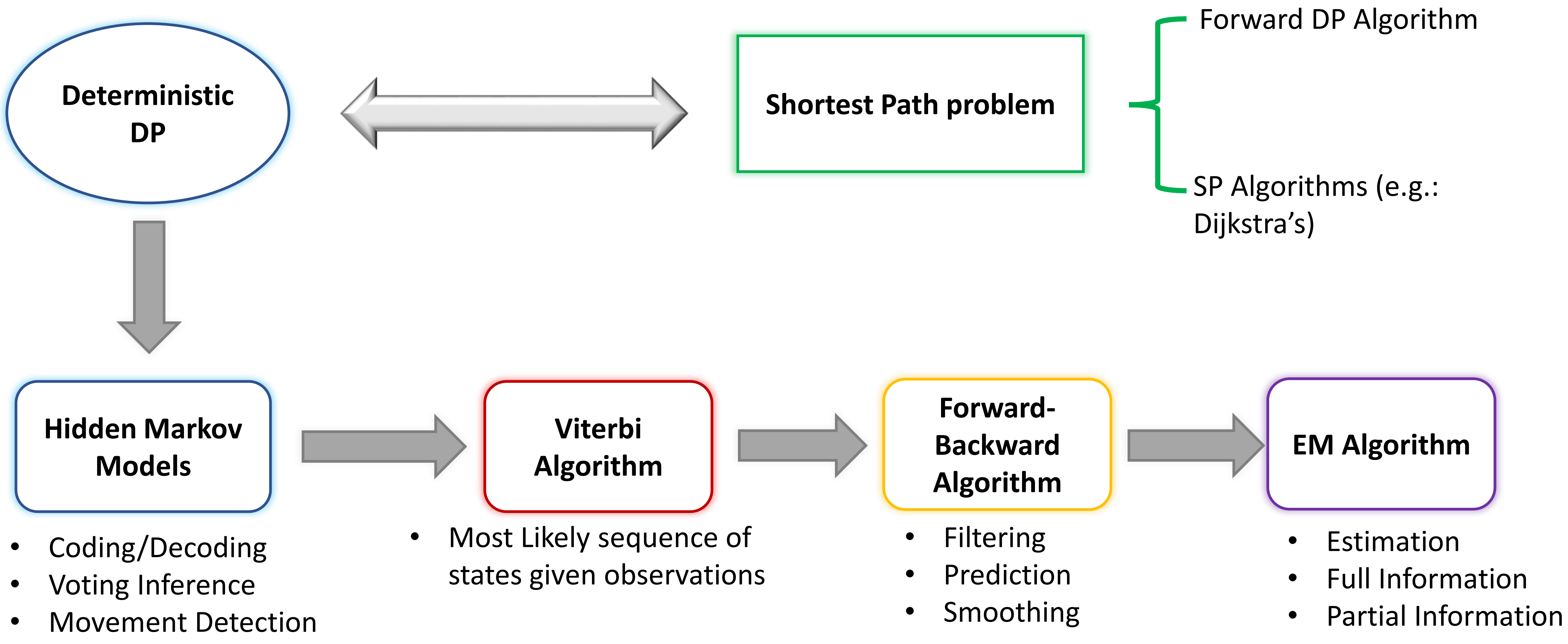
Recap of Deterministic DP and HMM



Stochastic Dynamic Programming

- Recall our DP formulation for problems with disturbances:

$$J^*(x_0) = \min_{\pi \in \Pi} \mathbb{E}_w \left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k) \right]$$

$$x_{k+1} = f_k(x_k, u_k(x_k), w_k), \forall k \in \{0, 1, ..., N-1\}$$

- Differently from the deterministic case, we need to obtain closed-loop policies. Hence our goal is to find the optimal policy:

$$\pi^* = \{\mu_0^*(x_0), ..., \mu_{N-1}^*(x_{N-1})\}$$

Stochastic Dynamic Programming

- Due to the disturbance vectors, forward DP will not work in general and we have to rely again the backward DP recursion:

$$J_N(x_n) = g_N(x_N)$$

$$J_i(x_i) = \min_{u_i \in U_i(x_i)} \left\{ \mathbb{E}_{w_i} \left[g_i(x_i, u_i, w_i) + J_{i+1}(f_i(x_i, u_i, w_i)) \right] \right\}, \forall i \in \{0, ..., N-1\}$$

- The key challenges here can be summarized as three questions:
  - How to compute the expectation w.r.t. to w_i?
  - How to perform the optimization on the right-hand side?
  - How to overcome the fact the above has to be done for **every** possible state?

Example: Optimal Stopping Problem

- Let's consider the problem of selling a house:

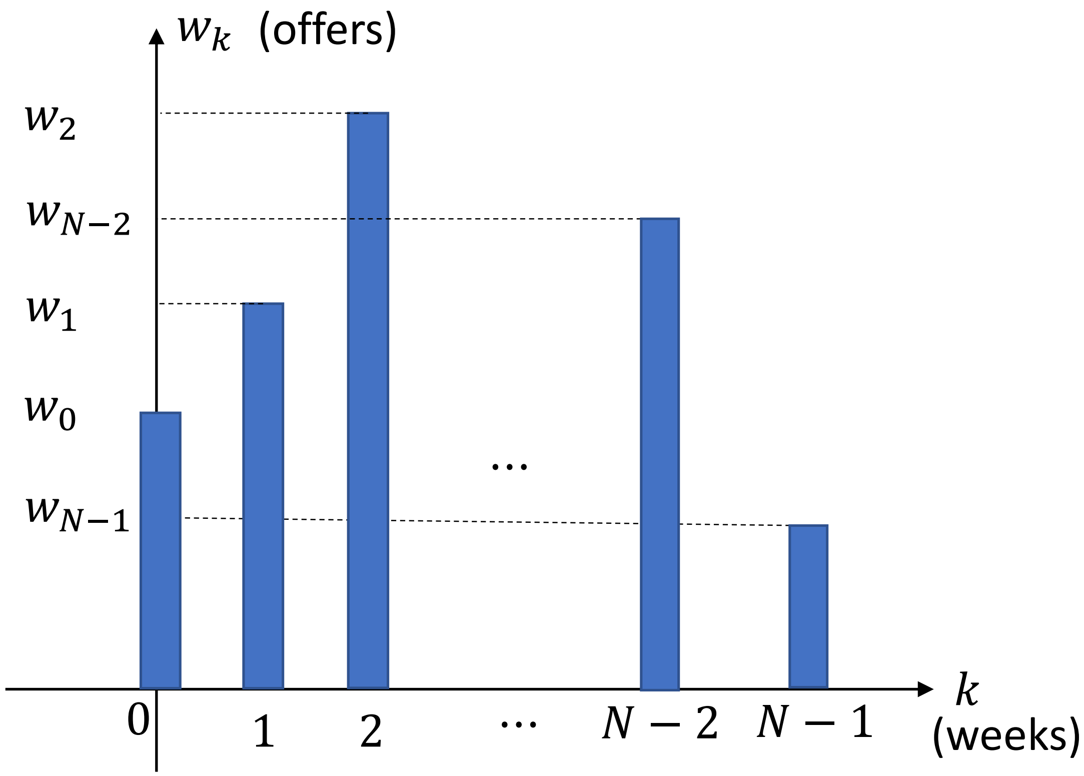
Reject the offer:

We continue to wait for future offers

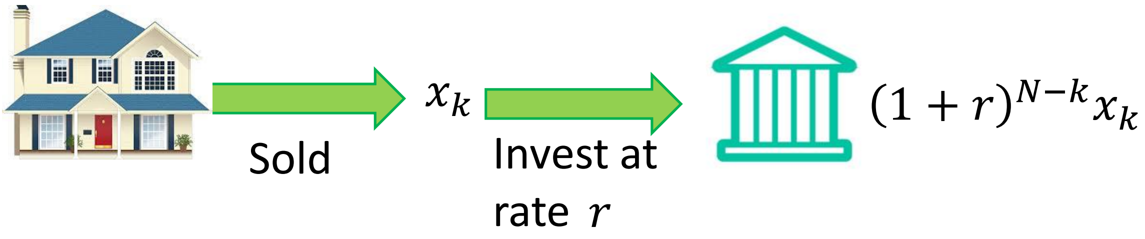
Accept the offer:

We sell the house

Put the money in the bank



If offer is accepted at some period k:



- The question is: What is the best policy (strategy) to sell the house to make most money?

Example: Optimal Stopping Problem

- This problem is an instance of a class of problems called Optimal Stopping Problems. Here the ``stopping” decision is the decision to sell the house.

- Upon selling, the “process” of receiving offers halts and we collect our money at the end of the investment horizon.

- We begin the DP analysis by defining the states and controls:

$$x_k \in \mathbb{R} \cup \{T\} \quad \begin{cases} \text{If } x_k = T\text{, then house has been sold at some time } k \leq N-1 \\ \text{If } x_k \neq T\text{, then house is yet to be sold, and the outstanding offer is of value } x_k \ (x_k = w_{k-1}). \end{cases}$$

$$u_k \in \{\text{“sell”}, \text{“do not sell”}\}$$

Example: Optimal Stopping Problem

- Now we can write out the dynamics:

$$x_{k+1} = f_k(x_k, u_k, w_k), , \forall k \in \{0, ..., N-1\}$$

- where the functions f_k are defined via the relation:

$$x_{k+1} = \begin{cases} T, \text{ if } x_k = T, \text{ or if } x_k \neq T \text{ and } u_k = \text{“sell”} \\ w_k, \text{ otherwise} \end{cases}$$

- And the cost functions:

$$g_N(x_N) = \begin{cases} x_N, \text{ if } x_N \neq T \\ 0, \text{ otherwise} \end{cases}$$

$$g_k(x_k, u_k, w_k) = \begin{cases} (1+r)^{N-k} x_k, \text{ if } x_k \neq T \text{ and } u_k = \text{“sell”} \\ 0, \text{ otherwise} \end{cases}$$

- Note that we enforced the house must be sold at the last period, if all offers were rejected.

Example: Optimal Stopping Problem

- Based on this modelling formulation, we can write out the DP recursion:

$$J_N(x_N) = \begin{cases} x_N, \text{ if } x_N \neq T \\ 0, \text{ if } x_N = T \end{cases}$$

$$J_k(x_k) = \begin{cases} \max\{(1+r)^{N-k} x_k, \mathbb{E}_{w_k}[J_{k+1}(w_k)]\}, \text{ if } x_k \neq T \\ 0, \text{ if } x_k = T \end{cases}$$

- Hence we can write the optimal policy as follows:

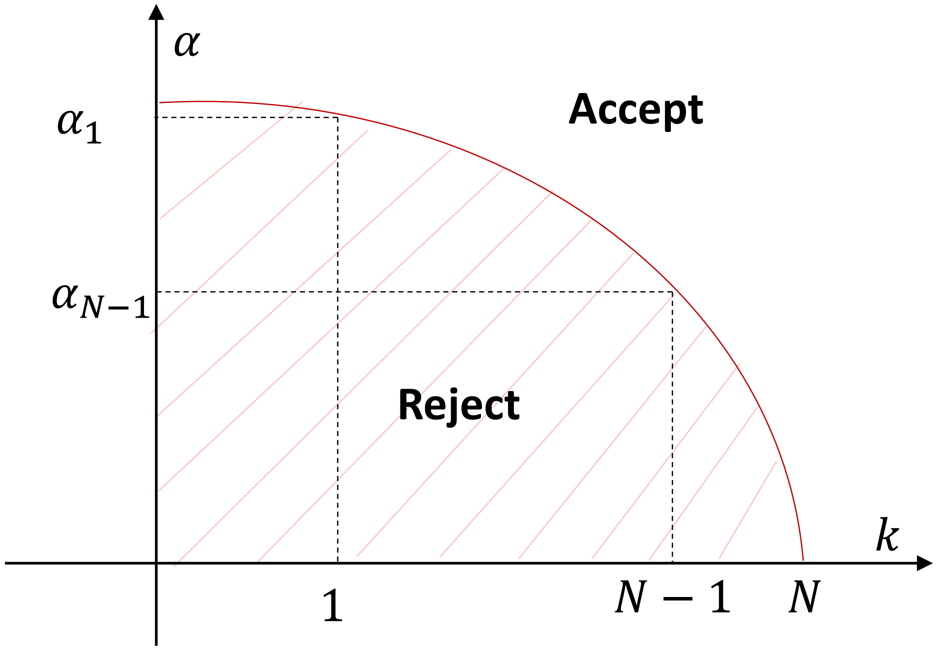
$$\mu^*(x_k) = \begin{cases} \text{“sell”, if } x_k > \alpha_k \\ \text{“do not sell”, if } x_k \leq \alpha_k \end{cases} \qquad \alpha_k = \frac{\mathbb{E}_{w_k}[J_{k+1}(w_k)]}{(1+r)^{N-k}}$$

Example: Optimal Stopping Problem

- If the offers w_k are i.i.d. it can be verified that:

$$\alpha_k \geq \alpha_{k+1}, \forall k \in \{0, \dots, N-1\}$$

- And we can draw the following graph:



- Where acceptance region is called the **stopping set**.

Linear Quadratic Regulator (LQR)

- We will study the classical LQR problem, which is a special type of Stochastic DP, where the dynamics are linear:

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \forall k \in \{0, 1, ..., N-1\}$$

- And the cost is quadratic:

$$\mathbb{E}_w \left[x_N^\top Q_N x_N + \sum_{k=0}^{N-1} (x_k^\top Q_k x_k + u_k^\top R_k u_k) \right]$$

- Where the matrices $Q_k's$ are symmetric p.s.d. and the matrices are $R_k's$ are symmetric p.d. .

Linear Quadratic Regulator (LQR)

- In addition, the disturbances $w_k's$ are independent of random vectors such that:
  - the probability distributions do not depend on x_k and u_k (so $w_k \sim \Pr_k(\cdot)$).

- Furthermore, each w_k has zero-mean and finite second moment:

$$\mathbb{E}[w_k] = 0,\ \mathbb{E}[w_k^2] < \infty,\ \forall k \in \{0, 1, ..., N-1\}$$

- We can consider a variation of LQR, where our goal is to follow some predetermined trajectory $(\bar{x}_0, \dots, \bar{x}_N, \bar{u}_0, \dots, \bar{u}_{N-1})$. And the cost function can be:

$$\mathbb{E}_w\left[(x_N - \bar{x}_N)^\top Q_N (x_N - \bar{x}_N) + \sum_{k=0}^{N-1} ((x_k - \bar{x}_k)^\top Q_k (x_k - \bar{x}_k) + u_k^\top R_k u_k)\right]$$

Linear Quadratic Regulator (LQR)

- As always, let's write the (backwards) DP recursion for the LQR problem:

$$J_N(x_N) = x_N^\top Q_N x_N$$

$$J_k(x_k) = \min_{u_k} \left\{ \mathbb{E}_{w_k} \left[x_k^\top Q_k x_k + u_k^\top R_k u_k + J_{k+1}(A_k x_k + B u_k + w_k) \right] \right\}$$

- To obtain the optimal policy we need to optimize the right-hand side.

- The LQR problem is nice because this optimization can be done efficiently (convex problem) and in closed form.

- This feature makes it the classical and (probably) the most study model for dynamic systems. It forms the "foundation" of optimal control and more complex DP models.

Linear Quadratic Regulator (LQR)

- Let's begin the recursion for $k = N - 1$:

$$J_{N-1}(x_{N-1}) = \min_{u_{N-1}} \Big\{ \mathbb{E}_{w_{N-1}} \big[x_{N-1}^\top Q_{N-1} x_{N-1} + u_{N-1}^\top R_{N-1} u_{N-1} +$$

$$(A_{N-1} x_{N-1} + B u_{N-1} + w_{N-1})^\top Q_N (A_{N-1} x_{N-1} + B u_{N-1} + w_{N-1}) \big] \Big\}$$

- Now expanding the last term:

$$J_{N-1}(x_{N-1}) = x_{N-1}^\top Q_{N-1} x_{N-1} + \min_{u_{N-1}} \Big\{ u_{N-1}^\top R_{N-1} u_{N-1} +$$

$$u_{N-1}^\top B_{N-1}^\top Q_N B_{N-1} u_{N-1} + 2 x_{N-1}^\top A_{N-1}^\top Q_N B_{N-1} u_{n-1} +$$

$$x_{N-1}^\top A_{N-1}^\top Q_N A_{N-1} x_{N-1} + \mathbb{E}_{w_{N-1}} \big[w_{N-1}^\top Q_N w_{N-1} \big] \Big\}$$

Linear Quadratic Regulator (LQR)

- Where we used the fact that:

$$\mathbb{E}_{w_{N-1}}[w_{N-1}^\top Q_N(A_{N-1}x_{N-1} + B_{N-1}u_{N-1})] = 0, \text{ since } \mathbb{E}[w_{N-1}] = 0$$

- And we pushed the expected value all the way to only contain terms that have w_{N-1}.

- Now differentiating w.r.t. u_{N-1} and setting the derivative equal to zero:

$$(R_{N-1} + B_{N-1}^\top Q_N B_{N-1})u_{N-1} + B_{N-1}^\top Q_N A_{N-1}x_{N-1} = 0$$

Linear Quadratic Regulator (LQR)

- Now by moving the term containing x_{N-1} to the right-hand side:

$$\mu^*(x_{N-1}) = u^*_{N-1} = -(R_{N-1} + B^\top_{N-1} Q_N B_{N-1})^{-1} B^\top_{N-1} Q_N A_{N-1} x_{N-1}$$

- Note that the above expression is **linear** in x_{N-1}. Hence if we define the matrix K_{N-1} as:

$$K_{N-1} = -(R_{N-1} + B^\top_{N-1} Q_N B_{N-1})^{-1} B^\top_{N-1} Q_N A_{N-1}$$

- Then the optimal closed-loop policy for $k = N - 1$ can be written as:

$$\mu^*_{N-1}(x_{N-1}) = K_{N-1} x_{N-1}$$

- which is a linear function of the state!

Linear Quadratic Regulator (LQR)

- Substituting the optimal policy back in the recursion for $k = N - 1$ we obtain:

$$J_{N-1}(x_{N-1}) = x_{N-1}^\top P_{N-1} x_{N-1} + \mathbb{E}_{w_{N-1}}\left[w_{N-1}^\top Q_N w_{N-1}\right]$$

- Where P_{N-1} is given by:

$$P_{N-1} = A_{N-1}^\top (Q_N - Q_N B_{N-1} (B_{N-1}^\top Q_N B_{N-1} + R_{N-1})^{-1} B_{N-1}^\top Q_N) A_{N-1} + Q_{N-1}$$

- and we note that the matrix P_{N-1} is symmetric and positive semidefinite since we can write:

$$x^\top P_{N-1} x = \min_u \left\{ x^\top Q_{N-1} x + u^\top R_{N-1} u + (A_{N-1} x + B_{N-1} u)^\top Q_N (A_{N-1} x + B_{N-1} u) \right\}$$

Linear Quadratic Regulator (LQR)

- Proceeding by induction to $N-2$, we can obtain a similar matrix P_{N-2}. Thus proceeding backwards for all $k \in \{0,1,\dots,N-1\}$, we can write the optimal closed-loop policy:

$$\mu_k^*(x_k) = K_k x_k$$

- where the matrix K_k (called the *control gain matrix*)is defined as:

$$K_k = -(R_k + B_k^\top P_{k+1} B_k)^{-1} B_k^\top P_{k+1} A_k$$

- and the symmetric positive semidefinite matrices P_k are given by the backwards recursion:

$$P_N = Q_N$$

$$P_k = A_k^\top (P_{k+1} - P_{k+1} B_k (B_k^\top P_{k+1} B_k + R_k)^{-1} B_k^\top P_{k+1}) A_k + Q_k, \; \forall k \in \{0, 1, \dots, N-1\}$$

Linear Quadratic Regulator (LQR)

- At the end of the recursion, we can write the optimal value function as:

$$J_0^*(x_0) = x_0^\top P_0 x_0 + \sum_{k=0}^{N-1} \mathbb{E}_{w_k}\left[w_k^\top P_{k+1} w_k\right]$$

- Several key remarks can be made:

  - The optimal closed-loop policy is **linear** in the states(This policy is often called the *linear feedback control law*).

  - The backwards recursion necessary to compute the matrices P_k can be done **very** efficiently.

  - The optimal value function is **quadratic** in the initial state x_0. The value function and the policy are simple and interpretable.

Linear Quadratic Regulator (LQR)

- Lastly:
  - The optimal closed-loop policy **does not** depend on the disturbance vectors $w_k{}'s$.

  - In particular, if we replace each $w_k{}'s$ by it's expected value, the optimal policy does not change.

  - In addition, even if the $w_k{}'s$ has non-zero mean, the optimal policy will only depend on the uncertainty via it's expectation (we leave the derivation as an exercise).

  - This phenomenon is called the ***Certainty Equivalence Principle*** and it appears on most (but not all) stochastic dynamic problems involving linear systems and quadratic costs.

The Riccati Equation

- Recall that the solution of the LQR problem can expressed as follows:

$$\mu_k^*(x_k) = K_k x_k$$

$$K_k = -(R_k + B_k^\top P_{k+1} B_k)^{-1} B_k^\top P_{k+1} A_k$$

$$P_N = Q_N$$

$$P_k = A_k^\top (P_{k+1} - P_{k+1} B_k (B_k^\top P_{k+1} B_k + R_k)^{-1} B_k^\top P_{k+1}) A_k + Q_k, \ \forall k \in \{0, 1, ..., N-1\}$$

- The last (backwards) recursion is so famous it has a name: **The discrete-time Riccati Equations.**

- Question: What happens to the Riccati equations if we take $k \to \infty$?

The Riccati Equation

- We will study Infinite-Horizon problems later when we cover Approximate DP and Reinforcement Learning, but as a preview, we will see that if we take $k \to \infty$, under mild assumptions the Riccati Equations will converge to the following algebraic form:

$$P = A^\top(P - PB(B^\top PB + R)^{-1}B^\top P)A + Q$$

- This equation above is called the **Algebraic Riccati Equation**. As we will show, this means that for the linear system:

$$x_{k+1} = Ax_k + Bu_k + w_k,\ \forall k \in \{0, ..., N-1\}$$

- When N is large, the optimal closed-loop policy will be stationary, that is $\pi^* = \{\mu^*, \mu^*, \dots, \mu^*\}$:

$$\mu^*(x) = Kx,\ K = -(B^\top KB + R)^{-1}B^\top PA$$

Controllability and Observability

- We take this opportunity to cover two very important concepts that are the core assumptions for the convergence of the Riccati Equations: **Controllability** and **Observability.**

Definition 1 *A pair of matrices (A, B), where A is an $n \times n$ matrix and B is an $n \times m$ matrix is said to be **controllable** os the $n \times nm$ matrix:*

$$\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}$$

has full rank (linearly independent rows).

*A pair os matrices (A, C), where C is an $m \times n$ is said to be **observable** if the pair $(A^\top, C^\top)$ is controllable.*

Controllability and Observability

- The notion of a controllable (A, B) can be explained intuitively: It states that for any initial state x_0, there exists a sequence of control vectors $(u_0, \dots, u_{N-1})$ that force the state x_n of the linear system:

$$x_{k+1} = Ax_k + Bu_k, \forall k \in \{0, ..., N-1\}$$

- to be equal to zero at time N.

- Observe that we successfully apply the dynamics to ``rollout" the dynamics, obtaining:

$$x_n = A^n x_0 + Bu_{n-1} + ABu_{n-2} + ..., A^{n-1}Bu_0$$

- where x_n is given explicitly as a function of the control sequence $(u_0, \dots, u_{N-1})$ and the initial state x_0 .

Controllability and Observability

- The previous is equivalent to (in matrix form):

$$x_n - A^n x_0 = \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} \begin{bmatrix} u_{n-1} \\ u_{n-2} \\ \vdots \\ u_0 \end{bmatrix}$$

- Now note that if (A, B) is controllable then the matrix on the right-hand side has full-rank.

- Then by appropriately selecting $(u_0, \dots, u_{N-1})$ we can obtain any vector in $\mathbb{R}^n$(which for example can be $-A^n x_0$, which would yield $x_n = 0$).

Controllability and Observability

- The notion of Observability can be seen intuitively in the context of estimation.

- Suppose we obtain some observations $z_0, z_1, \dots, z_{N-1}$of the form:

$$Z_k = Cx_k, \forall k \in \{0, ..., N-1\}$$

- It is possible to infer the initial state x_0 of the linear system $x_{k+1} = Ax_k$ by using the following:

$$\begin{bmatrix} z_{n-1} \\ \vdots \\ z_1 \\ z_0 \end{bmatrix} = \begin{bmatrix} CA_{n-1} \\ \vdots \\ CA \\ C \end{bmatrix} x_0$$

- In addition, we can verify that the above is equivalent to the property that, in the absence of control inputs, if \$$Cx_k \rightarrow 0$, then $x_k \rightarrow 0$.

Stability of Linear Systems

- Next, we present the very important notion of **Stability**: Suppose we utilize the stationary control policy:

$$\mu(x_k) = Kx_k,\ \forall k = 0, 1, \ldots$$

- And we replaced that in the linear systems definition, obtaining the *closed-loop system*:

$$x_{k+1} = (A + BK)x_k,\ \forall k = 0, 1, \ldots$$

- We say that the above system is **stable** if x_k tends to zero as $k \to \infty$.

- Again, by ``rolling-out" the system we can write:

$$x_k = (A + BK)^k x_0$$

Stability of Linear Systems

- So, given the closed-loop system:

$$x_k = (A + BK)^k x_0$$

- We can observe that it will be stable if and only if $(A + BK)^k \rightarrow 0$.

- This is equivalent to requiring that the eigenvalues of the matrix $(A + BK)$ are strictly within the unit circle.

- Lastly, we say that the pair of matrices $(A, B)$are **stabilizable** if there exists a matrix K such that $(A + BK)^k \rightarrow 0$.

Algebraic Riccati Equation

Proposition 1 *Let A be an $n \times n$ matrix, B be a $n \times m$ matrix, Q be and $n \times n$ positive semidefinite symmetric matrix. Consider the discrete-time Riccati Equation:*

$$P_{k+1} = A_k^\top (P_k - P_k B_k (B_k^\top P_k B_k + R_k)^{-1} B_k^\top P_k) A_k + Q_k, \ \forall k = \{0, 1, ...\}$$

where, w.l.o.g., the indices's are reversed: we start from an initial positive semidefinite matrix P_0. Assume that the pair (A, B) is controllable. Assume also that Q may be written as $Q = C^\top C$, where the pair (A, C) is observable. Then:

(a) There exists a positive definite symmetric matrix P such that, for every positive semidefinite symmetric initial matrix P_0 we have:

$$\lim_{k \to \infty} P_k = P$$

Further, P is the unique solution of the algebraic matrix equation:

$$P = A^\top (P - PB(B^\top PB + R)^{-1} B^\top P)A + Q$$

within the class of positive semidefinite symmetric matrices.

(b) The corresponding closed-loop system is stable; that is, the eigenvalues of the matrix:

$$D = A + BK$$

where:

$$K = -(B^\top KB + R)^{-1} B^\top PA$$

are strictly within the unit circle.