3 Markov Decision Processes with General State/Action Spaces

In this lecture, we consider Markov decision processes with uncountable state and action spaces. Before we proceed, a brief review on measure-theoretic probability is necessary.

3.1 Review on Probability and Measure Theory

The basic object in Kolmogorov's measure-theoretic formulation of probability theory is the *probability space* $(\Omega, \mathcal{F}, \mathbf{P})$, where Ω is a set called the *sample space*, $\mathcal{F}$ is a collection of subsets of Ω called a *σ-field*, and $\mathbf{P}$ is a function from $\mathcal{F}$ into $[0,1]$ called a *probability measure*.

The definition of *σ-algebra* or *σ-field* is given as follows.

Definition 3.1 (σ-algebra) *A $\sigma-$algebra is a non-empty collection $\mathcal{F}$ of subsets of Ω, i.e. $\mathcal{F} \subset 2^{\Omega}$, satisfying the following axioms:*
A1. $\Omega \in \mathcal{F}$.
A2. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
A3. If $A_1, A_2, \cdots$ is a countable sequence of elements in $\mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

From A1 and A2, we can conclude that $\varnothing \in \mathcal{F}$. From A2, A3 and De Morgan's laws, it is not difficult to see that for a sequence of elements $A_1, A_2, \cdots \in \mathcal{F}$, we also have $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$. Then we give the definition of *probability measure*.

Definition 3.2 (Probability measure) *A probability measure on $(\Omega, \mathcal{F})$ is a function $\mathbf{P} : \mathcal{F} \to [0,1]$ that satisfies the following axioms:*
A1. $\mathbf{P}(A) \geq 0$ for all $A \in \mathcal{F}$.
A2. If $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{F}$, then $\mathbf{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$.
A3. $\mathbf{P}(\Omega) = 1$.

From the preceding definition of probability space, we can also see that i) $\mathbf{P}(\varnothing) = 0$, ii) $\mathbf{P}(A^c) = 1 - \mathbb{P}(A)$, and iii) for an arbitrary sequence of sets $\{A_i\}_{i=1}^{\infty}$ in $\mathcal{F}$, $\mathbf{P}(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbf{P}(A_i)$. Interested readers can refer to [1] for the formal definitions of *algebra* and *measure*.

The construction of probability measures on a finite or a countable set is straightforward. It will be more interesting to investigate probability spaces based on $\Omega = \mathbb{R}$ or $\Omega = \mathbb{R}^n$. Before we take a closer look at these examples, the definition of *Borel σ-algebra* is given in an abstract sense.

Definition 3.3 (Borel σ-algebra) *Let Ω be a topological space. The Borel σ-algebra $\mathcal{B}(\Omega)$ is the smallest σ-algebra containing all open sets of Ω. Any element of $\mathcal{B}(\Omega)$ is called a Borel set.*

Therefore, a Borel set is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersections, and relative complements. For a collection of sets $\mathcal{A} \subset 2^{\Omega}$, we denote by $\sigma(\mathcal{A})$ to indicate the smallest σ-field of Ω that contains $\mathcal{A}$. Thus, denoting by $\mathcal{O}$ the collection of all open sets of Ω, we see that $\mathcal{B}(\Omega) = \sigma(\mathcal{O})$. Now we give the example of the Borel probability space based on the real line $\mathbb{R}$.

Example 3.1 (Probability on $\mathbb{R}$) We take the sample space $\Omega = \mathbb{R}$. Let $\mathcal{A}_0 = \{(a, b) : a < b, a, b \in \mathbb{R}\}$, the collection of all open intervals. The σ-algebra $\sigma(\mathcal{A}_0)$, denoted by $\mathcal{B}(\mathbb{R})$, is the smallest σ-algebra that contains all open intervals of $\mathbb{R}$. Any probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called a Borel probability measure. The same construction applies if Ω is any Borel subset of $\mathbb{R}$, such as, for example, the unit interval $[0, 1]$. We can define a probability measure λ on $[0, 1]$ by setting $\lambda((a, b)) = b - a$ for any open interval (a, b) with $0 \leq a < b \leq 1$. ■

Next, we need to define product spaces.

Definition 3.4 (Product σ-algebra) *Let $\{\Omega_i\}_{i\in\mathcal{I}}$ be a finite indexed collection of nonempty sets, $\Omega = \prod_{i\in\mathcal{I}} \Omega_i$, and $\pi_i : \Omega \to \Omega_i$ the coordinate maps: $\pi_i((\omega_j : j \in \mathcal{I})) = \omega_i$.If $\mathcal{F}_i$ is a σ-algebra on Ω_i for each i, the product σ-algebra on Ω is the σ-algebra generated by*

$$\{\pi_i^{-1}(A_i) : A_i \in \mathcal{F}_i, i \in \mathcal{I}\}.$$

We denote this σ-algebra by $\otimes_{i\in\mathcal{I}}\mathcal{F}_i$.

A *rectangle* in the product space Ω any set of the form $\prod_{iin\mathcal{I}} A_i$ with $A_i \in \mathcal{F}_i$. For the topological spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$, the product spaces $\Omega \times \Omega' = \{(\omega, \omega') : \omega \in \Omega, \omega' \in \Omega'\}$, and the rectangle $\text{Rect}(\mathcal{F} \times \mathcal{F}') = \{A \times A' : A \in \mathcal{F}, A' \in \mathcal{F}'\}$. The following example show a case when the sample spaces are real numbers $\mathbb{R}$.

Example 3.2 ($\mathbb{R}^2$ as a product space) Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two copies of $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The product space is $(\Omega\times\Omega', \mathcal{F}\otimes\mathcal{F}') = (\mathbb{R}^2, \mathcal{B}(\mathbb{R})\otimes\mathcal{B}(\mathbb{R}))$, and it can be shown that $\mathcal{B}(\mathbb{R})\otimes\mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$, the Borel-$\sigma$-algebra generated by all open subsets of $\mathbb{R}^2$.

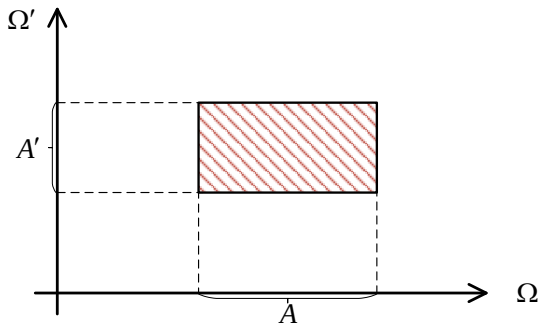

Figure 1: Product spaces and rectangle.

Lastly, we go over the definitions of measurable function and random variables. The definition of measurable function is given as follows.

Definition 3.5 (Measurable function) *If $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ are measurable spaces, a mapping $f : \Omega \to \Omega'$ is called $(\mathcal{F}, \mathcal{F}')-$measurable, or just measurable if*

$$f^{-1}(A) = \{x \in \Omega : f(x) \in A\} \in \mathcal{F}$$

for all $A \in \mathcal{F}'$.

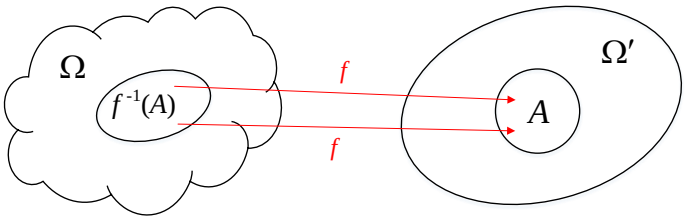

Figure 2: Measurable function.

For any $A, B \subset \Omega'$, we have the properties $f^{-1}(A^c) = (f^{-1}(A))^c$ and $f^{-1}(A \cup B) = g^{-1}(A) \cup g^{-1}(B)$. Random variable is defined as

Definition 3.6 (Random variable) *Let a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ be given. A random variable is a measurable function X from $(\Omega, \mathcal{F})$ to the real line $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, meaning that*

$$\{\omega \in \Omega : X(\omega) \leq c\} \in \mathcal{F}$$

for any $c \in \mathbb{R}$.

3.2 Controlled Markov Processes in general spaces

With the preceding concepts, we now formulate controlled Markov processes with general state/action spaces. Both the state space $\mathcal{X}$ and the action space $\mathcal{U}$ can be Borel spaces, which can be either a Borel subset of $\mathcal{R}^n$ with Borel σ-field or a finite or countable set with the powerset σ-field.

For $t = 0, 1, 2, \cdots$, the state transition kernels are defined as follows:

$$P_t(x, u, A) = \mathbf{P}[X_{t+1} \in A | X_t = x, U_t = u],$$

where $x \in \mathcal{X}$, $u \in \mathcal{U}$, and A is a Borel subset of $\mathcal{X}$. As a function of (x, u), $\mathbb{P}_t(x, u, A)$ is measurable for each fixed A, and $P_t(x, u, \cdot)$ is a Borel probability measure on the state space $\mathcal{X}$ for each fixed state-action pair $(x, u) \in \mathcal{X} \times \mathcal{U}$. In the following context, we will also use the notations $P_t(x, u, \mathrm{d}x')$ that encodes $P_t(x, u, A) = \int_A \mathbb{P}_t(x, u, \mathrm{d}x')$, or even $P_t(\mathrm{d}x'|x, u)$ and write $\mathbf{P}[X_{t+1} \in A | X_t = x, U_t = u] = \int_A P_t(\mathrm{d}x'|x, u)$. It is often the case that the transition kernels admit *probability densities*: for example, if $\mathcal{X} = \mathbb{R}^n$, then there exist nonnegative functions $p_t(x'|x, u)$, such that

$$P_t(x, u, A) = \int_A p_t(x'|x, u)\, \mathrm{d}x',$$

such that $\int_{\mathbb{R}^n} p_t(x'|x, u)\, \mathrm{d}x' = 1$. Before we introduce the policy and cost function, an example on state transition kernels of MDPs with general state/action spaces is discussed.

Example 3.3 (Transition densities with general state/action spaces) Consider a dynamical model described by

$$X_{t+1} = f_t(X_t, U_t) + W_t,$$

where $W_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2 I_n)$, and I_n is the $n \times n$ identity matrix. The transition probability densities satisfy

$$p_t(x'|x,u) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\|x' - f_t(x,u)\|^2\right),$$
$$P_t(A|x,u) = \int_A p_t(x'|x,u)\,\mathrm{d}x'.$$

■

For MDPs with general state/action spaces, a policy $g = (g_t)_{t\geq 0}$ is a sequence of measurable functions $g_t : \mathcal{X}_0^t \times \mathcal{U}_0^{t-1} \to \mathcal{U}$. The stagewise cost functions are also a measurable functions $c_t : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$, as is the terminal cost $c_t : \mathcal{X} \to \mathbb{R}$. The expected cost of g is defined as

$$J_T(g) = \mathbf{E}^g\left[\sum_{t=0}^{T-1} c_t(X_t, U_t) + c_T(X_T)\right].$$

As in the finite case, a Markov policy $g = (g_t)_{t\geq 0}$ is a collection of functions $g_t : \mathcal{X} \to \mathcal{U}$, such that the action at time t is a function of only the state at time t.

Now, Blackwell's principle of irrelevant information is valid for general Borel spaces, therefore exactly the same proof as in the finite case can be used to show that Markov policies are optimal. Accordingly, we can use dynamic programming to find optimal policies: defining the value functions $V_t : \mathcal{X} \to \mathbb{R}$ for $t = 0, 1, 2, \cdots, T$, the Bellman recursion for MDPs with general (Borel) state/action spaces starts with $V_T = c_T$ and, for $t = T-1, T-2, \ldots, 0$,

$$\begin{aligned} V_t(x) &= \inf_{u\in\mathcal{U}} \left\{c_t(x,u) + \mathbf{E}[V_{t+1})(X_{t+1}|X_t = x, U_t = u]\right\} \\ &= \inf_{u\in\mathcal{U}} \left\{c_t(x,u) + \int_{\mathcal{X}} c(x',u) P_t(\mathrm{d}x'|x,u)\right\}, \end{aligned}$$

and the optimal policy $g_t^*(x) = \arg\min_u V_t(x)$. Now we revisit a well-known optimal control problem, the linear quadratic regulator (LQR) problem, in a stochastic setting. We first formulate the stochastic LQ problem and immediately show the key results, and the proof will be derived in next lecture.

Example 3.4 (Stochastic LQR problem) Consider the scalar state and action spaces, $\mathcal{X} = \mathcal{U} = \mathbb{R}$. The dynamical model is described by

$$X_{t+1} = aX_t + bU_t + W_t, \tag{1}$$

where the disturbances $W_0, W_1, \ldots$ are independent with $\mathbf{E}[W_t] = 0$ and $\mathbf{E}[W_t^2] = \sigma^2$. The cost function is defined as $c_t(x,u) = qx^2 + ru^2$ for $t = 0, 1, 2, \cdots, T-1$ and $c_T(x,u) = q_T x^2$, where $q, q_T \geq 0$ and $r > 0$. The objective of this LQR problem is to find the optimal policy g^* minimizing

$$J(g) = \mathbf{E}^g\left[\sum_{t=0}^{T-1}(qX_t^2 + rU_t^2) + q_T X_T^2\right].$$

We will show the following: The optimal policy is linear in x and satisfies

$$g_t^* = G_t x, \tag{2}$$

where G_t is the gain at time t. The value function is quadratic in x and satisfies

$$V_t(x) = K_t x^2 + \sum_{s=t+1}^{T} K_s \sigma^2. \tag{3}$$

Moreover, G_t and K_t can be computed recursively from $K_T = q_T$, $K_t = q + K_{t+1}a^2[1 - K_{t+1}b^2/(r + K_{t+1}b^2)]$, and $G_t = -K_t ab/(r + K_t b^2)$. The detailed deviations of these results are given in the following subsection. ■

3.3 The scalar Linear Quadratic Regulator problem: the stationary case

The linear quadratic regulator (LQR) is a classic example of a control system with continuous state and action spaces. For now we assume that the dynamics are stationary (time-invariant). We take $\mathcal{X} = \mathcal{U} = \mathbb{R}$ and $X_{t+1} = aX_t + bU_t + W_t$, where $\{W_t\}_{t=0}^{T}$ are independent with $\mathbf{E}[W_t] = 0, \mathbf{E}[W_t^2] = \sigma^2$. The cost of the system at each time $t = 0, 1, \ldots, T-1$ is $c_t(x, u) = c(x, u) = qx^2 + ru^2$ for some $q \geq 0$ and $r > 0$, and the terminal cost is $c_T(x) = q_T x^2, q_T \geq 0$. Intuitively, the qx^2 term in $c(x, u)$ penalizes for the state to deviate from 0 and the ru^2 term penalizes the "effort" required to control the system. The finite-horizon LQR problem is then to minimize the expected cost:

$$\min_g J(g) \tag{4}$$

$$J(g) = \mathbf{E}^g \left[\sum_{t=0}^{T-1} qX_t^2 + rU_t^2 + q_T X_T^2 \right]. \tag{5}$$

In the remainder of this section, we will prove the following theorem:

Theorem 3.1 *The optimal policies and value functions for Eq. (4) at time step $t = 0, \ldots, T-1$ are of the form*

$$g_t^*(x) = G_t x \tag{6}$$

$$V_t(x) = K_t x^2 + \sum_{s=t+1}^{T} K_s \sigma^2 \tag{7}$$

, where G_t, K_t can be computed recursively with the following relations:

$$K_T = q_T \tag{8}$$

$$G_t = -\frac{K_{t+1}ab}{r + K_{t+1}b^2} \tag{9}$$

$$K_t = q + K_{t+1}a^2 \left(1 - \frac{K_{t+1}b^2}{r + K_{t+1}b^2}\right), \tag{10}$$

for $t = 0, \ldots, T-1$.

This theorem tells us that the optimal control policy for the LQR is a linear function of the state and the value function is a quadratic function of the state. In order to prove the theorem, we need to prove a separate lemma, which will be convenient for solving Eq. (4) later.

Lemma 3.1 (Completion of squares) *Consider the function*

$$f(x,u) = c_1u^2 + 2c_2xu$$

where $c_1 > 0$. *Then, for any* x, $\min_{u\in\mathbb{R}} f(x,u)$ *is achieved uniquely at* $u^* = -\frac{c_2}{c_1}x$ *and*

$$\min_{u\in\mathbb{R}} f(x,u) = -\frac{c_2^2x^2}{c_1}.$$

Proof: Expand f as

$$\begin{aligned} f(x,u) &= c_1u^2 + 2c_2xu \\ &= c_1(u^2 + 2\frac{c_2}{c_1}xu + \frac{c_2^2x^2}{c_1^2} - \frac{c_2^2x^2}{c_1^2}) \\ &= c_1(u + \frac{c_2}{c_1}x)^2 - \frac{c_2^2x^2}{c_1} \\ &\geq -\frac{c_2^2x^2}{c_1}. \end{aligned}$$

Equality holds if and only if $u = -\frac{c_2}{c_1}x$, in which case $\min_{u\in\mathbb{R}} f(x,u) = -\frac{c_2^2x^2}{c_1}$. ■

Now we can prove Thm. 3.1 by backward induction:

Proof: [Thm. 3.1] We have $V_T(x) = q_Tx^2$. For $t = T-1$,

$$\begin{aligned} Q_{T-1}(x,u) &= c(x,u) + \mathbf{E}[V_T(X_T)|X_{T-1} = x, U_{T-1} = u] \\ &= qx^2 + ru^2 + \mathbf{E}[q_TX_T^2|X_{T-1} = x, U_{T-1} = u] \\ &= qx^2 + ru^2 + \mathbf{E}[q_T(ax + bu + W_t)^2] \\ &= qx^2 + ru^2 + q_T(ax + bu)^2 + q_T\sigma^2 \\ &= (q + q_Ta^2)x^2 + r + (q_Tb^2)u^2 + 2q_Tabxu + q_T\sigma^2. \end{aligned}$$

Let $k_T := q_T$:

$$\begin{aligned} V_{T-1}(x) &= \min_{u\in\mathbb{R}} Q_t(x,u) \\ &= (q + K_Ta^2)x^2 + K_T\sigma^2 + \min_{u\in\mathbb{R}}\{(r + K_Tb^2)u^2 + 2K_Taxu + K_T\sigma^2\} \end{aligned}$$

Since $r + K_Tb^2 > 0$, we can apply Lemma 3.1 with $c_1 = r + K_Tb^2, c_2 = K_Tab$ to get

$$\begin{aligned} V_{T-1}(x) &\geq (q + K_Ta^2)x^2 + K_T\sigma^2 - \frac{(K_Tab)^2}{r + K_Tb^2}x^2 \\ &= (q + K_Ta^2(1 - \frac{K_Tb^2}{r + K_Tb^2}))x^2 + K_T\sigma^2 \\ &= K_{T-1}x^2 + K_T\sigma^2, \end{aligned}$$

where equality is achieved uniquely by $u^* = G_{T-1}x = -\frac{K_T ab}{r+K_T b^2}x$. Notice that $K_{T-1} = q + K_T a^2(1 - \frac{K_T b^2}{r+K_T b^2}) \geq 0$ since $K_T \geq 0$, therefore the theorem holds for $t = T-1$.

Suppose at time step $t+1$, $V_{t+1}(x) = K_{t+1}x^2 + \sum_{s=t+1}^{T} K_s\sigma^2$ and $g^*_{t+1}(x) = G_{t+1}x$ for K_{t+1}, G_{t+1} defined in 8. Then at time t we have

$$\begin{aligned}
V_t(x) &= \min_{u\in\mathbb{R}}\{qx^2 + ru^2 + \mathbf{E}[V_{t+1}(X_{t+1})|X_t = x, U_t = u]\} \\
&= \min_{u\in\mathbb{R}}\left\{qx^2 + ru^2 + K_{t+1}\mathbf{E}[(ax+bu+W_t)^2] + \sum_{s=t+1}^{T} K_s\sigma^2\right\} \\
&= \min_{u\in\mathbb{R}}\left\{qx^2 + ru^2 + K_{t+1}(ax+bu)^2 + K_{t+1}\mathbf{E}[W_t^2] + \sum_{s=t+1}^{T} K_s\sigma^2\right\} \\
&= (q + K_{t+1}a^2)x^2 + \min_{u\in\mathbb{R}}\{(r + K_{t+1}b^2)u^2 + 2K_{t+1}abxu\} + K_t\sigma^2 + \sum_{s=t+1}^{T} K_s\sigma^2 \\
&= (q + K_{t+1}a^2)x^2 + \min_{u\in\mathbb{R}}\{(r + K_{t+1}b^2)u^2 + 2K_{t+1}abxu\} + \sum_{s=t}^{T} K_s\sigma^2.
\end{aligned}$$

Since $r + K_{t+1}b^2 > 0$ by inductive assumptions, we apply Lemma 3.1)with $c_1 = r + K_{t+1}b^2, c_2 = K_{t+1}ab$:

$$\min_{u\in\mathbb{R}}\{(r + K_{t+1}b^2)u^2 + 2K_{t+1}abxu\} = -\frac{K_{t+1}^2a^2b^2}{r + K_{t+1}b^2}x^2,$$

where equality holds when $g_t^*(x) = G_t x, G_t = -\frac{K_{t+1}ab}{r+K_{t+1}b^2}$. As a result,

$$\begin{aligned}
V_t(x) &= (q + K_{t+1}a^2 - -\frac{K_{t+1}^2a^2b^2}{r + K_{t+1}b^2})x^2 + \sum_{s=t+1}^{T} K_s\sigma^2 \\
&= K_t x^2 + \sum_{s=t+1}^{T} K_s\sigma^2.
\end{aligned}$$

Therefore the inductive hypothesis is true for $s = t$ and Theorem 3.1 is proved. ■

3.4 The vector Linear Quadratic Regulator problem

The LQR can be generalized to vector state and action spaces. Let $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{U} = \mathbb{R}^m$. Let the matrices $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}$ be given. The linear dynamics of the system is given by

$$X_{t+1} = AX_t + BU_t + W_t, \tag{11}$$

where $W_0, W_1, ..., W_T$ is a sequence of independent random vectors in $\mathbb{R}^n$ with $\mathbf{E}[W_t] = 0, \mathbf{E}[W_t W_t^\top] = \Sigma$. The cost function for each time step is defined as:

$$c_T(x,u) = x^\top Q_T x \tag{12}$$

$$c_t(x,u) = x^\top Q x + u^\top R u, \tag{13}$$

for t = 0, ..., T-1, where the $n \times n$ matrices Q, Q_T are symmetric and positive semidefinite, while the $m \times m$ matrix R is positive definite. The vector LQR is to minimize the expected cost:

$$\min_g J_T(g) \tag{14}$$

$$J_T(g) = \mathbf{E}^g \left[\sum_{t=0}^{T-1} X_t^\top Q X_t + U_t^\top R U_t + X_T^\top Q_T X_T \right]. \tag{15}$$

As in the scalar case, the optimal policy can be obtained via dynamic programming, and an analogous theorem holds for the vector case:

Theorem 3.2 *The optimal policies and value functions for Eq. (14) at time $t = 0, \ldots, T-1$ are:*

$$g_t^*(x) = G_t x \tag{16}$$

$$V_t(x) = x^\top K_t x + \sum_{s=t+1}^{T} \mathrm{tr}(K_s \Sigma), \tag{17}$$

where the matrices G_t, K_t can be computed recursively as follows: $K_T = Q_T$ and, for $t = T-1, T-2, \ldots, 0$,

$$G_t = -(R + B^\top K_{t+1} B)^{-1} B^\top K_{t+1} A \tag{18}$$

$$K_t = Q + A^\top (K_{t+1} - K_{t+1} B (R + B^\top K_{t+1} B)^{-1} B^\top K_{t+1}) A \tag{19}$$

Again as in the scalar case, we need the completion-of-squares lemma, which is slightly more involved for the vector case:

Lemma 3.2 (Completion of squares for vectors) *Consider the function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, given by*

$$f(x,u) = u^\top F_1 u + 2 x^\top F_2 u,$$

where the matrix F_1 is positive definite. Then, for any fixed x, the minimum of $f(x,u)$ over $u \in \mathbb{R}^m$ is achieved uniquely by $u^* = -F_1^{-1} F_2^\top x_2$, and

$$\min_{u \in \mathbb{R}^m} f(x,u) = -x^\top F_2 F_1^{-1} F_2^\top x.$$

Proof: Expand f as

$$f(x,u) = u^\top F_1 F_1^{-1} F_1 u + 2 x^\top F_2 F_1^{-1} F_1 u + x^\top F_2 F_1^{-1} F_2^\top x - x^\top F_2 F_2^\top F_1^{-1} x \tag{20}$$

$$= (F_1 u + F_2^\top x)^\top F_1^{-1} (F_1 u + F_2^\top x) - x^\top F_2 F_1^{-1} F_2^\top x \tag{21}$$

$$\geq -x^\top F_2 F_1^{-1} F_2^\top x, \tag{22}$$

where evidently equality is achieved if and only if $u = -F_1^{-1} F_2^\top x$. ■

Proof: [Thm. 3.2] We start the backward induction with $V_T(x) = c_T(x) = x^\top Q_T x$. Set $K_T = Q_T$. Then, for $t = T-1$,

$$V_{T-1}(x) = \min_{u\in\mathbb{R}^m}\{x^\top Qx + u^\top Ru + \mathbf{E}[X_T^\top Q_T X_T]\} \tag{23}$$
$$= \min_{u\in\mathbb{R}^m}\{x^\top Qx + u^\top Ru + \mathbf{E}[(Ax+Bu+W_T)^\top Q_T(Ax+Bu+W_T)]\} \tag{24}$$
$$= x^\top Qx + \min_{u\in\mathbb{R}^m}\{(Ax+Bu)^\top Q_T(Ax+Bu) + \mathbf{E}[W_T^\top Q_T W_T]\} \tag{25}$$
$$= x^\top(Q+A^\top Q_T A)x + \min_u\{u^\top(R+B^\top Q_T B)u + x^\top A^\top K_T Bu + (A^\top K_T Bu)^\top\} + \tag{26}$$
$$\mathrm{tr}(K_T\Sigma)\} \tag{27}$$
$$= x^\top(Q+A^\top Q_T A)x + \min_u\{u^\top(R+B^\top Q_T B)u + 2x^\top A^\top K_T Bu\} + \mathrm{tr}(K_T\Sigma)\} \tag{28}$$
$$= x^\top(Q+A^\top Q_T A)x + L + \mathrm{tr}(K_T\Sigma), \tag{29}$$

where $x^\top A^\top K_T Bu = u^\top B^\top K_T Ax$ by trace theorem, $\mathrm{tr}(x^\top Cu) = \mathrm{tr}(u^\top C^\top x)$. Let $Q_T := K_T$ and apply Lemma. (3.2) on L with $F_1 = R + B^\top K_T B, F_2 = A^\top K_T B$:

$$L \geq -x^\top A^\top K_{t+1}B(R+B^\top K_{t+1}B)^{-1}B^\top K_{t+1}Ax, \tag{30}$$

with equality when $g_{T-1}(X) = G_{T-1}X = (R+B^\top K_{t+1}B)^{-1}B^\top K_{t+1}Ax$. Plug Eq. (30) into Eq. (23),

$$\begin{aligned} V_{T-1}(x) &= x^\top[Q + A^\top(K_T - K_T B(R+B^\top K_T B)^{-1}B^\top K_T)A]x + \mathrm{tr}(K_T\sigma) \\ &= x^\top K_{T-1}x + \mathrm{tr}(K_T\Sigma). \end{aligned}$$

Assume the statements hold for time $s = t+1$. Then at time t:

$$\begin{aligned} V_t(x) &= x^\top Qx + \min_{u\in\mathbb{R}^m}\{(Ax+Bu)^\top K_{t+1}(Ax+Bu)\} + \mathbf{E}[W_{t+1}^\top K_{t+1}W_{t+1}] + \sum_{s=t+2}\mathrm{tr}(K_s\Sigma) \\ &= x^\top(Q+A^\top K_{t+1}A)x + \min_{u\in\mathbb{R}^m}\{u^\top(R+B^\top K_{t+1}B)u + 2x^\top A^\top K_{t+1}Bu\} + \sum_{s=t+1}\mathrm{tr}(K_s\Sigma) \\ &\geq x^\top(Q+A^\top K_{t+1}A)x - x^\top A^\top K_{t+1}B(R+B^\top K_{t+1}B)^{-1}B^\top K_{t+1}Ax + \sum_{s=t+1}\mathrm{tr}(K_s\Sigma) \\ &= x^\top K_t x + \sum_{s=t+1}^{T}\mathrm{tr}(K_s\Sigma), \end{aligned}$$

where the inequality uses Lemma 3.2. Equality is achieved when $u^* = -(R+B^\top K_t B^\top)^{-1}B^\top K_t Ax = G_t x$. Therefore, the theorem holds for $t = 0, ..., T-1$.

■

It can also be shown that the matrices K_t are positive semidefinite, and therefore the value functions V_t are *convex*. This will be explored in a homework problem.

References

[1] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, 1999.