Control of Stochastic Systems
Lecture 8
Optimal Stochastic Control
with Complete Observations
on a Finite Horizon

Jan H. van Schuppen

29th of April 2025

Delft University of Technology
Delft, The Netherlands

Outline

Outline

Example Control

Problem. Control of freeway traffic flow

- In The Netherlands there has been installed on freeways or motorways, a freeway control and signalling system.
- Matrix boards display advisory speeds to drivers.
- Measurements of detection loops in the road surface.
  Provide information of the passage and of the speed of a each car.
- Problem of control.
  Determine how to set advisory speed
  so as to maximize the traffic flow?
  Traffic queues to be avoided or
  formation of such queues to be postponed.
  When a traffic queue occurs, the road capacity is much lower.
- Project 1986 – 1990 at CWI
  in cooperation with the government agency Rijkswaterstaat.

Example Control

Problem. Control of freeway traffic flow

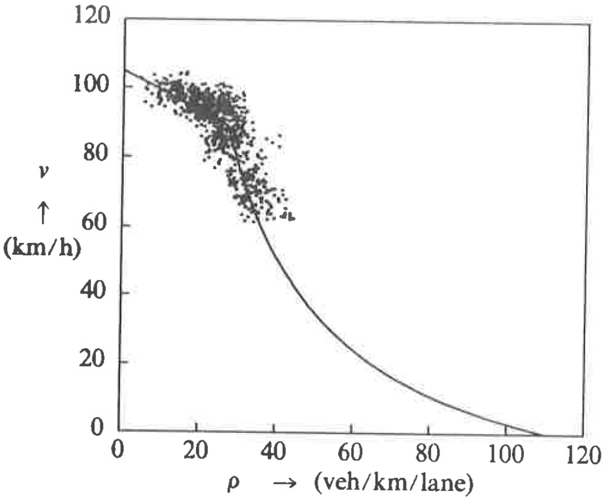
Simplified model of one section, continuous-time, stochastic control system. Description in terms of a stochastic differential equation.

$$
\begin{aligned}
d\rho(t) &= \frac{1}{Ll}[\lambda_0 - l\rho(t)v(t)]dt + \sigma_1 dw_1(t), \\
dv(t) &= -\frac{1}{t_r}[v(t) - v_e(\rho(t))]dt + \sigma_2 dw_2(t); \\
\rho(t) &\quad \text{car density in veh/km.lane,} \\
v(t) &\quad \text{average speed in km/h of all cars in section;} \\
(\rho_s, v_s),\ (\rho_u, v_u) &\quad \text{steady states of deterministic system;} \\
&\quad \text{behavior of fluctuations;} \\
&\quad \text{if car density is too high, the system goes unstable.}
\end{aligned}
$$

Effect of advisory speed is primarily a variance reduction. See figures.

Example Control

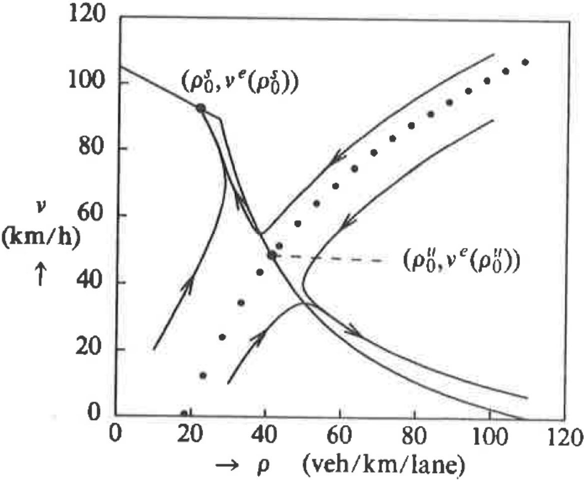
Problem. Control of freeway traffic flow



Example Control

Problem. Control of freeway traffic flow

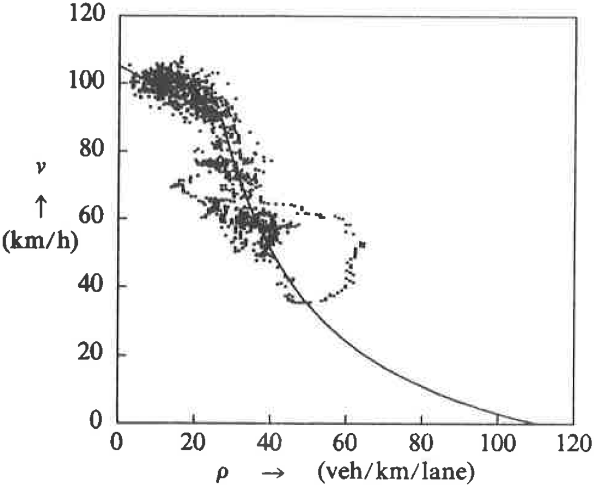
State-space trajectories of uncontrolled deterministic system.
Traffic density (veh/km.lane) and average speed (km/h).



Example Control

Problem. Control of freeway traffic flow

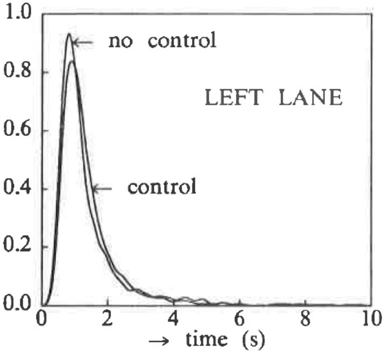
State space trajectories of actual congested traffic.
Traffic density (veh/km.lane) and average speed (km/h).

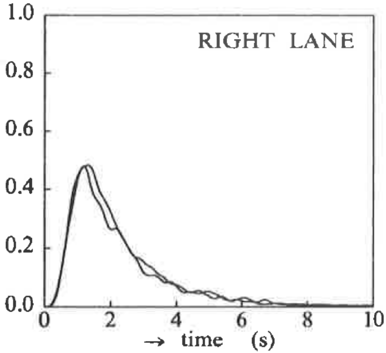


Example Control

Problem. Control of freeway traffic flow

Headway time is duration at a road location
between the arrival times of two subsequent vehicles.
Figure of the probability densities of headway time,
without and with control, on a motorway with two lanes.





Outline

Problem. Optimal stochastic control on a finite horizon

$$T = \mathbb{N}_{t_1} = \{0, 1, 2, \dots, t_1\},\ t_1 \in \mathbb{Z}_+,$$

$$X = \mathbb{R}^{n_x},\ U = \mathbb{R}^{n_u},$$

$$x(t+1) = f(t, x(t), u(t), v(t)),\ x(0) = x_0,$$

$$G = \left\{ \begin{array}{l} g = \{g_0,\ g_1,\ \dots,\ g_{t_1-1}\} \mid \forall\, t \in T\backslash\{t_1\} \\ g_t : X^{t+1} \to U \ \text{ is a measurable map} \end{array} \right\},$$

$$x^g(t+1) = f(t, x^g(t), g_t(x^g(0:t)), v(t)),\ x^g(0) = x_0,$$

$$u^g(t) = g_t(x^g(0:t));\ \ J : G \to \mathbb{R}_+ \cup \{+\infty\},$$

$$J(g) = E\left[\left(\sum_{s=0}^{t_1-1} b(s, x^g(s), g_s(x^g(0:s))\right) + b_1(x^g(t_1))\right],$$

$$b : T \times X \times U \to \mathbb{R}_+,\ b_1 : X \to \mathbb{R}_+,\ \text{measurable};$$

$$\inf_{g\in G} J(g);$$

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

Comments on problem

- Recursive stochastic control system, nonlinear in general.
- Call b the cost rate and b_1 the terminal cost.
- Call J^* the value and call $g^* \in G$ an optimal control law.
- $J^* \in \mathbb{R}_+ \cup \{+\infty\}$ exists because of positive costs.
- $g^* \in G$ may or may not exist.
  Example, from optimization, $\inf_{u \in (0,1)} \; u = 0$; note that $0 \notin (0, 1)$.
- Call for any $\epsilon \in (0, \infty)$
  $g_\epsilon \in G$ an ϵ-optimal control law if

  $$J^* < J(g_\epsilon) < J^* + \epsilon.$$

  Not discussed further. Used in information theory.
- Is $J(g^*) < J(g_z)$
  where $g_z = 0$ and if $J(g^*) \neq J(g_z)$?
  In general one needs a condition of stochastic controllability.

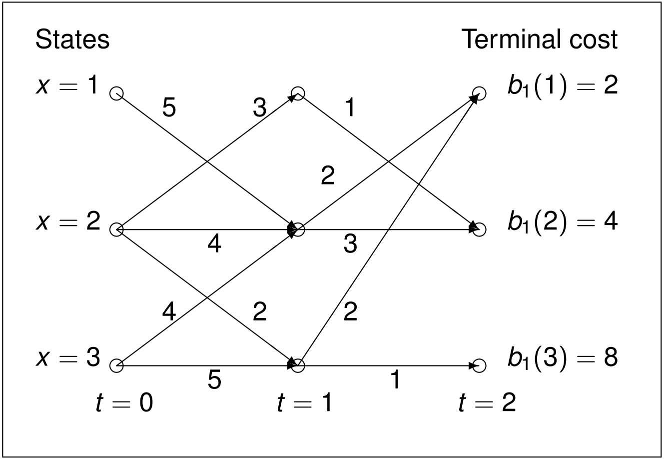
Approaches to Optimal Control

- Approaches to deterministic and stochastic optimal control:
  - Maximal principle, by Pontryagin etal.
    (1962 English translation of Russian original)
    Provides only necessary conditions of optimality.
  - Dynamic programming, by R. Bellman (1957) and others.
    Provides sufficient and necessary conditions.
- Dynamic programming preferred by lecturer.
- General approach to dynamic programming
  using P-essential infimum of C. Striebel (1976).
  Concept of conditional optimality.
  Relation with martingale theory.
  See Section 16.2.
- Approach of Steve Shreve (1978) for discrete-time problems,
  attention for measurability of control laws.
  See Section 12.5.

Outline

Dynamic Programming

Example. Introduction to dynamic programming (1)



Dynamic Programming

Example. Introduction DP (2)

- The value function $V(t, x)$ denotes the minimal cost-to-go from $(t,\ x)$ to a tuple $(t_1,\ x_1)$ at the terminal time $t_1 = 2$ for a state $x_1 \in X$.
- The value function at the terminal time $t_1 = 2$ is

$$\begin{aligned} V(t_1,\ x_V) &= b_1(x_V) \\ &= \begin{cases} 2, & \text{if } x = 1, \\ 4, & \text{if } x = 2, \\ 8, & \text{if } x = 3. \end{cases} \end{aligned}$$

Notation $x_V \in X$ and $u_V \in U$.

Dynamic Programming

Example. Introduction DP (3)

- Compute the value function by a backward recursion, from the terminal time $t_1 = 2$ to the initial time $t = 0$.
- The backward recursion of the value function is

$$\begin{aligned} & \forall\ t = t_1 - 1,\ t_1 - 2,\ \ldots,\ 0, \\ V(t, x_V) &= \min_{u_V \in U(x_V)} \{b(t, x_V, u_V) + V(t+1, x(t+1))\} \\ &= \min_{u_V \in U(x_V)} \{b(t, x_V, u_V) + V(t+1, f(t, x_V, u_V))\}. \end{aligned}$$

Dynamic Programming

Example. Introduction DP (4)

Compute the minimal cost on the remaining horizon $V(t, x_V)$.

$$
\begin{aligned}
V(t, x_V) &= \min_{u_V \in U(x_V)} \{b(t, x_V, u_V) + V(t+1, f(t, x_V, u_V))\};\\
V(1,1) &= \min_{u_V \in U(1)=\{2\}} \{b(1,1,2) + V(2,2)\}\\
&= \min_{u_V \in U(1)=\{2\}} \{b(1,1,2) + b_1(2)\} = 1 + 4 = 5,\\
V(1,2) &= \min_{u_V \in U(1,2)=\{1,2\}} \{b(1,2,u_V) + V(2, f(1,2,u_V))\}\\
&= \min\{2+2, 3+4\} = \min\{4,\ 7\} = 4,\\
V(1,3) &= \min_{u_V \in U(1,3)=\{1,3\}} \{b(1,3,u_V) + V(2, f(1,3,u_V))\}\\
&= \min\{2+2, 1+8\} = \min\{4,\ 9\} = 4;\\
g^*(1, x) &= \mathrm{argmin}_{u_V \in U(x_V)} \{\ldots\} = \begin{cases} 2, & \text{if } x = 1,\\ 1, & \text{if } x = 2,\\ 1, & \text{if } x = 3. \end{cases}
\end{aligned}
$$

Dynamic Programming

Comments. Introduction to dynamic programming (5)

- Further computations, see Example 12.3.1 in book.
- One computes the value function $V(t,\ x)$ for all times $t \in T$ and for all states $x \in X$.
- During the backward recursion one also computes the optimal control law $g^*(t, x)$ for all $(t,\ x)$.
- You may already have used dynamic programming. For example, to compute the time of departure for a trip so as to arrive at a destination before a particular time.

Outline

Notation

$$
\begin{aligned}
& E[V(t+1, f(t, x^g(t), u^g(t), v(t)))|F^{x^g(t),u^g(t)}] \\
& E[V(t+1, f(t, x_V, u_V, v(t))|F^{x_V,u_V}] \\
& = \int V(t+1, f(t, x_V, u_V, w))\ f_{v(t)}(dw),
\end{aligned}
$$

$f_{v(t)}$ denotes the pdf of $v(t)$; $v(t)$ independent of $(x^g(t),\ u^g(t))$.

Assumption. Finite cost

$$
\begin{aligned}
& \forall\ g \in G, \\
& E[b(t, x^g(t), u^g(t))] < \infty,\ \forall\ t \in T \backslash \{t_1\}, \\
& E[b_1(x^g(t_1))] < \infty; \\
G_{fc} &= \{g \in G|\ J(g) < \infty,\ \text{finite costs of } g \text{ holds}\}; \\
G &= G_{fc} \neq \emptyset,\ \text{assumed.}
\end{aligned}
$$

If $G_{fc} = \emptyset$ then costs of control laws are all equal, no distinction.

Def. Conditional cost to go

$$
\begin{aligned}
&\forall\ g \in G_{fc},\ \text{define},\\
x^g(t+1) &= f(t, x^g(t), g_t(x^g(0:t)), v(t)),\ x^g(0) = x_0,\\
u^g(t) &= g_t(x^g(0:t)),\\
&J : G_{fc} \times \Omega \times T \to \mathbb{R}_+,\\
J(g,t) &= E[\sum_{s=t}^{t_1-1} b(s, x^g(s), u^g(s)) + b_1(x^g(t_1)) | F_t^{x^g}]\\
&= E[\sum_{s=t}^{t_1-1} b(s, x^g(s), g_s(x^g(0:s))) + b_1(x^g(t_1)) | F_t^{x^g}],\\
J(g) &= E[J(g,0)],\ \text{by a property of conditional expectation.}
\end{aligned}
$$

Call $J(g,\ t)$ the conditional cost-to-go
of $g \in G_{fc}$ from $(t,\ x^g(t))$ at time $t \in T$ till the terminal time t_1.

Procedure. Dynamic programming on finite horizon (1)

1. Define $V(t_1, x_V) = b_1(x_V)$ for all $x_V \in X$.
2. For $t = t_1 - 1, t_1 - 2, \ldots, \; 2, \; 1, \; 0$, define for all $x_V \in X$,

(2.A)

$$V(t, x_V) = \inf_{u_V \in U(t, x_V)} \{b(t, x_V, u_V) + \\ + E[V(t+1, f(t, x_V, u_V, v(t)))|F^{x_V, u_V}]\} \\ = \inf \;\; h(t, \; x_V, \; u_V);$$

(2.B) if $\forall \; x_V \in X, \; \exists \; u^* \in U(t, x_V) \subseteq U$ such that

$$b(t, x_V, u^*) + E[V(t+1, f(t, x_V, u^*, v(t)))|F^{x_V, u^*}] \\ = \inf_{u_V \in U(t, x_V)} \{b(t, x_V, u_V) + \\ + E[V(t+1, f(t, x_V, u_V, v(t)))|F^{x_V, u_V}]\},$$

then define

$$g_t^*(x_V) = u^*; \;\; g_t^* : X \to U;$$

note, g_t^* depends only on x_v, not on past states.

Procedure. Dynamic programming on finite horizon (2)

(3) Check if $\forall\, t \in T$, $g_t^*(.) : X \to U$ and $V(t,.) : X \to \mathbb{R}_+$ are measureable functions; stop if not.
For details,
see the Sections 12.5 and 12.15 of the lecture notes and the introduction of the book [12] of the reference list of Chapter 12.

(4) Output (V, g^*).
One may write $g^* : T \times X \to U$.

Comments on Procedure DP (1)

- The dynamic programming procedure is a backward recursion.
- Note that one calculates $V(t,x)$ and $g^*(t,x)$ for all $(t,\ x) \in T \times X$, regardless of what the states will be during the control system operation.
- The input space $U(t,x_V) \subseteq U$ for all $t \in T$ may depend on the time t and on the current state x_V. Useful in examples.
- At each time $t \in T$, one infimizes the sum of the cost rate and the current estimate of the future minimal cost on the horizon $\{t+1,\ \ldots,\ t_1\}$.
- Note that in Step (2.B) one assumes the existence of a minimizer hence that the infimum is attained. See for sufficient conditions for existence, Section 12.4.
- Step 3 on measurable functions is rather technical, see Section 12.5.

Comments on Procedure DP (2)

- Principle of optimality is mentioned in the literature.
  Principle is a restriction of choices.
  Principle of optimality does not make sense to lecturer.
  See proof of theorem below.
- In this lecture, the time axis
  $T = \mathbb{N}_{t_1} = \{0,\ 1,\ 2,\ \ldots,\ t_1\}$ is a totally ordered set.
  Dynamic programming procedure
  well defined by backward recursion.
- If the time set is a partially-ordered set,
  then the dynamic programming procedure requires attention.
- Example, T is a graph.
  E.G. Dijkstra's algorithm for the shortest path in a weighted graph
  is dynamic programming on a partially-ordered set.
- Dynamic programming of decentralized control
  has not been properly defined.

Theorem. Sufficient and necessary condition for optimality - 1

Theorem 12.6.4 of the lecture notes.

(a) Sufficient condition. A lower bound on the value J^*.
Consider V as produced by the Procedure of DP. Then,

$$\begin{aligned}
& \forall\, g \in G, \\
V(t, x^g(t)) \leq J(g,t) &= E\left[\sum_{s=t}^{t_1-1} b(s, x^g(s), u^g(s)) + b_1(x^g(t_1)) | F_t^{x^g(t)}\right], \\
E[V(0,x_0)] &\leq E[J(g,0)] = J(g), \\
E[V(0,x_0)] &\leq J^* = \inf_{g \in G} J(g).
\end{aligned}$$

- Expectation of the value function $E[V(0,x_0)]$ is a lower bound of the infimal cost J^*.
- Value function $V(t, x^g(t))$ at state of trajectory $x^g(t)$ is almost surely a lower bound of the conditional cost-to-go $J(g,t)$. C. Striebel (1976) uses the term conditional optimality at time $t \in T$.

Theorem. Sufficient and necessary condition for optimality - 2

(b c) Existence of an optimal control law.

$$
\begin{aligned}
& \text{if } \forall\ x_V,\ \exists\ u^* \in U(t, x_V), \\
& \text{such that infimal cost is attained,} \\
g_t^*(x_V) &= u^*,\ \text{and if } g_t^* : X \to U,\ \text{is measurable, then} \\
V(t, x^{g^*}(t)) &= J(g^*, t); \quad \text{if this holds for all } \ t \in T \backslash \{t_1\},\ \text{then,} \\
J(g^*) &= E[J(g^*, 0)] = E[V(0, x_0)] \\
&\leq J^* = \inf_{g \in G} J(g) \leq J(g^*), \\
J^* &= E[V(0,\ x_0)] = J(g^*).
\end{aligned}
$$

Thus $g^* \in G$ is an optimal control law.
It is a unique optimal control law
if, for all $t \in T$, strict convexity holds in the DP recursion.
g^* is a Markov control law; by definition,
$g_t^*(x^{g^*}(t))$ does not depend on the past states $x^{g^*}(0 : t-1)$.

Dynamic Programming

Theorem. Sufficient and necessary condition for optimality - 3

- The value function:
  - was defined as the minimal cost-to-go from $(0,\ x_0)$ to $(t_1,\ x_1)$;
  - was defined as the outcome of the Procedure Dynamic Programming.

  Are these definitions consistent?
- The equality,

$$E[V(0, x_0)] = J(g^*) = J^* = \inf_{g \in G} J(g),$$

  justifies calling V, defined in the Procedure DP, the value function of the stochastic optimal control problem.

Dynamic Programming

Theorem. Sufficient and necessary condition for optimality - 4

(d) Necessary condition.
If there exist V satisfying the DP procedure and
if there exists a Markov control law $g^* \in G_M$ which is optimal
then,

$$\begin{aligned} V(t_1, x^{g^*}(t_1)) &= b_1(x^{g^*}(t_1)), \text{ and} \\ V(t, x^{g^*}(t)) &= b(t, x^{g^*}(t), g_t^*(x^{g^*}(t))) + \\ &\quad + E[V(t+1, f(t, x^{g^*}(t), g_t(x^{g^*}(t)), v(t)))|F_t^{x^{g^*}}], \\ &\forall\, t \in T(0 : t_1 - 1). \end{aligned}$$

Thus infima in the DP procedure are attained
for $u^*(t) = g_t^*(x^{g^*}(t))$.
The above relations hold for the optimal state trajectory x^{g^*}!

Proof of theorem based on two lemmas stated next.
Proof outline due to D. Blackwell (U. California at Berkeley).

Lemma. Comparison principle

$$
\begin{aligned}
& V : T \times X \to \mathbb{R}_+; \text{ if} \\
V(t_1, x_V) \le {} & b_1(x_V), \forall x_V \in X, \\
V(t, x_V) \le {} & b(t, x_V, u_V) + E[V(t+1, f(t, x_V, u_V, v(t)))|F_t^{x_V, u_V}], \\
& \forall\ x_V \in X,\ \forall\ u_V \in U(t, x_v),\ \forall\ t = 0, 1, \ldots, t_1 - 1; \\
& \text{then,} \\
V(t, x^g(t)) \le {} & J(g, t) \text{ a.s. } \forall\ t \in T,\ \forall\ g \in G.
\end{aligned}
$$

Thus $V(t, x^g(t))$ is a lower bound of $J(g, t)$
for all control laws $g \in G$ and for all times $t \in T$.

Proof of Lemma. Comparison principle - 1

$$
\begin{aligned}
V(t_1, x^g(t_1)) &\leq b_1(x^g(t_1)), \text{ by assumption,} \\
&= E[b_1(x^g(t_1))|F_{t_1}^{x^g}], \\
&\quad \text{by a result of conditional expectation,} \\
&= J(g, t_1), \text{ by definition of } J(g, .).
\end{aligned}
$$

Suppose that the inequality holds for all $t+1, t+2, \ldots, t_1$.
Inequality is to be proven for t.

Proof of Lemma. Comparison principle - 2

$$
\begin{aligned}
& V(t, x^g(t)) \\
& \leq b(t, x^g(t), u^g(t)) + E[V(t+1, f(t, x^g(t), u^g(t), v(t)))|F_t^{x^g}], \\
& \quad \text{by the assumption for } V \text{ and } F^{u^g(t)} \subseteq F_t^{x^g}, \\
& = b(t, x^g(t), u^g(t)) + E[V(t+1, x^g(t+1))|F_t^{x^g}], \\
& \leq b(t, x^g(t), u^g(t)) + E[J(g, t+1)|F_t^{x^g}], \text{ by the induction hypothesis} \\
& = b(t, x^g(t), u^g(t)) + \\
& \quad + E[\, E[\sum_{s=t+1}^{t_1-1} b(s, x^g(s), u^g(s)) + b_1(x^g(t_1))|\, F_{t+1}^{x^g}]|\, F_t^{x^g}], \\
& \quad \text{by definition of } J(g, t+1), \\
& = E[\sum_{s=t}^{t_1-1} b(s, x^g(s), u^g(s)) + b_1(x^g(t_1))|F_t^{x^g}], \text{ by conditional expectation,} \\
& = J(g, t), \text{ by definition of } J(g, t).
\end{aligned}
$$

Lemma. Value function of a Markov control law

Let $g \in G_M$ be a Markov control law. Define,

$$\begin{aligned} V^g &: T \times X \to \mathbb{R}_+, \\ V^g(t_1, x_V) &= b_1(x_V),\ \forall\ x_V \in X, \\ V^g(t, x_V) &= b(t, x_V, g_t(x_V)) + E[V^g(t+1, f(t, x_V, g_t(x_V), v(t)))|F^{x_V}]., \\ &\forall\ x_V \in X,\ \forall\ t \in T(0 : t_1 - 1). \end{aligned}$$

Then the following equalities hold,

$$V^g(t, x^g(t)) = J(g, t),\ a.s.\ \forall\ t \in T.$$

Proof is similar to that of previous lemma except that equality is achieved everywhere due to existence of Markov control law.

Proof of theorem.(a) - 1

Let V be constructed by the dynamic programming procedure.
Then V satisfies the conditions of V in the Lemma comparison principle, hence,

$$\begin{aligned} & \forall\, g \in G, \\ V(t, x^g(t)) &\leq J(g, t), \text{ a.s. } \forall\, t \in T; \\ E[V(0, x_0)] &\leq E[J(g, 0)] = J(g), \\ & \text{by def. of } J(g, 0) \text{ and of } J(g), \\ E[V(0, x_0)] &\leq \inf_{g \in G} J(g) = J^*, \text{ since } g \in G \text{ was arbitrary.} \end{aligned}$$

Proof of Theorem.(a) - 2

Let $g^* \in G$ be determined by the DP procedure.
Hence it achieves the infima of the DP procedure.
Then V satisfies the conditions for V^{g^*} of second lemma.
From the second lemma then follows that,

$$
\begin{aligned}
V(t, x^{g^*}(t)) &= J(g^*, t) \text{ a.s., } \forall t \in T. \Rightarrow \\
J(g^*) &= E[J(g^*, 0)], \text{ by def. } J(g^*, 0), \\
&= E[V(0, x_0)], \text{ by second lemma} \\
&\leq J^*, \text{ by lemma comparison principle,} \\
&\leq J(g^*), \text{ by definition of } J^*, \Rightarrow \\
J(g^*) &= J^* = \inf_{g \in G} J(g).
\end{aligned}
$$

Hence $g^* \in G$ is an optimal control law.

Necessity proof quite detailed, not suitable for slides.

Outline

Problem. LQG Complete observations - Finite horizon - 1

Gaussian stochastic control system,

$$
\begin{aligned}
x(t+1) &= A(t)x(t) + B(t)u(t) + M(t)v(t),\ x(0) = x_0, \\
z(t) &= C_z(t)x(t) + D_z(t)u(t),\ \ \forall\ t \in T(0:t_1-1), \\
z(t_1) &= C_z(t_1)x(t_1); \\
& T(0:t_1),\ X = \mathbb{R}^{n_x},\ U = \mathbb{R}^{n_u},\ Z = \mathbb{R}^{n_z},\ n_x,\ n_u, n_z \in \mathbb{Z}_+, \\
& x_0 \in G(0, Q_0),\ v(t) \in G(0, I), \\
& \forall\ t \in T(0:t_1-1),\ \mathrm{rank}(D_z(t)) = n_u\ \Rightarrow\ D_z(t)^T D_z(t) \succ 0.
\end{aligned}
$$

Consider past-state information pattern $\{F_t^x,\ t \in T\}$ and corresponding set G of control laws. Closed-loop system with $g \in G$ is,

$$
\begin{aligned}
x^g(t+1) &= A(t)x^g(t) + B(t)g_t(x^g(0:t)) + M(t)v(t),\ x^g(0) = x_0, \\
z^g(t) &= C_z(t)x^g(t) + D_z(t)g_t(x^g(0:t)),\ \forall\ t \in T(0:t_1-1), \\
u^g(t) &= g_t(x^g(0:t)).
\end{aligned}
$$

Problem. LQG Complete observations - Finite horizon - 2

$$
\begin{aligned}
& \inf_{g \in G} J(g), \quad J : G \to \mathbb{R}_+, \\
J(g) &= E \left[\left(\sum_{s=0}^{t_1 - 1} z^g(s)^T z^g(s) \right) + z^g(t_1)^T z^g(t_1) \right], \\
b_1(x_V) &= z_V^T z_V = x_V^T C_z(t_1)^T C_z(t_1) x_V; \\
b(t, x_V, u_V) &= z(t)^T z(t) = \left(\begin{array}{c} x_V \\ u_V \end{array} \right)^T Q_{cr}(t) \left(\begin{array}{c} x_V \\ u_V \end{array} \right) \succeq 0, \\
Q_{cr}(t) &= \left(\begin{array}{cc} C_z^T(t) C_z(t) & C_z(t)^T D_z(t) \\ D_z^T(t) C_z(t) & D_z(t)^T D_z(t) \end{array} \right) \succeq 0, \\
& C_z(t_1)^T C_z(t_1) \succeq 0, \\
& \text{it is assumed that } D_z^T(t) D_z(t) \succ 0, \ \forall \ t \in T \backslash \{t_1\}.
\end{aligned}
$$

Assumption. Controllability and observability

The time-varying Gaussian stochastic control system is on the interval:
supportable,
stochastically controllable, and stochastically observable from z,

$$
\begin{aligned}
n_x &= \mathrm{rank}(\sum_{s=0}^{t_1-1} \Phi(s,0)M(s)M(s)^T\Phi(s,0)^T); \\
n_x &= \mathrm{rank}(\sum_{s=0}^{t_1-1} \Phi(s,0)B(s)B(s)^T\Phi(s,0)^T); \\
n_x &= \mathrm{rank}(\sum_{s=0}^{t_1-1} \Phi(s,0)^T C_z(s)^T C_z(s)\Phi(s,0)); \\
x(t) &= A(t-1)x(t-1),\ x(0) = x_0, \\
x(t) &= \Phi(t,0)x_0, \\
\Phi(t,0) &= A(t-1)A(t-2)\ldots A(1)A(0).
\end{aligned}
$$

Optimal Control Laws

Def. LQG optimal control law (1)

Define the backward control Riccati recursion by the formulas,

$$
\begin{aligned}
Q_c(t_1) &= C_z^T(t_1)C_z(t_1), \;\; Q_c : T \rightarrow \mathbb{R}_{pds}^{n_x \times n_x}, \\
Q_c(t) &= A(t)^T Q_c(t+1)A(t) + C_z^T(t)C_z(t) + \\
&\quad - [A(t)^T Q_c(t+1)B(t) + C_z^T(t)D_z(t)] \times \\
&\quad \times [B(t)^T Q_c(t+1)B(t) + D_z^T(t)D_z(t)]^{-1} \times \\
&\quad \times [A(t)^T Q_c(t+1)B(t) + C_z^T(t)D_z(t)]^T, \\
&\forall\, t \in T(0 : t_1 - 1); \\
&\text{then, } \forall\, t \in T(0 : t_1 - 1), \\
&Q_c(t) \in \mathbb{R}_{pds}^{n_x \times n_x}, \\
&[B(t)^T Q_c(t+1)B(t) + D_z^T(t)D_z(t)] \succeq D_z(t)^T D_z(t) \succ 0.
\end{aligned}
$$

In literature mostly restricted to the case $C_z^T(t)D_z(t) = 0$ for all times.

Def. LQG optimal control law (2)

Define the LQG optimal control law by the formula,

$$
\begin{aligned}
g^*_{LQG,co,fh}(t, x_V) &= F(t, Q_c(t+1))\, x_V, \\
&\quad g^*_{LQG,co,fh} : T \times X \to U, \text{ where,} \\
F(t, Q_c(t+1)) &= -[B(t)^T Q_c(t+1) B(t) + D_z^T(t) D_z(t)]^{-1} \times \\
&\quad \times [A(t)^T Q_c(t+1) B(t) + C_z^T(t) D_z(t)]^T, \\
&\quad F : T \times \mathbb{R}^{n_x \times n_x}_{pds} \to \mathbb{R}^{n_u \times n_x}.
\end{aligned}
$$

Note that $g^* = g^*_{LQG,co,fh} \in G_M$ is a Markov control law.

Theorem. LQG Complete observations - Finite horizon - 1

(a)

- The control law $g^*_{LQG,co,fh}$ is the optimal control law of the LQG optimal stochastic control problem with complete observations on a finite horizon.
- Call g^* the optimal control law LQG-CO-FH, call Q_c the solution of the backward control Riccati recursion.
- The optimal control law $g^*(t, x)$ is linear in the state x for all times $t \in T$.
- The linear control law g^* is optimal in the set of nonlinear Borel measurable control laws. (This generalizes the theorem with respect to optimality with respect to only linear control laws.)

Theorem. LQG Complete observations - Finite horizon - 2

(b) The value function and the optimal cost are given by,

$$
\begin{aligned}
V(t, x_V) &= x_V^T Q_c(t) x_V + r(t), \quad r : T \to \mathbb{R}_+, \\
r(t_1) &= 0, \\
r(t) &= r(t+1) + \mathrm{tr}\left(M(t)^T Q_c(t+1) M(t)\right) \\
&= \sum_{s=t}^{t_1-1} \mathrm{tr}(M(s)^T Q_c(s+1) M(s)); \\
J^*_{LQG,co,fh} &= J(g^*_{LQG,co,fh}) \\
&= E[V(0, x_0)] = E[x_0^T Q_c(0) x_0] + r(0) \\
&= \mathrm{tr}(Q_c(0)\ Q_{x_0}) + r(0).
\end{aligned}
$$

(c)
- $g^*_{LQG,co,fh}$ does depend on the functions $(A,\ B)$ and (C_z, D_z) but does not depend on the function M.
- However, value $J^*_{LQG,co,fh}$ depends also on the function M.

Lemma. Optimization of a quadratic function (1)

$$X = \mathbb{R}^{n_x},\ U = \mathbb{R}^{n_u},$$

$$h(x,u) = \begin{bmatrix} x \\ u \end{bmatrix}^T Q \begin{bmatrix} x \\ u \end{bmatrix},\ h : X \times U \to \mathbb{R}_+,$$

$$\inf_{u \in U} h(x,u),\ \forall\, x \in X,$$

$$Q = \begin{bmatrix} Q_{xx} & Q_{xu} \\ Q_{xu}^T & Q_{uu} \end{bmatrix} \in \mathbb{R}_{pds}^{(n_x+n_u)\times(n_x+n_u)},\ 0 \prec Q_{uu} \in \mathbb{R}_{spds}^{n_u \times n_u};$$

$$0 \prec Q_{uu} \ \Rightarrow\ \forall\, x \in X,\ \lim_{\|u\| \to \infty} h(x,u) = +\infty;$$

$$\frac{\partial h(x,u)}{\partial u} = 2u^T Q_{uu} + 2x^T Q_{xu},$$

$$\frac{\partial^2 h(x,u)}{\partial u^2} = 2Q_{uu} \succ 0.$$

Thus $h(x,u)$ is, for all $x \in X$, a continuous and strictly-convex function in u.

Lemma. Optimization of a quadratic function (2)

$$
\begin{aligned}
0 &= \frac{\partial h(x,u)}{\partial u} = 2u^T Q_{uu} + 2x^T Q_{xu} \ \Rightarrow \ 0 = Q_{uu} u + Q_{xu}^T x, \\
u^* &= -Q_{uu}^{-1} Q_{xu}^T x; \ \ \text{alternatively} \\
h(x,u) &= (u-u^*)^T Q_{uu} (u-u^*) + x^T (Q_{xx} - Q_{xu} Q_{uu}^{-1} Q_{xu}^T) x, \\
& \inf_{u \in U} \ h(x,u) \\
&= \inf_{u \in U} \ \left\{ (u-u^*)^T Q_{uu} (u-u^*) + x^T (Q_{xx} - Q_{xu} Q_{uu}^{-1} Q_{xu}^T) x \right\} \\
&= x^T (Q_{xx} - Q_{xu} Q_{uu}^{-1} Q_{xu}^T) x = h(x, \ u^*).
\end{aligned}
$$

The minimizer is u^* as specified above.

Proof of LQG optimal control law - 1

The dynamic programming procedure is applied.

$$
\begin{aligned}
V(t_1, x_V) &= b_1(x_V) = x_V^T C_z(t_1)^T C_z(t_1) x_V, \\
Q_c(t_1) &= C_z(t_1)^T C_z(t_1), \\
r(t_1) &= 0, \text{ then,} \\
V(t_1, x_V) &= x_V^T Q_c(t_1) x_V + r(t_1), \;\; Q_c(t_1) = Q_c(t_1)^T \succeq 0.
\end{aligned}
$$

By induction. Suppose that for $s = t+1, \ldots, t_1$,

$$
V(s, x_V) = x_V^T Q_c(s) x_V + r(s), \;\; Q_c(s) = Q_c(s)^T \succeq 0.
$$

It will be shown that then this also holds for $s = t$.
The DP procedure prescribes to solve,

$$
\inf_{u_V \in U(t, x_V)} \left\{ z^g(t)^T z^g(t) + E[V(t+1, f(t, x_V, u_V, v(t))) | F^{x_V, u_V}] \right\}.
$$

Proof of LQG optimal control law - 2

$$
\begin{aligned}
& E[V(t+1, f(t, x_V, u_V, v(t)))|F^{x_V,u_V}] \\
& = E[V(t+1, x(t+1))|F^{x_V,u_V}] \\
& = E[x(t+1)^T Q_c(t+1)x(t+1) + r(t+1)|\ F^{x_V,u_V}] \\
& \quad \text{by the induction hypothesis for } V(t+1,\ x_v) \text{ at } s = t+1, \\
& = E[\big(A(t)x_V + B(t)u_V + M(t)v(t)\big)^T Q_c(t+1)\times \\
& \quad \times(A(t)x_V + B(t)u_V + M(t)v(t)) + r(t+1) \mid F^{x_V,u_V}], \\
& \quad \text{by the system recursion,} \\
& = \begin{bmatrix} x_V \\ u_V \end{bmatrix}^T \begin{bmatrix} A(t)^T Q_c(t+1)A(t) & A(t)^T Q_c(t+1)B(t) \\ B(t)^T Q_c(t+1)A(t) & B(t)^T Q_c(t+1)B(t) \end{bmatrix} \begin{bmatrix} x_V \\ u_V \end{bmatrix} \\
& \quad + \operatorname{tr}\left(Q_c(t+1)M(t)M(t)^T\right) + r(t+1), \\
& \quad \text{because } E[v(t)] = 0,\ E[v(t)v(t)^T] = I.
\end{aligned}
$$

Proof of LQG optimal control law - 3

Define,

$$
\begin{aligned}
H_{11}(t) &= A(t)^T Q_c(t+1)A(t) + C_z^T(t)C_z(t),\\
H_{12}(t) &= A(t)^T Q_c(t+1)B(t) + C_z^T(t)D_z(t),\\
H_{22}(t) &= B(t)^T Q_c(t+1)B(t) + D_z^T(t)D_z(t),\\
H(t) &= \begin{bmatrix} H_{11}(t) & H_{12}(t) \\ H_{12}(t)^T & H_{22}(t) \end{bmatrix};\\
& \quad H(t) = H(t)^T \succeq 0,\ Q_c(t+1) \succeq 0,\ \Rightarrow\\
H_{22}(t) &= H_{22}(t)^T = B(t)^T Q_c(t+1)B(t) + D_z(t)^T D_z(t)\\
&\succeq D_z(t)^T D_z(t) \succ 0.
\end{aligned}
$$

by the induction hypothesis and the assumption on D_z.

Proof of LQG optimal control law - 4

$$
\begin{aligned}
& V(t, x_V) \\
&= \inf_{u_V \in U(t,x_V)} \left\{ z^g(t)^T z^g(t) + E[V(t+1, f(t, x_V, u_V, v(t)))|F^{x_V}] \right\} \\
&= \inf_{u_V \in U(t,x_V)} \left\{ \begin{array}{l} \begin{bmatrix} x_V \\ u_V \end{bmatrix}^T H(t) \begin{bmatrix} x_V \\ u_V \end{bmatrix} + \\ +\mathrm{tr}\left(Q_c(t+1)M(t)M(t)^T\right) + r(t+1) \end{array} \right\} \\
&= \inf_{u_V \in U(t,x_V)} \left\{ \begin{array}{l} \begin{bmatrix} x_V \\ u_V + H_{22}(t)^{-1}H_{12}(t)^T x_V \end{bmatrix}^T \times \\ \times \begin{bmatrix} H_{11}(t) - H_{12}(t)H_{22}^{-1}(t)H_{12}^T(t) & 0 \\ 0 & H_{22}(t) \end{bmatrix} \times \\ \times \begin{bmatrix} x_V \\ u_V + H_{22}(t)^{-1}H_{12}(t)^T x_V \end{bmatrix} + \\ +\mathrm{tr}\left((Q_c(t+1)M(t)M(t)^T\right) + r(t+1) \end{array} \right\},
\end{aligned}
$$

Proof of LQG optimal control law - 5

$$
\begin{aligned}
&= x_V^T [H_{11}(t) - H_{12}(t) H_{22}(t)^{-1} H_{12}(t)^T] x_V + \\
&\quad + \operatorname{tr}\left(Q_c(t+1) M(t) M(t)^T\right) + r(t+1), \\
&= x_V^T Q_c(t) x_V + r(t), \text{ if} \\
u_V^* &= -H_{22}^{-1}(t) H_{12}(t)^T x_V, \\
Q_c(t) &= H_{11}(t) - H_{12}(t) H_{22}(t)^{-1} H_{12}(t)^T, \\
r(t) &= r(t+1) + \operatorname{tr}\left(Q_c(t+1) M(t) M(t)^T\right).
\end{aligned}
$$

Proof of LQG optimal control law - 6

Define,

$$
\begin{aligned}
g^*_{LQG,co,fh} &: T \times X \rightarrow U, \\
g^*_t(x_V) = u^* &= -H_{22}(t)^{-1} H_{12}(t)^T x_V \\
&= -[B(t)^T Q_c(t+1) B(t) + D_z(t)^T D_z(t)]^{-1} \times \\
&\quad \times [A(t)^T Q_c(t+1) B(t) + C_z(t)^T D_z(t)]^T x_V.
\end{aligned}
$$

- $H_{22}(t) \succ 0$ for all $t \in T$,
  implies that the value of the optimization $u^* = g^*_t(x_V)$ is unique,
  hence the control law $g^*_{LQG,co,fh}$ is unique.
- The control law $g^*_t(.)$ for all $t \in T(0 : t_1 - 1)$ is a linear function
  hence is a Borel measurable function.
- The sufficient condition of the theorem
  implies that g^* is the optimal control law.

Problem. Gambling with a logarithmic reward

$$x : \Omega \times T \to \mathbb{R}_+, \; u : \Omega \times T \to \mathbb{R}_+, \; v : \Omega \times T \to \{0, \; 1\},$$

$$x(t) = \text{capital of a gambler at time } t \in T,$$

$$u(t) = \text{bid of the gambler at time } t \in T \text{ is } \; u(t) \in (0, \; x(t)),$$

$$v(t) = \begin{cases} 1, & \text{if bid a success,} \\ 0, & \text{if bid not a success,} \end{cases}$$

$$p = P(\{v(t) = 1\}),$$

$$1 - p = P(\{v(t) = 0\}), \; p \in \left(\frac{1}{2}, \; 1\right) \; \text{ (which is not realistic),}$$

$$\begin{aligned} x(t+1) = {} & [x(t) - u(t) + 2u(t)] \; I_{\{v(t)=1\}} + \\ & + [x(t) - u(t)] \; I_{\{v(t)=0\}}, \; x(0) = x_0 \in (0, \; \infty), \end{aligned}$$

$$J(g) = E[\ln(x^g(t_1))], \; \text{at terminal time only,}$$

$$\sup_{g \in G} \; J(g).$$

Proposition. Gambling with a logarithmic reward

The value function and the optimal control law are,

$$\begin{aligned} V(t,\ x_V) &= \ln(x_V) + (t_1 - t)\ c(p), \\ c(p) &= p \ln(p) + (1-p) \ln(1-p) + \ln(2), \\ g^*(x_V) &= 2\left(p - \frac{1}{2}\right)\ x_V. \end{aligned}$$

- T.M. Cover (1984) formulated and solved this problem.
- The optimal control law is a linear function of the state.
- Note that the value function for every time has the same analytic form in x_V, $V(t,\ x_V) = \ln(x_V) + (t_1 - t)c(p)$. The set of logarithmic value functions is invariant with respect to the dynamic programming operator.
- See Problem 12.9.12 of the book.

Proof (1)

$$
\begin{aligned}
V(t_1,\ x_V) &= \ln(x_V), \\
&\quad \text{suppose that for } s = t,\ t+1,\ \ldots,\ t_1, \\
V(s,\ x_V) &= \ln(x_V) + (t_1 - s)\ c(p), \\
c(p) &= p \ln(p) + (1-p) \ln(1-p) + \ln(2), \\
V(t-1,\ x_V) &= \sup_{u_v \in U}\ E[V(t, f(x_V,\ u_V,\ v(t)))|\ F^{x_V,\ u_V}] \\
&= \sup_{u_v \in U}\ E[\ln(x_V - u_V + 2u_V)\ I_{\{v(t)=1\}} + \\
&\quad + \ln(x_V - u_V)\ I_{\{v(t)=0\}} + (t_1 - t)c(p)|\ F^{x_V, u_V}] \\
&= \sup_{u_v \in U}\ p \ln(x_V + u_V) + (1-p) \ln(x_V - u_V) + (t_1 - t)\ c(p) \\
&= \sup_{u_v \in U}\ H(x_V, u_V).
\end{aligned}
$$

Proof (2)

$$\frac{dH(x_V,\ u_V)}{du_V} = \frac{p}{x_V + u_V} - \frac{1-p}{x_V - u_V},$$

$$\frac{d^2H(x_V,\ u_V)}{du_V^2} = -\frac{p}{(x_V + u_V)^2} - \frac{1-p}{(x_V - u_V)^2} < 0,$$

$$H(x_V,\ .) : U \to \mathbb{R} \text{ is strictly concave, } \forall\ x_V \in (0,\ +\infty),$$

$$0 = \frac{dH(x_V,\ u_V)}{du_V} = \frac{p}{x_V + u_V} - \frac{1-p}{x_V - u_V},$$

$$g^*(x_V) = u_V^* = 2\left(p - \frac{1}{2}\right)\ x_V,$$

$$p \in \left(\frac{1}{2},\ 1\right)\ \Rightarrow\ u_V^* \in (0,\ 1)x_V = (0,\ x_V),$$

$$\begin{aligned} H(x_V,\ g^*(x_V)) &= \ln(x_V) + c(p) + (t_1 - t))\ c(p) \\ &= \ln(x_V) + (t_1 - (t-1))\ c(p). \end{aligned}$$

The result follows by induction.

Optimal Control Laws

Special cases of optimal stochastic control problems with explicit optimal control laws?

By explicit one means: either a computational procedure or existence of an analytic formula for V and for g^*.

1. LQG problem with complete observations on a finite horizon.
2. Finite stochastic control systems.
   See Section 12.8 and Homework Set 8.
3. The discrete-time portfolio selection problem on a finite horizon.
   No explicit control law, a numerical computation is required.
4. A gambling problem. See Example 12.9.12.
5. LEQG. A Gaussian stochastic control system
   with as cost function the expectation of an exponential function
   with in the exponent a sum of quadratic forms in x and u.
   Section 12.11.

Outline

Problem. Invariant subsets of value functions

Does there exist a subset of functions on the state set

$$
\begin{aligned}
& V_{inv} \subseteq \{V : X \to \mathbb{R}\} \text{ such that,} \\
& (1) \quad V(t_1, .) \in V_{inv} \text{ and} \\
& (2) \quad \forall\, t \in T \backslash \{0\}, \\
& \qquad V(t, .) \in V_{inv} \;\Rightarrow\; V(t-1, .) \in V_{inv}?
\end{aligned}
$$

- The analytic form of V_{inv}
  yields the analytic form of the optimal control law.
- The smaller the subset V_{inv}, the better for control theory.

Example 1. LQG problem

$$V(t, x) = x^T Q_c(t) x + r(t).$$

Invariant subset of value functions consists of
functions which are quadratic in x with in addition a time function,
parametrized by the functions Q_c and r.

Example 2. A gambling problem with logarithmic award

$$V(t,\ x) = \ln(x) + (t_1 - t)c(p).$$

Invariant subset of value functions consists of
logarithmic functions in x with in addition a time function.

Example 3. A finite stochastic control system

A subset of value functions satisfying a particular condition.
See Exercise 2 of Homework Set 8.

Def. Special case of a Gaussian stochastic control system

$$
\begin{aligned}
x(t+1) &= \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + Bu(t) + Mv(t),\ x(0) = x_0, \\
z(t) &= \begin{bmatrix} C_{z,1} & 0 \end{bmatrix} x(t) + D_z u(t),\ \ \forall\ t \in T \backslash \{t_1\}, \\
z(t_1) &= \begin{bmatrix} C_{z,1} & 0 \end{bmatrix} x(t), \\
& (A_{11},\ C_{z,1}) \text{ observable pair.}
\end{aligned}
$$

Realization theory of linear systems
yields transformation to the above form,
for observability of the state from the controlled output z.

Theorem. Dependence of an optimal control law on the state

Consider the special case of a Gaussian stochastic control system.
If an optimal control law and a value function exist then,

$$g^*(t,\ (x_1,\ x_2)) = g^*(t,\ (x_1,\ 0)),\ \ \forall\,(x_1,\ x_2) \in X,\ \forall t \in T\backslash\{t_1\},$$
$$V(t,\ (x_1,\ x_2)) = V(t,\ (x_1,\ 0)),\ \ \forall\,(x_1,\ x_2) \in X,\ \forall t \in T.$$

- Optimal control and the value function depend only on the state component which is observable by the controlled output.
- System theory of optimal control.
- Theorem is a form of invariance of the set of value functions.
- See Theorem 12.10.1 and Corollary 12.10.2 of the book.

Comments. Perspectives

- How to find the analytic form of the value function from the equation?
- Few formulas known. Example $V(t,\ x) = x^T Q(t)x + r(t)$.
- Numerical approximation of value function.
  Often used in mathematical finance.
- System theoretic approach.
  Use the concepts
  of a Hamiltonian system and of a port-Hamiltonian system.
  Then one starts with an analytic formula for $h(x,\ u)$.
  See literature.
- Not further discussed in this lecture.

Outline

Remarks on dynamic programming

- R. Bellman has popularized dynamic programming,
  see his books published in 1957 and 1962.
- Dynamic programming is a very useful technique.
- Determination of an analytic value function
  depends on the particular case considered,
  depends on the stochastic control system and on the cost function.
- The invariance of a value function is related to
  the existence of a Lyapunov function for stability.
- Course participants best learn to apply dynamic programming,
  also for the case of a finite stochastic system.
- Course participants best learn
  the proof and
  how to use dynamic programming for optimal control problems.

Remarks about the dynamic programming procedure

- Sufficient conditions for
  existence of a minimizer $u_V^* \in U(t, x_V)$.
  See Section 12.4.
  Optimization theory, see Section 17.7.
- Condition on measurability of an optimal control law.
  Formulated by S.E. Shreve.
  Sufficient conditions in Section 12.5.
  In examples to be checked.
- Approach of C. Striebel (1976 book) using P-essential infimum.
  See Section 16.2 of the book.
- Condition of stochastic controllability needed.
  Condition not always stated in the literature.
  For a stochastic control problem on a finite horizon,
  this issue is related to the question:
  Is $J(g^*) < J(g_z)$? Where $g_z = 0$ for all states and for all times.

Overview of Lecture 8

- Dynamic programming procedure.
  Based on a backward time recursion.
- Theorem
  (a) relation of the value function and the infimal cost;
  (b) relation of the value function and the cost-to-go.
- Proof of theorem.
- Solution of LQG optimal stochastic control problem
  in terms of the backward control Riccati recursion and
  the LQG-CO-FH optimal control law.
- Solution of a gambling problem with a logarithmic reward.
- Concept of invariance of a subset of value functions.
- Dependence of the value function and of the optimal control law
  on the observable part of the state via the controlled output.