Control of Stochastic Systems
Lecture 9
Optimal Stochastic Control with Complete Observations on an Infinite Horizon

Jan H. van Schuppen

1st May 2025

Delft University of Technology
Delft, The Netherlands

Outline

Outline

Control on an infinite-horizon

- Infinite-horizon case if the time index set is $T = \mathbb{N} = \{0, 1, \ldots\}$.
  May be called a half-infinite horizon.
- For many engineering control problems
  the control law is expected to be operating for relatively long periods.
  Moreover, the effects near the end point of the time horizon
  are not of interest.
- An infinite-horizon control problem is then appropriate.
- The optimal control law is simple to implement
  if it does not explicitly depend on time.

Def. Cost functions

Consider the time-invariant recursive stochastic control system,

$$x(t+1) = f(x(t), u(t), v(t)),\ x(0) = x_0,\ \ b : X \times U \to \mathbb{R}_+,$$
$$G_{ti,m} = \{g : X \to U |\ g \text{ measurable}\};\ \ u^g(s) = g(x^g(s)).$$

Define respectively:
– the average cost function,
– the discounted cost function with discount rate $r \in (0,1)$, and
– the total cost function according to the formulas,

$$J_{ac}(g) = \limsup_{t\to\infty} \frac{1}{t} E[\sum_{s=0}^{t-1} b(x^g(s), u^g(s))],\ J_{ac} : G_{ti,m} \to \mathbb{R}_+ \cup \{+\infty\},$$

$$J_{dc}(g) = E[\sum_{s=0}^{\infty}\ r^s\ b(x^g(s), u^g(s))],\ J_{dc} : G_{ti,m} \to \mathbb{R}_+ \cup \{+\infty\},$$

$$J_{tc}(g) = E[\sum_{s=0}^{\infty} b(x^g(s), u^g(s))],\ J_{tc} : G_{ti,m} \to \mathbb{R}_+ \cup \{+\infty\}.$$

Which cost criterion for a problem?

- Which cost function is most suitable
  for a considered stochastic control problem?
- Average cost function.
  Near and distant future costs have equal weight.
  Much used in engineering.
- Discounted cost function.
  Near future costs are relatively more important than
  distant future costs.
  Much used in economics.
  Not treated in this lecture, see the book.
- Total cost function.
  Not much used. Useful?

Control synthesis - Distinguish the cases

Either the state set, or the input set, or the cost function are:

- Finite.
  A finite state set and a finite input set.
- Countable.
  A countable infinite state set and a countable or finite input set.
- Positive.
  Uncountable state space and an uncountable input space
  with a positive cost function.
- Negative.
  Uncountable state space and an uncountable input space
  with a negative cost function.

The cases require different analytical methods.
Below use is made of the positive case,
which includes the finite case and the countable case,
both with positive cost.

Outline

Problem. Optimal stochastic control with complete observations and with average cost (1)

$$
\begin{aligned}
x(t+1) &= f(x(t), u(t), v(t)),\ x(0) = x_0, \\
& n_x,\ n_u,\ n_v \in \mathbb{Z}_+,\ T = \mathbb{N},\ X = \mathbb{R}^{n_x},\ U = \mathbb{R}^{n_u}, \\
& v : \Omega \times T \to \mathbb{R}^{n_v}, \\
& \text{independent and stationary sequence,} \\
G_{tv} &= \left\{ \begin{array}{l} g = (g_0, g_1(.), \ldots, g_t(.), \ldots) |\ \forall\ t \in T, \\ g_t : X^{t+1} \to U,\ \text{Borel measurable} \end{array} \right\}; \\
& \text{closed-loop control system,} \\
x^g(t+1) &= f(x^g(t), g_t(x^g(0:t)), v(t)),\ x^g(0) = x_0, \\
u^g(t) &= g_t(x^g(0:t)) = g_t(x^g(0), x^g(1), \ldots, x^g(t)), \\
x^g(0:t) &= (x^g(0), x^g(1), \ldots, x^g(t)).
\end{aligned}
$$

Set G_{tv} considered is of time-varying control laws.
The engineering interest is in time-invariant control laws.

Problem. Optimal stochastic control with average cost (2)

$$
\begin{aligned}
& J_{ac} : G_{tv} \rightarrow \mathbb{R}_+ \cup \{\infty\},\ b : X \times U \rightarrow \mathbb{R}_+, \\
J_{ac}(g) &= \limsup_{t_1 \rightarrow \infty} \frac{1}{t_1} E\left[\sum_{s=0}^{t_1-1} b(x^g(s), g_s(x^g(0:s)))\right] \in \mathbb{R}_+ \cup \{+\infty\}; \\
G_{ac,f} &= \{g \in G_{tv} |\ J_{ac}(g) < \infty\}, \\
& \inf_{g \in G_{ac,f}} J_{ac}(g).
\end{aligned}
$$

Comments

- If $G_{ac,f} = \emptyset$ then no comparison of control laws is possible!
- Assumption of stochastic controllability may be needed for the condition $G_{ac,f} \neq \emptyset$.
- The condition $G_{ac,f} \neq \emptyset$ is discussed for each example separately.

Comments. Control theory issues (discussed below)

- How to derive
  the dynamic programming equation of average cost?
- Does there exist a pair (J_{ac}^*, V)
  which is a solution of the dynamic programming equation?
  Existence requires stochastic controllability.
- Does the dynamic programming equation
  have a unique solution?
- How to compute
  a solution of the dynamic programming equation?
- How to derive an optimal control law?

Def. Average-cost dynamic programming equation (DPE-AC)

$(J^*_{ac}, V), \quad J^*_{ac} \in \mathbb{R}_+, \; V : X \to \mathbb{R}_+,$

assume that $\forall \; x_v \in X, \;\; \exists \; U(x_V) \subseteq U$;

define the DPE-AC for $(J^*_{ac}, \; V)$ by the formula

$$\begin{aligned} & J^*_{ac} + V(x_V) \\ & = \inf_{u_V \in U(x_V)} \{b(x_V, u_V) + E[V(f(x_V, u_V, v(t)))|F^{x_V,u_V}]\}, \; \forall \; x_V \in X; \end{aligned}$$

recall the notation,

$$E[V(f(x_V, u_V, v(t))|F^{x_V,u_V}] = \int V(f(x_V, u_V, w)) \; f_{v(t)}(dw).$$

Also called the Bellman-Hamilton-Jacobi equation (BHJ equation) of optimal stochastic control.
R. Bellman (USA, active till about 1975),
R.W. Hamilton (Ireland, 1805-1865),
C.G.J. Jacobi (Switzerland, 1804-1851), all mathematicians.

Comments

- Call J^*_{ac} the value and call V the value function of the average-cost dynamic programming equation (DPE-AC).
- The equation DPE-AC is for $(J^*_{ac},\ V)$.
- Needed is a procedure to compute or to calculate $(J^*_{ac},\ V)$.
- Note in the DPE-AC, the infimization of the sum of:
  – the cost rate $b(x_V, u_V)$ of one time step and
  – the current estimate of the optimal cost-to-go at the next state.

Proposition. Uniqueness of the solution of the DPE-AC

(a) If $(J^*_{ac,1}, V)$ and $(J^*_{ac,2}, V)$ are two solutions of DPE-AC then $J^*_{ac,1} = J^*_{ac,2}$.

(b) If (J^*_{ac}, V) is a solution of DPE-AC and $c \in \mathbb{R}_+$ then $(J^*_{ac}, V + c)$ is another solution.

Proof is simple, see lecture notes.

Procedure. Optimal control law from value function

Assume $G_{ac,f} \neq \emptyset$.

1. Determine a solution (J^*_{ac}, V) of the DPE-AC, existence assumed.

$$J^*_{ac} \in \mathbb{R}_+, \ V : X \to \mathbb{R}_+,$$
$$J^*_{ac} + V(x_V) = \inf_{u_V \in U(x_V)} \{b(x_V, u_V) + E[V(f(x_V, u_V, v(t)))|F^{x_V,u_V}]\};$$

2.

$$\begin{aligned} & \text{if } \forall\, x_V \in X, \ \exists\, u^* \in U(x_V) \text{ such that infimum is attained,} \\ & b(x_V, u^*) + E[V(f(x_V, u^*, v(t))|F^{x_V,u^*}] \\ = & \inf_{u_V \in U(x_V)} \{b(x_V, u_V) + E[V(f(x_V, u_V, v(t))|F^{x_V,u_V}]\}, \\ & \text{then define} \end{aligned}$$
$$g^*(x_V) = u^*, \ \ g^* : X \to U.$$

3. Is g^* a measurable function? If not then stop.
4. Output $(J^*_{ac}, \ V, \ g^*)$.

Theorem. Sufficient condition for optimal control law.

Assume $G_{ac,f} \neq \emptyset$.

(a) Assume (a.1) $\exists\ (J^*_{ac},\ V)$ a solution of the DPE-AC;

$$\text{(a.2) } E[V(x(t))] < \infty,\ \forall\ t \in T,\ \text{ and}$$

$$\text{(a.3) } \forall\ g \in G_{ac,f},\ \limsup_{t_1 \to \infty} \frac{1}{t_1} E[V(x^g(t_1))] = 0.\ \text{ Then}$$

$$J^*_{ac} \leq J_{ac}(g),\ \forall\ g \in G_{ac,f};\ \text{ hence,}$$

$$J^*_{ac} \leq \inf_{g \in G_{ac,f}} J_{ac}(g).$$

(b) If, in addition to (a),
(b.1) the infima in the DPE-AC are all attained,
(b.2) if g^* is a measurable function, and
(b.3) if $g^* \in G_{ac,f}$,
then the function $g^* \in G_{ac,f}$ is an optimal control law and

$$J^*_{ac} = J_{ac}(g^*) = \inf_{g \in G_{ac,f}} J_{ac}(g).$$

Proof of Theorem (1)

Let $g \in G_{ac,f}$, and $t_1 \in T$, and $s \in \{0, 1, \ldots, t_1 - 1\}$.

$$
\begin{aligned}
& b(x^g(s), u^g(s)) + E[V(f(x^g(s), u^g(s), v(s)))|F_s^{x^g}] \\
\geq & \inf_{u_V \in U(x^g(s))} \left\{ b(x^g(s), u_V) + E[V(f(x^g(s), u_V, v(s)))|F_s^{x^g}] \right\} \\
= & J_{ac}^* + V(x^g(s)), \text{ by DPE-AC,} \\
\Rightarrow & \sum_{s=0}^{t_1-1} \left(b(x^g(s), u^g(s)) + E[V(x^g(s+1))|F_s^{x^g}] \right) \\
\geq & \sum_{s=0}^{t_1-1} [J_{ac}^* + V(x^g(s))] = t_1 J_{ac}^* + \sum_{s=0}^{t_1-1} V(x^g(s)); \\
\Rightarrow &
\end{aligned}
$$

Proof of Theorem (2)

$$
\begin{aligned}
J_{ac}^* \leq\ & \frac{1}{t_1} \sum_{s=0}^{t_1-1} E[b(x^g(s), u^g(s))] + \\
& + \frac{1}{t_1} \sum_{s=0}^{t_1-1} E[E[V(x^g(s+1)) | F^{x^g(s), u^g(s)}]] - \frac{1}{t_1} \sum_{s=0}^{t_1-1} E[V(x^g(s))] \\
J_{ac}^* \leq\ & \limsup_{t_1 \to \infty} \frac{1}{t_1} \sum_{s=0}^{t_1-1} E[b(x^g(s), u^g(s))] + \\
& + \limsup_{t_1 \to \infty} \frac{1}{t_1} E[V(x^g(t_1))] - \liminf_{t_1 \to \infty} \frac{1}{t_1} E[V(x(0))] \\
=\ & \limsup_{t_1 \to \infty} \frac{1}{t_1} \sum_{s=0}^{t_1-1} E[b(x^g(s), u^g(s))] = J_{ac}(g), \text{ by assumptions,} \\
J_{ac}^* \leq\ & \inf_{g \in G_{ac,f}} J_{ac}(g).
\end{aligned}
$$

Proof of Theorem (3)

(b) If infima of DPE-AC are all attained
then equality holds in the first inequality
hence equality holds also in the other inequalities.

$$J_{ac}(g^*) = J^*_{ac} \leq \inf_{g \in G_{ac,f}} J_{ac}(g) \leq J_{ac}(g^*),$$
$$J_{ac}(g^*) = J^*_{ac} = \inf_{g \in G_{ac,f}} J_{ac}(g).$$

Thus J^*_{ac} is the value and $g^* \in G_{ac,f}$ is an optimal control law.

How to check whether a candidate solution is actually a solution of the DPE-AC?

1. Formulate a candidate solution $(J_{ac}^*,\ V)$.
2. Calculate the right-hand side of the DPE-AC using the candidate solution.
3. Check whether the right-hand side of DPE-AC equals the left-hand side of the DPE-AC, which is $J_{ac}^* + V(x_V)$. If so then the candidate solution is a solution of the DPE-AC.

How to determine a candidate solution (J_{ac}^*, V) of the DPE-AC?

- Run a few steps of the finite-horizon case DP procedure. Formulate a conjecture about the analytic form of the value function $V(x_V)$.
- Check whether the candidate solution is a solution.

Control problems with optimal control laws in explicit form?

- Gaussian stochastic control system with a quadratic cost rate.
  See below.
- Finite stochastic control system.
  See below.
- LEQG problem.
- Examples with a countable state set.

Control theory for discounted cost

Different dynamic programming equation.
Otherwise analogous to average cost.
Proofs are simpler, using the theory of a fix point equation.

Outline

Problem. Linear-Quadratic-Gaussian (LQG) optimal stoc. control with complete observations on an infinite-horizon with average cost

$$
\begin{aligned}
x(t+1) &= Ax(t) + Bu(t) + Mv(t),\ x(0) = x_0, \\
z(t) &= C_z x(t) + D_z u(t), \quad n_u \le n_z,\ D_z \in \mathbb{R}^{n_z \times n_u}, \\
&\quad \mathrm{rank}(D_z) = n_u \ \Rightarrow\ \mathrm{rank}(D_z^T D_z) = n_u \ \Rightarrow\ D_z^T D_z \succ 0, \\
&\quad \text{past-state information structure, } G_{tv} \text{ set of control laws;} \\
x^g(t+1) &= Ax^g(t) + Bg_t(x(0:t)) + Mv(t), x^g(0) = x_0, \\
z^g(t) &= C_z x^g(t) + D_z g_t(x(0:t)), \\
u^g(t) &= g_t(x^g(0:t));\ \ J_{ac} : G_{tv} \to \mathbb{R}_+ \cup \{+\infty\}, \\
J_{ac}(g) &= \limsup_{t_1 \to \infty} \frac{1}{t_1} E\left[\sum_{s=0}^{t_1-1} z^g(s)^T z^g(s)\right], \\
Q_{cr} &= \begin{pmatrix} C_z^T C_z & C_z^T D_z \\ (C_z^T D_z)^T & D_z^T D_z \end{pmatrix}.
\end{aligned}
$$

Assumption. LQG average cost

$$A_c = A - B(D_z^T D_z)^{-1} D_z^T C_z \in \mathbb{R}^{n_x \times n_x},$$

$$C_c^T C_c = C_z^T C_z - C_z^T D_z (D_z^T D_z)^{-1} D_z^T C_z \in \mathbb{R}_{pds}^{n_x \times n_x},$$

$C_c \in \mathbb{R}^{n_x \times n_x}$ is determined by a factorization.

Assumption (a) holds if:
(1) $(A,\ B)$ is a controllable pair; and
(2) $(A_c,\ C_c)$ is an observable pair.
Assumption (b) holds if:
(1) $(A,\ B)$ is a stabilizable pair; and
(2) $(A_c,\ C_c)$ is a detectable pair.

Comments

- Assumption (a) implies Assumption (b). Converse not true.
- $D_z^T C_z = 0$ implies that $(A_c,\ C_c) = (A, C_z)$
  hence (a.2) holds if (A, C_z) is an observable pair.
  (b.2) similarly.

Proposition. LQG finite average cost

Define the set of linear Markov control laws,

$$G_{LM} = \left\{ \begin{array}{l} g : X \rightarrow U | \ \exists \ F \in \mathbb{R}^{n_u \times n_x} \text{ such that} \\ g(x) = Fx \text{ and } \mathrm{spec}(A + BF) \subset \mathbb{D}_o \end{array} \right\}.$$

(a) Assumption (b) holds. Then,

$$\begin{aligned} &\exists \ F \in \mathbb{R}^{n_u \times n_x}, \ \mathrm{spec}(A+BF) \subset \mathbb{D}_o, \quad \text{see Prop. 21.2.11 of LN;} \\ \Rightarrow \ & g(x) = Fx, \ g \in G_{LM} \neq \emptyset. \end{aligned}$$

(b) Assumption (b) holds. Then,

$$\begin{aligned} \forall \ g \in G_{LM}, \ g(x) = Fx, \ \ \exists \ Q_x^g \in \mathbb{R}^{n_x \times n_x}_{pds}, \\ Q_x^g = (A+BF) Q_x^g (A+BF)^T + MM^T, \\ J_{ac}(g) = \mathrm{tr}\left((C_z + D_z F) Q_x^g (C_z + D_z F)^T \right) < \infty, \\ \Rightarrow g \in G_{ac,f}. \end{aligned}$$

Proof of Proposition (1)

(a) This follows from
the assumption that $(A,\ B)$ is a stabilizable pair and
from control theory of linear systems (see Chapter 21).
(b) The closed-loop system satisfies $\text{spec}(A + BF) \subset \mathbb{D}_o$.
Then there exists an invariant measure,

$$\begin{aligned}
x(t) &\in G(0, Q_x^g); \\
Q_x^g &= (A + BF) Q_x^g (A + BF)^T + MM^T; \\
Q_x^g(t+1) &= (A + BF) Q_x^g(t)(A + BF) + MM^T,\ Q_x^g(0) = Q_{x_0}, \\
Q_x^g &= \lim_{t \to \infty} Q_x^g(t), \\
Q_x^g &= \lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} Q_x^g(s),\ \text{see Section 22.1 of lecture notes;}
\end{aligned}$$

Proof of Proposition (2)

$$
\begin{aligned}
z^g(t) &= C_z x^g(t) + D_z u^g(t) = (C_z + D_z F) x^g(t), \\
E[z^g(t)^T z^g(t)] &= \mathrm{tr}((C_z + D_z F) Q_x^g(t) (C_z + D_Z F)^T), \\
J_{ac}(g) &= \lim_{t_1 \to \infty} \frac{1}{t_1} \sum_{s=0}^{t_1 - 1} E[z^g(s)^T z^g(s)] \\
&= \lim_{t_1 \to \infty} \frac{1}{t_1} \sum_{s=0}^{t_1 - 1} \mathrm{tr}\left((C_z + D_z F) Q_x^g(s) (C_z + D_z F)^T\right) \\
&= \mathrm{tr}((C_z + D_z F) Q_x^g (C_z + D_z F)^T) < \infty, \\
&\quad \text{by Proposition 17.5.9 of the book.}
\end{aligned}
$$

Def. Control algebraic Riccati equation

Define the control algebraic Riccati equation (CARE) with side conditions,

$$\begin{aligned}
Q_c^* &= f_{CARE}(Q_c^*) \\
&= A^T Q_c^* A + C_z^T C_z + \\
&\quad - [A^T Q_c^* B + C_z^T D_z][B^T Q_c^* B + D_z^T D_z]^{-1}[A^T Q_c^* B + C_z^T D_z]^T, \\
(1) &\quad \text{spec}(A + BF(Q_c^*)) \subset \mathbb{D}_o, \\
(2) &\quad Q_c^* \in \mathbb{R}_{pds}^{n_x \times n_x}; \\
&\quad \text{where} \\
F(Q_c^*) &= -[B^T Q_c^* B + D_z^T D_z]^{-1}[A^T Q_c^* B + C_z^T D_z]^T \in \mathbb{R}^{n_u \times n_x}, \\
&\quad f_{CARE} : \mathbb{R}^{n_x \times n_x} \to \mathbb{R}^{n_x \times n_x}.
\end{aligned}$$

Comments

- Assumption (b) implies that there exists a solution of CARE. Hence also Assumption (a) implies the existence.
- Equation $Q_c = f_{CARE}(Q_c)$ does not have a unique solution.
- The equation has a unique solution Q_c^* with the side conditions: $\text{spec}(A + BF(Q_c^*)) \subset \mathbb{D}_o$ and $Q_c^* \in \mathbb{R}_{pds}^{n_x \times n_x}$.
  In fact, either of these side conditions implies the other.
- See Section 22.2 of the book for algebraic Riccati equations of filtering and of control.
- Lecture 11 provides more details on the Riccati equation.
- The Matlab command idare
  can compute a numerical approximation of Q_c^*.
  Check Matlab whether $C_z^T D_z = 0$ is required.
  If required then transform according to formulas.
  Check whether you use the discrete-time or
  the continous-time CARE called icare!

Def. Optimal LQG average cost control law

Define the optimal LQG average cost control law
for a solution Q_c^* of CARE with side conditions,

$$g^*(x) = g^*_{LQG,co,ac}(x) = F(Q_c^*)\, x,$$
$$F(Q_c^*) = -[B^T Q_c^* B + D_z^T D_z]^{-1}[A^T Q_c^* B + C_z^T D_z]^T \in \mathbb{R}^{n_u \times n_x};$$
$$\text{then } \mathrm{spec}(A + BF(Q_c^*)) \subset \mathbb{D}_o,$$
$$\text{because of definition of } Q_c^*.$$

Theorem. LQG average cost (1)

Assumption (b) holds.

(a) There exists a solution Q_c^* of the control algebraic Riccati equation (CARE) with side conditions.

(b) The control law $g^*_{LQG,co,ac} = g^*$ is well defined and satisfies

$$\begin{aligned} &\mathrm{spec}(A + BF(Q_c^*)) \subset \mathbb{D}_o, \\ &J_{ac}(g^*) < \infty \;\Rightarrow\; g^* \in G_{ac,f} \neq \emptyset. \end{aligned}$$

(c) The control law g^* depends on the system matrices $(A,\ B,\ C_z,\ D_z)$ but not on M; the value J_{ac}^* depends on $(A,\ B,\ C_z,\ D_z,\ M)$.

Theorem. LQG average cost (2)

(d) The solution of the average-cost dynamic programming equation is

$$
\begin{aligned}
V(x_V) &= x_V^T Q_c^* x_V, \\
J_{ac}^* &= \mathrm{tr}(M^T Q_c^* \, M), \\
J_{ac}^* &= J_{ac}(g^*), \\
g^*(x^g) &= g^*_{LQG,co,ac}(x^g) = F(Q_c^*) x^g, \\
F(Q_c^*) &= -[B^T Q_c^* B + D_z^T D_z]^{-1} [A^T Q_c^* B + C_z^T D_z]^T.
\end{aligned}
$$

The optimal control law g^* is a linear function of the state.
It is optimal over the set of nonlinear Borel-measurable control laws.

Theorem. LQG average cost (3)

(e) Optimality over the subset of linear control laws.
For any linear control law g_s, not necessarily optimal, satisfying a stability condition, it is true that

$$
\begin{aligned}
g_s(x^g) = F_s x^g, &\ \ \text{spec}(A + BF_s) \subset \mathbb{D}_o \text{ assumed,} \\
& J_{ac}(g^*) - J_{ac}(g_s) \\
& = \mathrm{tr}(M^T Q_c^* M) - \mathrm{tr}(M^T Q_c^{g_s} M) \\
& = \mathrm{tr}(M^T (Q_c^* - Q_c^{g_s}) M) \leq 0, \\
Q_c^{g_s} &= (A + BF_s)^T Q_c^{g_s} (A + BF_s) + (C_z + D_z F_s)^T (C_z + D_z F_s).
\end{aligned}
$$

Note that the latter equation is a control Lyapunov equation, note the transposes.
Note the relation with matrix inequality and stochastic realization!
Call

$$J_{ac}(g^*) - J_{ac}(g_s) \leq 0,$$

the regret for using g_s rather than g^*.

Proof of LQG average-cost (1)

(a) This follows from the assumptions and from Th. 22.2.2.
(b) This follows from the definition of Q_c^* and from a proposition.
(c) This follows directly from the formulas.
(d) The closed-loop system is a Gaussian system,

$$\begin{aligned}
x^g(t+1) &= (A + BF(Q^*))x^g(t) + Mv(t),\ x^g(0) = x_0, \\
&\quad \text{spec}(A + BF(Q^*)) \subset \mathbb{D}_o; \\
&\quad x^g(t) \in G(m^g(t),\ Q_{x^g}(t)),\ \forall\ t \in T, \\
&\quad \lim_{t \rightarrow +\infty} Q_{x^g}(t) = Q_{x^g}(+\infty) \in \mathbb{R}_{pds}^{n_x \times n_x}; \\
E[V(x^g(t))] &= E[x^g(t)^T Q_c^* x^g(t)] = \mathrm{tr}(Q_c^* Q_{x^g}(t))) < +\infty,\ \forall\ t \in T, \\
&\quad \text{assumption (a.2) holds;} \\
&\quad \limsup_{t \rightarrow +\infty} \frac{E[V(x^g(t))]}{t} = \lim \frac{\mathrm{tr}(Q_c^* Q_{x^g}(t))}{t} = 0.
\end{aligned}$$

Hence assumption (a.3) holds.

Proof of LQG average-cost (2)

(d) It will be shown below that the pair (J_{ac}^*, V) specified in the theorem is a solution of the dynamic programming equation,

$$J_{ac}^* + V(x_V) = \inf_{u_V \in U} \left\{ \begin{bmatrix} x_V \\ u_V \end{bmatrix}^T Q_{cr} \begin{bmatrix} x_V \\ u_V \end{bmatrix} + E_g[V(x(t+1))|F^{x_V,\, u_V}] \right\}.$$

A calculation shows that

$$\begin{aligned}
& E_g[V(x(t+1))|\, F^{x_V, u_V}] \\
& = E[x(t+1)^T Q_c^* x(t+1)|\, F^{x_V,\, u_V}] \\
& = E_g[(Ax_V + Bu_V + Mv(t))^T Q_c^* (Ax_V + Bu_V + Mv(t))|F^{x_V,\, u_V}] \\
& = \begin{bmatrix} x_V \\ u_V \end{bmatrix}^T \begin{bmatrix} A^T Q_c^* A & A^T Q_c^* B \\ (A^T Q_c^* B)^T & B^T Q_c^* B \end{bmatrix} \begin{bmatrix} x_V \\ u_V \end{bmatrix} + \mathrm{tr}(M^T Q_c^* M);
\end{aligned}$$

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{12}^T & H_{22} \end{bmatrix} = \begin{bmatrix} A^T Q_c^* A + C_z^T C_z & A^T Q_c^* B + C_z^T D_z \\ (A^T Q_c^* B + C_z^T D_z)^T & B^T Q_c^* B + D_z^T D_z \end{bmatrix}.$$

Proof of LQG average-cost (3)

$$
\begin{aligned}
&\inf_{u_V \in U} \left\{ \begin{bmatrix} x_V \\ u_V \end{bmatrix}^T Q_{cr} \begin{bmatrix} x_V \\ u_V \end{bmatrix} + E_g[V(x(t+1)) | F^{x_V,\ u_V}] \right\} \\
&= \inf_{u_V \in U} \left\{ \begin{bmatrix} x_V \\ u_V \end{bmatrix}^T H \begin{bmatrix} x_V \\ u_V \end{bmatrix} + \mathrm{tr}(M^T Q_c^* M) \right\} \\
&= \inf_{u_V \in U} \left\{ \begin{bmatrix} x_V \\ u_V + H_{22}^{-1} H_{12}^T x_V \end{bmatrix}^T \begin{bmatrix} Q_c^* & 0 \\ 0 & H_{22} \end{bmatrix} \begin{bmatrix} x_V \\ u_V + H_{22}^{-1} H_{12}^T x_V \end{bmatrix} + \mathrm{tr}(M^T Q_c^* M) \right\}, \\
&\quad \text{by } Q_c^* = f_{CARE}(Q_c^*), \\
&= x_V^T Q_c^* x_V + \mathrm{tr}(M^T Q_c^* M), \\
&\quad \text{by } H_{22} \succ 0 \text{ and for } u_V^* = -H_{22}^{-1} H_{12}^T x_V, \\
&= V(x_V) + J_{ac}^*.
\end{aligned}
$$

Hence (J_{ac}^*, V) is a solution of the DPE-AC.

Proof of LQG average-cost (4)

From the calculation above follows that

$$
\begin{aligned}
g^*(x_V) &= \arg\min_{u_V \in U} \left\{ \begin{bmatrix} x_V \\ u_V \end{bmatrix}^T Q_{cr} \begin{bmatrix} x_V \\ u_V \end{bmatrix} + E_g[V(x(t+1))|F^{x_V,\, u_V}] \right\} \\
&= u_V^* = -H_{22}^{-1} H_{12}^T x_V \\
&= -[B^T Q_c^* B + D_z^T D_z]^{-1} [A^T Q_c^* B + C_z^T D_z]^T x_V = F(Q_c^*) x_V; \\
u_V^* &= g^*(x_V) = F(Q_c^*) x_V,
\end{aligned}
$$

is a time-invariant linear Markov control law.

Finally, use the theorem stated for a general stochastic control system.

Procedure. Computation LQG optimal control law with complete observations and for average cost

1. Check whether conditions (b) or (a) are satisfied.
2. Compute the matrix Q_c^*
   which is a solution of the control algebraic Riccati equation
   with side conditions.
   Use Matlab, but check whether you compute the solution
   of the control Riccati equation or of the filter Riccati equation.
3. Compute

$$\begin{aligned} g^*(x) &= F(Q_c^*)x, \quad \text{the optimal control law and} \\ J_{ac}^* &= \mathrm{tr}(M^T Q_c^* M), \ \text{the value.} \end{aligned}$$

4. Output $(F(Q_c^*),\ J_{ac}^*,\ Q_c^*)$.

Remarks. Controllability and observability

- Controllability versus stabilizability.
- Observability versus detectability.

Proposition. Non controllable GSCS (1)

Consider a time-invariant Gaussian stochastic control system.
Assume the system is supportable, stochastically observable, not stochastically controllable, but stochastically stabilizable.
Then there exists a state space transformation to the form,

$$\overline{x}(t) = Sx(t), \quad S \in \mathbb{R}_{nsng}^{n_x \times n_x},$$

$$\overline{x}(t+1) = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} \overline{x}_1(t) \\ \overline{x}_2(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u(t) + Mv(t), \; x(0) = x_0,$$

$$\text{spec}(A_{22}) \subset \mathbb{D}_o, \text{ by stabilizability.}$$

Note that the second state component $\overline{x}_2$ is not affected by the input at all!

Proposition. Non controllable GSCS (2)

Consider any linear control law and the associated closed-loop system,

$$u(t) = g(\overline{x}(t)) = F\overline{x}(t) = \begin{bmatrix} F_1 & F_2 \end{bmatrix} \overline{x}(t), \; F \in \mathbb{R}^{n_u \times n_x},$$

$$\overline{x}^g(t+1) = \begin{bmatrix} A_{11} + B_1 F_1 & A_{12} + B_1 F_2 \\ 0 & A_{22} \end{bmatrix} \overline{x}^g(t) + Mv(t),$$

$$\text{spec}(A + BF) = \text{spec}(A_{11} + B_1 F_1) \cup \text{spec}(A_{22}) \subset \mathbb{D}_o.$$

- The LQG optimal control law is a linear control law.
- The dynamics of the closed-loop system always include the dynamics A_{22} of the stabilizable part $\overline{x}_2^g$!
- Stochastic controllability implies that: the state component $\overline{x}_2$ is absent from the system; and $\text{spec}(A_{11} + B_1 F_1)$ is placed according to the average cost.
- Conclusion: A system which is stochastically controllable may have a better closed-loop dynamics than a system which is only stochastically stabilizable!

Proposition. Non observable GSCS (1)

Consider a time-invariant Gaussian stochastic control system.
Assume the system is supportable, stochastic controllable,
not stochastic observable, but stochastically detectable.
Then there exists a state space transformation to the form,

$$
\begin{aligned}
\overline{x}(t) &= Sx(t), \;\; S \in \mathbb{R}_{nsng}^{n_x \times n_x}, \\
\overline{x}(t+1) &= \begin{bmatrix} A_{11} & 0 \\ A_{12} & A_{22} \end{bmatrix} \begin{bmatrix} \overline{x}_1(t) \\ \overline{x}_2(t) \end{bmatrix} + Bu(t) + Mv(t), \; x(0) = x_0, \\
z(t) &= \begin{bmatrix} C_{z,1} & 0 \end{bmatrix} \overline{x}(t) + D_z u(t), \\
& (A_{11}, \; C_{z,1}) \text{ observable pair}, \\
& \text{spec}(A_{22}) \subset \mathbb{D}_o \text{ by stochastic detectability.}
\end{aligned}
$$

Note that the state component $\overline{x}_2$
does not affect the controlled output z at all!

Proposition. Non observable GSCS (2)

Consider particular linear control law and the associated closed-loop system,

$$
\begin{aligned}
u^{g_a}(t) &= g_a(\overline{x}(t)) = F_a\overline{x}(t) = \begin{bmatrix} F_{a,1} & 0 \end{bmatrix} \overline{x}(t), \; F_a \in \mathbb{R}^{n_u \times n_x}, \\
\overline{x}^{g_a}(t+1) &= \begin{bmatrix} A_{11} + B_1 F_{a,1} & 0 \\ A_{21} + B_2 F_{a,1} & A_{22} \end{bmatrix} \overline{x}^{g_a}(t) + Mv(t), \\
z^{g_a}(t) &= \begin{bmatrix} C_{z,1} & 0 \end{bmatrix} \overline{x}^{g_a}(t) + D_z u^{g_a}(t), \\
\text{spec}(A + BF_a) &= \text{spec}(A_{11} + B_1 F_{a,1}) \cup \text{spec}(A_{22}) \subset \mathbb{D}_o.
\end{aligned}
$$

- Control objective only: minimize cost function.
  Then $F_{a,1}$ is useful.
- However, the dynamics of the state of the closed-loop system always include the dynamics A_{22} of the stabilizable part $\overline{x}_2$!
  This is not good for stability of the closed-loop system.
- Control objective: minimize cost and have good stability.
  Then it is necessary that
  $g_b(x) = F_b x$ with $F_b = \begin{bmatrix} F_{b,1} & F_{b,2} \end{bmatrix}$ and $F_{b,2} \neq 0$.

Problem. Has the value decreased?

1. Is $J_{ac}^* = J_{ac}(g^*) < J_{ac}(g_z)$
   if $g_z = 0$ and $J_{ac}^* \neq J_{ac}(g_z)$?
   Yes, if the control system is stochastically controllable.
2. Is it true that

$$g_z = 0,\ \forall\, \epsilon \in (0,1)\ \exists\ g_\epsilon^* \in G \text{ such that}$$

$$0 < \frac{J_{ac}(g_\epsilon^*)}{J_{ac}(g_z)} < \epsilon? \ \Leftrightarrow\ 0 < J_{ac}(g_\epsilon^*) < \epsilon\, J_{ac}(g_z)?$$

- Answer is no in general!
- There exists an optimal control problem where an ϵ exists only near 1.
- Check the linear system from the input u to the controlled output z.
  If there does exist a zero of this linear system
  inside the unit disc near the unit circle
  then there is a lower bound on the fraction.
- See Section 22.2, p. 806–808.
  J.C. Doyle has written about this.

Examples. List

Optimal stochastic control with complete observations,
on a half-infinite horizon, with average cost.

1. The control of the paper machine,
   cost function is modified to include the input variable,
   on a half-infinite horizon.
   K.J. Aström used a different formulation with partial observations.
2. Control for course keeping of ship at sea, see Lecture 7.
3. Homework Set 9, Exercise 1.

Outline

Problem. Finite case of the optimal stochastic control problem with average cost

$$\begin{aligned} T = \mathbb{N} = \{0,1,2,\ldots\},&\ X,\ U,\ \text{finite sets}, \\ x(t+1) = f(x(t),u(t),v(t)),&\, x(0) = x_0; \\ & f : X \times U \times W_v \rightarrow X;\ x_0 : \Omega \rightarrow X \text{ is a random variable}, \\ & v : \Omega \times T \rightarrow \mathbb{R}^{n_v},\ \{v(s),\ \forall\, s \in T\},\ \text{independent rvs.} \end{aligned}$$

- Past state information pattern $\{F_t^x,\ t \in T\}$.
- Set G_{tv} of time-varying past-state control laws, for all $t \in T$, $g_t : X^{t+1} \rightarrow U$.
- Closed-loop stochastic control system,

$$\begin{aligned} x^g(t+1) &= f(x^g(t), g(t, x^g(0:t)), v(t)), x^g(0) = x_0, \\ u^g(t) &= g(t, x^g(0:t)). \end{aligned}$$

Problem. Optimal stochastic control problem with average cost

- Cost rate,

$$b : X \times U \to \mathbb{R}_+, \quad X,\ U \text{ finite sets,}$$
$$b_{max} = \max_{(x,u)\in X\times U} b(x,u) \in \mathbb{R}_+, \text{ exists.}$$

- Average cost function is finite,

$$J_{ac}(g) = \limsup_{t_1\to\infty} \frac{1}{t_1} E_g[\sum_{s=0}^{t_1-1} b(x^g(s), u^g(s))] \leq b_{max} < +\infty,$$
$$J_{ac} : G \to \mathbb{R}_+.$$

Every control law has finite average cost!

Determine an optimal control law $g^* \in G$ such that,

$$J_{ac}^* = J_{ac}(g^*) = \inf_{g\in G} J_{ac}(g).$$

Comments. Optimal stochastic control problem with average cost

- There exists an example (Example 13.2.19)
  for which a time-varying control law
  has a strictly lower cost
  than any time-invariant control law!
- S. Ross has formulated this example (see his book 1970).
- There is no control synthesis procedure
  for time-varying control laws.
- Below attention is restricted to time-invariant control laws.

Comments. Control theory issues

- How to derive the dynamic programming equation of average cost?
- Does there exist a pair (J_{ac}^*, V) which is a solution of the dynamic programming equation? Existence requires stochastic controllability.
- Does the dynamic programming equation have a unique solution?
- How to compute a value function of the dynamic programming equation?
- How to derive an optimal control law $g^* \in G$ from the value function?
- Why is stochastic controllability not used in the classical literature?

Notation,

$$
\begin{aligned}
P(x_1, x_0, u_a) &= P(\{x(t+1) = x_1\} | \{x(t) = x_0,\ u(t) = u_a\}) \in \mathbb{R}_+, \\
A(u_a)_{(x_1,x_0)} &= P(\{x(t+1) = x_1\} | \{x(t) = x_0,\ u(t) = u_a\}), \\
A(g)_{(x_1,x_0)} &= P(\{x(t+1) = x_1\} | \{x(t) = x_0,\ u(t) = g(x_0)\}).
\end{aligned}
$$

Procedure. Optimal control law from the DPE-AC

Data: $n_x \in \mathbb{R}_+$, $b : X \times U \to \mathbb{R}_+$ and $P : X \times X \times U \to \mathbb{R}^{n_x \times n_x}$.

1. Solve the dynamic programming equation of average cost

$$J_{ac}^* + V(x_V) = \min_{u_V \in U(x_V)} \{b(x_V, u_V) + \sum_{x_1 \in X} P(x_1, x_V, u_V) V(x_1)\}$$

   for $V : X \to \mathbb{R}$ and $J_{ac}^* \in \mathbb{R}$.
   If a solution V exists then it is called the value function.

2.

$$\begin{aligned} & \forall\ x_V,\ \exists\ u^* \in U(x_V),\ \text{exists because } U \text{ is a finite set,} \\ & b(x_V, u^*) + \sum_{x_1 \in X} P(x_1, x_V, u^*) V(x_1) \\ = & \min_{u_V \in U(x_V)} \{b(x_V, u_V) + \sum_{x_1 \in X} P(x_1, x_V, u_V) V(x_1)\},\ \text{then,} \end{aligned}$$

$$g^*(x_V) = u^*,\ \text{where } g^* : X \to U.$$

Theorem. Sufficient condition for optimality

Consider the optimal stochastic control problem
for a finite-stochastic control system with average cost and
with complete observations on an infinite horizon.

1. If there exists a solution (J^*_{ac}, V)
   of the dynamic programming equation of average cost, and
2. if the control law $g^* \in G$ is produced
   by the procedure stated above,

then

(a) g^* is an optimal control law and

(b) J^*_{ac} is the value of the problem.

Def. Reducible and irreducible stochastic matrices
A stochastic matrix $P \in \mathbb{R}_{st}^{n_x \times n_x}$ is said to be reducible
if there exists a permutation matrix $Q \in \mathbb{R}^{n \times n}$ such that

$$QPQ^{-1} = \begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix},$$
$$P_{11} \in \mathbb{R}_{+}^{n_1 \times n_1},\ P_{22} \in \mathbb{R}_{+}^{n_2 \times n_2},\ n_1,\ n_2 \in \mathbb{Z}_{+},\ n_1 + n_2 = n;$$

with the diagonal block matrices P_{11} and P_{22} being square matrices.
If there does not exist such a permutation matrix
then P is said to be irreducible.

See Chapter 18 of the book for reducible and irreducible matrices.
Concept formulated by G. Frobenius (1908, 1909).
Note that for a permutation matrix $Q \in \mathbb{R}_{perm}^{n \times n}$, $Q^{-1} = Q^T$.

Theorem. Solution DPE (1)

Consider average-cost optimal stochastic control problem - Finite case.
Assume that the transition matrix $P(g)$ is irreducible
for all time-invariant Markov control laws $g \in G_M$.

(a) Existence value and value function.

$$\begin{aligned}
& \exists\ J_{ac}^* \in \mathbb{R}, \ \ \exists\ V : X \to \mathbb{R}, \\
& \text{which are a solution of the DP equation,} \\
J_{ac}^* + V(x_V) = & \min_{u_V \in U} \{ b(x_V,\ u_V) + \sum_{x_1 \in X} P(x_1,\ x_V,\ u_V) V(x_1) \}, \\
& \forall\ x_V \in X.
\end{aligned}$$

(b) Uniqueness value and value function.
The solution V of the dynamic programming equation
is unique up to an additive constant.
The minimal cost, $J_{ac}^* \in \mathbb{R}$, is unique.

Theorem. Solution DPE (2)

(c) Sufficient condition for optimality.
Consider a solution $(J_{ac}^*,\ V)$ of the DPE-AC.
Let g^* be as constructed in Step (2) of the DP Procedure.
Then $g^* \in G_{SM}$ is an optimal control law and J_{ac}^* is the value.

(d) Necessity condition for optimality.
If $g_a \in G_{SM}$ is such that

$$J_{ac}(g_a) = \inf_{g \in G} J_{ac}(g),$$

then the dynamic programming equation holds with $V(g_a)$ and g_a attains the minimum in Step (2) of the average cost dynamic programming procedure.

Proof analogous to the case of a Gaussian stoc. control system.

Problem. How to solve the DP equation?

Available procedures:

1. The value iteration procedure;
   may not converge in a finite number of steps; and
2. the policy iteration procedure (control-law iteration);
   a condition implies it converges in a finite number of steps.

Comments

- A policy is a term of
  the research area of operations research (part of mathematics)
  which is identical to a control law.
- Experience indicates that most effective is
  a mixed procedure in which alternatingly
  a finite number of steps of both procedures are used.
- The mixed procedure has been applied successfully
  in the research group of the lecturer.

Procedure. Policy iteration (1; control law iteration)

Data: $b : X \times U \to \mathbb{R}_+$, $A : U \to \mathbb{R}_{st}^{n_x \times n_x}$,
$g_0 : X \to U$ a time-invariant Markov control law.

1. Initialization. Let $m = 0$. Solve the equation

$$\begin{aligned} & J(g_0)1_{n_x} + V(g_0) = b(g_0) + A(g_0)V(g_0), \\ \Leftrightarrow\ & J(g_0)1_{n_x} - b(g_0) = [A(g_0) - I]V(g_0). \end{aligned}$$

for $V(g_0) \in \mathbb{R}^n$ and $J(g_0) \in \mathbb{R}$ where,

$$b(g_0) = \begin{bmatrix} b(x_1, g_0(x_1)) \\ \vdots \\ b(x_n, g_0(x_n)) \end{bmatrix},\ V(g_0) = (.),\ 1_{n_x} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Subtract the last equation from the preceeding $(n-1)$ equations to obtain a set of $(n-1)$ equations without $J(g_0)$ from which $V(g_0)$ can be determined up to an additive constant. Then compute $J(g_0)$.

Procedure. Policy iteration (2)

2 For $m := m + 1$ while

$$J(g_{m-1})1_{n_x} + V(g_{m-1}) > \min_{g \in \mathbb{R}^{n_x}} \{b(g) + A(g)V(g_{m-1})\}, \text{ do,}$$

(2.1) Let $g_m = \arg\min_{g \in \mathbb{R}^{n_x}} \{b(g) + A(g)V(g_{m-1})\}$.
(2.2) Solve the equation

$$J(g_m)e + V(g_m) = b(g_m) + A(g_m)V(g_m)$$

for $J(g_m) \in R$ and $V(g_m) \in \mathbb{R}^{n_x}$.

3 Let $g^* = g_m$.

Proposition

Assume that the $A(g) \in \mathbb{R}_{st}^{n_x \times n_x}$ is irreducible $\forall\, g \in G_{SM}$.
Then the policy iteration procedure
yields an optimal control law in a finite number of steps.

Problem. Control of inventory (1)

Consider a finite stochastic control system with,

$$X = \{1, 2\},\ U = \{u_1, u_2\},$$

$$A(u_1) = \frac{1}{4}\begin{bmatrix} 3 & 3 \\ 1 & 1 \end{bmatrix},\ A(u_2) = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix},$$

$$b(1, u_1) = 2,\ b(1, u_2) = 0.5,$$

$$b(2, u_1) = 1,\ b(2, u_2) = 3.$$

Policy iteration procedure is applied.

1. Initialization.
   Let $g_0(1) = u_1$ and $g_0(2) = u_2$, and $g_0 : X \to U$.
   Let $m = 0$.

Problem. Control of inventory (2)

(2) Solve the equations for $V(g_0)$,

$$J(g_0)1 + V(g_0) = b(g_0) + A(g_0)V(g_0), \quad V(g_0,1) = 0;$$

$$J(g_0)\begin{bmatrix}1\\1\end{bmatrix} + \begin{bmatrix}V(g_0,1)\\V(g_0,2)\end{bmatrix} = \begin{bmatrix}2\\3\end{bmatrix} + \frac{1}{4}\begin{bmatrix}3 & 1\\1 & 3\end{bmatrix}\begin{bmatrix}V(g_0,1)\\V(g_0,2)\end{bmatrix},$$

$$\Rightarrow$$

$$\begin{bmatrix}V(g_0,1)\\V(g_0,2)\end{bmatrix} = \begin{bmatrix}0\\2\end{bmatrix},$$

$$J(g_0) = 2.5.$$

(3) Determine $g_1 : X \to U$ such that

$$b(g_1) + A(g_1)v(g_0) = \min_{g \in \mathbb{R}^n}\{b(g) + A(g)V(g_0)\}.$$

$$m = 1.$$

Problem. Control of inventory (3)

$$\begin{aligned}
&\min\{b(1,u_1)+A_{11}(u_1)V(g_0,1)+A_{1,2}(u_1)V(g_0,2),\\
&\ b(1,u_2)+A_{11}(u_2)V(g_0,1)+A_{1,2}(u_2)V(g_0,2)\}\\
&=\min\{2+\frac{1}{4}\times 2, 0.5+\frac{3}{4}\times 2\}=\min\{2.5,2\}=2.\\
&\min\{b(2,u_1)+A_{21}(u_1)V(g_0,1)+A_{22}(u_1)V(g_0,2),\\
&\ b(2,u_2)+A_{21}(u_2)V(g_0,1)+A_{22}(u_2)V(g_0,2)\}\\
&=\min\{1+\frac{1}{4}\times 2, 3+\frac{3}{4}\times 2\}=\min\{1.5,4.5\}=1.5;
\end{aligned}$$

$$g_1(1)=u_2,\ g_1(2)=u_1.$$

Problem. Control of inventory (4)

$$\begin{aligned}
J(g_0)e + V(g_0) &= 2.5 \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 4.5 \end{bmatrix} \\
&> \begin{bmatrix} 2 \\ 1.5 \end{bmatrix} = \min_{g \in R^2} \{ b(g) + A(g)V(g_0) \} \\
&= b(g_1) + A(g_1)V(g_0). \\
&\quad \text{Solve for } V(g_1),\ J(g_1), \\
J(g_1)1_{n_x} + V(g_1) &= b(g_1) + A(g_1)V(g_1), V(g_1, 1) = 0, \\
J(g_1)1_{n_x} + \begin{bmatrix} V(g_1,1) \\ V(g_1,2) \end{bmatrix} &= \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} V(g_1,1) \\ V(g_1,2) \end{bmatrix}, \\
V(g_1, 1) &= 0, \\
V(g_1) &= \left(\begin{bmatrix} 0 \\ \frac{1}{3} \end{bmatrix} \right),\ J(g_1) = 0.75.
\end{aligned}$$

Problem. Control of inventory (5)

$m = 2$. Determine, $g_2 : X \to U$ such that

$$b(g_2) + A(g_2)V(g_1) = \min_{g \in \mathbb{R}^2} \{b(g) + A(g)V(g_1)\},$$

$$\min\{b(1,u_1) + A_{11}(u_1)V(g_1,1) + A_{12}(u_1)V(g_1,2),$$
$$b(1,u_2) + A_{11}(u_2)V(g_1,1) + A_{12}(u_2)V(g_1,2)\}$$
$$= \min\{2 + \frac{1}{4} \times \frac{1}{3}, 0.5 + \frac{3}{4} \times \frac{1}{3}\} = \min\{2.08, 0.75\} = 0.75,$$

$$\min\{b(2,u_1) + A_{21}(u_1)V(g_1,1) + A_{22}(u_1)V(g_1,2),$$
$$b(2,u_2) + A_{21}(u_2)V(g_1,1) + A_{22}(u_2)V(g_1,2)\}$$
$$= \min\{1 + \frac{1}{4} \times \frac{1}{3}, 3 + \frac{3}{4} \times \frac{1}{3}\} = \min\{1.08, 3.25\} = 1.08,$$

$g_2(1) = u_2,\ g_2(2) = u_1.$

$g_2 = g_1,\ \ g^* = g_1$ is an optimal control law

$J_{ac}^* = 0.75,\ \ V = (V(1), V(2)) = (0, 1/3).$

Outline

Stochastic control by system approximation

What to do with an optimal stochastic control problem
for which no analytic solution for the value function exists?

- Consider a continuous-time stochastic control system with a continuous state-space.
- Approximate that system by a discrete-time finite-state stochastic system.
- Apply the DP procedure for the optimal stochastic control problem of the finite stochastic control system (FSCS).
- Transform the optimal control from the discrete case to an optimal control law for the continuous state-space stochastic control system.
- Approach above to be used if no analytic results seem possible.
- Convergence analysis proven in H.J. Kushner, 1971 and P. Dupuis, H. Kushner, 2001 (2nd Ed.).

Further reading

- Average cost.
  D.P. Bertsekas, 1976 (Chapter 8).
  P.R. Kumar, P. Varaiya 1986 (Sections 4.5, 8.5, 8.6).
  H. Kushner, 1971 (Chapter 6, Section 9.4).
  D. Blackwell, 1970.
- Discounted cost.
  D.P. Bertsekas, 1976.
  D. Blackwell, 1965 (two references in this year).
- Continuous-time optimal stochastic control.
  W.H. Fleming, R.W. Rishel (1975; book).
  Uses PDE theory.

Overview of Lecture 9

- Optimal stochastic control problems
  with complete observations and on an infinite horizon.
  Average cost and discounted cost.
- Control theory for average cost functions.
  Dynamic programming equation for the value function and the value.
  From the value function to the optimal control law.
- Case Gaussian stochastic control system.
  Usefulness of stochastic controllability and of stochastic observability.
- Case finite stochastic control system.
  The value iteration procedure.
- System approximation and optimal stochastic control.