETH Zürich

Derivation of stoch. HJB

cont'd

Taylor series expansion of Value function

$$\Delta V^*(t,\mathbf{x}) \approx \frac{dV^*(t,\mathbf{x})}{dt}\Delta t$$

$$= E\left\{\frac{\partial V^*(t,\mathbf{x})}{\partial t}\Delta t + \left(\frac{\partial V^*(t,\mathbf{x})}{\partial \mathbf{x}}\right)^T \dot{\mathbf{x}}\Delta t + \frac{1}{2}\dot{\mathbf{x}}^T\frac{\partial^2 V^*(t,\mathbf{x})}{\partial \mathbf{x}^2}\dot{\mathbf{x}}\Delta t^2\right\}$$

Plug in

$$\dot{\mathbf{x}}(t) = \mathbf{f}_t(\mathbf{x}(t),\mathbf{u}(t)) + \mathbf{B}(t)\mathbf{w}(t)$$

using shorthand notation $V^*_{\mathbf{z}} := \frac{\partial V^*}{\partial \mathbf{z}}$ $\mathbf{f} := \mathbf{f}_t(\mathbf{x}(t),\mathbf{u}(t))$

$$\frac{dV^*}{dt}\Delta t = E[V^*_t\Delta t + V^{*T}_{\mathbf{x}}(\mathbf{f}+\mathbf{B}\mathbf{w})\Delta t + \frac{1}{2}(\mathbf{f}+\mathbf{B}\mathbf{w})^T V^*_{\mathbf{xx}}(\mathbf{f}+\mathbf{B}\mathbf{w})\Delta t^2]$$

A D R L

ETH Zürich

$$\frac{dV^*}{dt}\Delta t = E[V_t^*\Delta t + V_{\mathbf{x}}^{*T}(\mathbf{f}+\mathbf{Bw})\Delta t + \frac{1}{2}(\mathbf{f}+\mathbf{Bw})^T V_{\mathbf{xx}}^*(\mathbf{f}+\mathbf{Bw})\Delta t^2]$$

$E[\mathbf{w}] = 0$

known with certainty

$\Delta t \neq 0$ divide

$$\frac{dV^*}{dt} = V_t^* + V_{\mathbf{x}}^{*T}\mathbf{f} + \frac{1}{2}\mathrm{Tr}\{E[(\mathbf{f}+\mathbf{Bw})^T V_{\mathbf{xx}}^*(\mathbf{f}+\mathbf{Bw})]\Delta t\}$$

known with certainty

$\mathrm{tr}[\mathbf{AB}] = \mathrm{tr}[\mathbf{BA}]$

$$\frac{dV^*}{dt} = V_t^* + V_{\mathbf{x}}^{*T}\mathbf{f} + \frac{1}{2}\mathrm{Tr}\left[V_{\mathbf{xx}}^* E[(\mathbf{f}+\mathbf{Bw})^T(\mathbf{f}+\mathbf{Bw})]\Delta t\right]$$

expand

$$\frac{dV^*}{dt} = V_t^* + V_{\mathbf{x}}^{*T}\mathbf{f} + \frac{1}{2}\mathrm{Tr}\left[V_{\mathbf{xx}}^*\left(\mathbf{ff}^T\Delta t + 2\mathbf{f}E(\mathbf{w}^T)\mathbf{B}^T\Delta t + \mathbf{B}E(\mathbf{ww}^T)\mathbf{B}^T\Delta t\right)\right]$$

= 0 (mean free noise)

rearrange

$$\frac{dV^*}{dt} = V_t^* + V_{\mathbf{x}}^{*T}\mathbf{f} + \frac{1}{2}\mathrm{Tr}\left[V_{\mathbf{xx}}^*\left(\mathbf{ff}^T\Delta t + \mathbf{BWB}^T\delta(t)\Delta t\right)\right]$$

Stoch. HJB

$$\frac{dV^*}{dt} = V_t^* + V_{\mathbf{x}}^{*T}\mathbf{f} + \frac{1}{2}\mathrm{Tr}\left[V_{\mathbf{xx}}^*\left(\mathbf{f}\mathbf{f}^T\Delta t + \mathbf{BWB}^T\delta(t)\Delta t\right)\right]$$

Assuming that $\lim_{\Delta t \to 0} \delta(t)\Delta t = 1$, and taking the limit as $\Delta t \to 0$,

$$\frac{dV^*}{dt} = V_t^* + V_{\mathbf{x}}^{*T}\mathbf{f} + \frac{1}{2}\mathrm{Tr}\left[V_{\mathbf{xx}}^*\mathbf{BWB}^T\right]$$

Replace LHS with
(using intermediate result S32)

$$\frac{dV^*(t,\mathbf{x})}{dt} = \beta V^*(t,\mathbf{x}) - L(\mathbf{x},\mathbf{u}^*(t))$$

$$\beta V^*(t,\mathbf{x}) - V_t^*(t,\mathbf{x}) = \min_{\mathbf{u}(t)}\left\{L(\mathbf{x},\mathbf{u}(t)) + V_{\mathbf{x}}^{*T}\mathbf{f}_t(\mathbf{x},\mathbf{u}(t)) + \frac{1}{2}\mathrm{Tr}[V_{\mathbf{xx}}^*\mathbf{B}(t)\mathbf{W}(t)\mathbf{B}^T(t)]\right\}$$

Hamilton Jacobi Bellman Equation

$$V^*(t_f,\mathbf{x}) = \Phi(\mathbf{x})$$

A D R L

ETH Zürich

'Stochastic' Hamilton Jacobi Bellman Equation

$$\beta V^*(t,\mathbf{x}) - V_t^*(t,\mathbf{x}) = \min_{\mathbf{u}(t)} \left\{ L(\mathbf{x},\mathbf{u}(t)) + V_{\mathbf{x}}^{*T} \mathbf{f}_t(\mathbf{x},\mathbf{u}(t)) + \frac{1}{2}\mathbf{Tr}[V_{\mathbf{xx}}^* \mathbf{B}(t)\mathbf{W}(t)\mathbf{B}^T(t)] \right\}$$

add'l cost

compare to deterministic HJB: $\beta V^* - \frac{\partial V^*}{\partial t} = \min_{\mathbf{u}\in\mathbf{U}} \left\{ L(\mathbf{x},\mathbf{u}) + \left(\frac{\partial V^*}{\partial \mathbf{x}}\right)^T \mathbf{f}(\mathbf{x},\mathbf{u}) \right\}$

	Discrete Time	Continuous Time
Stochastic System	Optimization Problem: $\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n, \mathbf{u}_n) + \mathbf{w}_n$ $\mathbf{w}_n \sim P_{\mathbf{w}}(\cdot \mid \mathbf{x}_n, \mathbf{u}_n)$ $\min_{\mathbf{u}_{0 \to N-1}} E\{\alpha^N \Phi(N) + \sum_{k=0}^{N-1} \alpha^k L(\mathbf{x}_k, \mathbf{u}_k)\} \quad \alpha \in [0,1]$ Stochastic Bellman equation: $V^*(n, \mathbf{x}) = \min_{\mathbf{u}_n}\{L(x, u_n) + \alpha E[V^*(n+1, \mathbf{x}_{n+1})]\}$ Infinite horizon: $\alpha \in [0,1)$ $\Phi(N) = 0$ V^* is not function of time.	Optimization Problem: $d\mathbf{x} = \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t)dt + \mathbf{B}(\mathbf{x}_t, \mathbf{u}_t)d\mathbf{w}_t$ $\mathbf{w}_t \sim \mathcal{N}(0, \Sigma)$ $\min_{\mathbf{u}_{0 \to t_f}} E\{e^{-\beta t_f}\Phi(t_f) + \int_0^{t_f} e^{-\beta t} L(\mathbf{x}_t, \mathbf{u}_t)dt\}$ Stochastic HJB equation: $\beta V^*(t, \mathbf{x}) - V_t^*(t, \mathbf{x}) = \min_{\mathbf{u}_t}\{L(\mathbf{x}, \mathbf{u}_t) + V_x^{*T}(t, \mathbf{x})\mathbf{f}(\mathbf{x}, \mathbf{u}_t) + \frac{1}{2} Tr[V_{\mathbf{xx}}^*(t, \mathbf{x})\mathbf{B}\Sigma\mathbf{B}^T]\}$ Infinite horizon: $\Phi(t_f) = 0$ V^* is not function of time.
Deterministic System	Optimization Problem: $\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n, \mathbf{u}_n)$ $\min_{\mathbf{u}_{0 \to N-1}} \{\alpha^N \Phi(N) + \sum_{k=0}^{N-1} \alpha^k L(\mathbf{x}_k, \mathbf{u}_k)\} \quad \alpha \in [0,1]$ Bellman equation: $V^*(n, \mathbf{x}) = \min_{u_n}\{L(\mathbf{x}, \mathbf{u}_n) + \alpha V^*(n+1, \mathbf{x}_{n+1})\}$ Infinite time horizon: $\alpha \in [0,1)$ $\Phi(N) = 0$ V^* is not function of time.	Optimization Problem: $d\mathbf{x} = \mathbf{f}(\mathbf{x}_{(t)}, \mathbf{u}_{(t)})dt$ $\min_{\mathbf{u}_{0 \to t_f}} \{e^{-\beta t_f}\Phi(t_f) + \int_0^{t_f} e^{-\beta t} L(\mathbf{x}_t, \mathbf{u}_t)dt\}$ HJB equation: $\beta V^*(t, \mathbf{x}) - V_t^*(t, \mathbf{x}) = \min_{\mathbf{u}_{(t)}}\{L(\mathbf{x}_t, \mathbf{u}_t) + V_{\mathbf{x}}^{*T}(t, \mathbf{x})\mathbf{f}(\mathbf{x}, \mathbf{u})\}$ Infinite time horizon: $\Phi(t_f) = 0$ V^* is not function of time.

Stochastic, Discrete → Continuous: Section 1.3.3

Deterministic, Discrete → Continuous: Section 1.3.1

Discrete, Stochastic → Deterministic: $\mathbf{w}_n \sim P_{\mathbf{w}}(\cdot \mid \mathbf{x}_n, \mathbf{u}_n) = \delta(\mathbf{w})$

Continuous, Stochastic → Deterministic: $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{0})$ i.e. $\Sigma = \mathbf{0}$

L2 L3