## 1.4 Dynamic programming, general setup

$$x_{t+1} = f(x_t, u_t, t), t = \{0, ..., T-1\}$$

Policy: $\Pi^* = \{u_0, ..., u_{t-1}\}, u_t = u_t(x_t)$ Cost:

$$J_\pi(x_0) = \sum g_t(x_t, u_t(x_t) + g_T(x_T)$$

Optimal cost:

$$\min J^*(x_0) = \min_\pi J_\pi(x_0)$$

Optimal policy, $\Pi^* = \{u_0*, ..., u^*_{t-1}\}$ such that $J^*(x_0) = J_{\pi^*}(x_0)$

## 1.5 Principle of optimality

Let $\Pi = \{u_0*, ...u^*_{t-1}\}$ be an optimal policy and look at the "tail subproblem". At time $\tau$, suppose that $x_\tau = z$. The tail subproblem is

$$\begin{array}{l} \text{minimize}_{u_\tau, ..., u_{\tau-1}} \sum_{t=\tau}^{T-1} g_t(x_t, u_t) + g_T(x_T) =: V_\tau(z) \\ \text{such that } x_\tau = z \\ \text{and } x_{t+1} = f(x_t, u_t, t) \end{array} \tag{10}$$

where $V_\tau(z)$ is the cost-to-go function.

**Theorem 1** *The tail policy $\Pi = \{u^*_0, ...u^*_{t-1}\}$ is optimal for the tail subproblem.*

The interpretation is that the optimality of the future does not depend on the past. The general strategy is to start with the shortest tail subproblem, solve it and keep working backwards. We now specialize to our problem.

Let $\tau = T, x_T = z$. Then $V_T(z) = z^\top Q_T z$, no control at $\tau = T$. Now we apply the dynamic programming principle, $\tau = T-1, x_{T-1} = z$,

$$\begin{aligned} V_{t-1}(z) &= \min_{u_{T-1}} z^\top Qz + u_{T-1}^\top R u_{T-1} + x_T^\top Q_T x_T, x_T = Ax_{T-1} + Bu_{T-1} \\ &= \min_{u_{T-1}} z^\top Qz + u_{T-1}^\top R u_{T-1} + (Az + Bu_{T-1})^\top Q_T (Az + Bu_{T-1}) \end{aligned} \tag{11}$$

We can solve for the optimal control input $u_{T-1}$ very easily:

$$u^*_{T-1} = u^*_{T-1}(z) = -(R + B^\top Q_T B)^{-1} B^\top Q_T A z$$

and

$$
\begin{aligned}
V_{T-1}(z) &= z^\top (Q + A^\top Q_T B(R + B^\top Q_T B)^{-1} B^\top Q_T A) z \\
&= z P_{T-1} z
\end{aligned} \tag{12}
$$

Note that

$$V_T = z^\top Q_T z =: z^\top P_T z$$

and

$$P_{T-1} = Q + A^\top P_T B(R + B^\top P_T B)^{-1} B^\top P_T A$$

The idea: Guess that $V_\tau(z) = z^\top P_\tau z$, for some $P_\tau$ and figure out how $P_\tau$ changes. In general,

$$u_t^* = -(R + B^\top Q_T B)^{-1}) B^\top P_{t+1} A x_t =: K_t x_t$$

with

$$P_t = Q + A^\top P_{t+1} A - A^\top P_{t+1} B(R + B^\top P_{t+1} B)^{-1} B^\top P_T A, V_t(z) = z^\top P_t z$$

with optimal cost,

$$J^*(x_0) = V_0(x_0) = x_0^\top P_0 x_0$$

1. Riccati recursion runs backwards in time
2. Feedback law can be precomputed independent of $x_0$.
3. Complexity is of order $O(Tn^3)$ instead of $O(T^3 np^2)$.

The general strategy is that we solve for $V_{T-1}$ and guess the form of $V_\tau$, then prove by induction. Note that we do not have any assumptions on $(A, B)$, nothing about stability or stabilizability. This is possible because we are looking at finite time horizon. An extreme case is that we have an unstable plant and $B = 0$ - this is still well defined, but this has terrible performance.

Looking at the cost function again:

$$J(u) = \sum_{t=0}^{N} x_t^\top Q x_t + u_t^\top R u_t + x_t^\top Q_T x_t$$

$V_t^* = \min_{u_t=0} z_t^\top Q_T z = z_t^\top P_T z.$
Now applying the DP principle,

$$\begin{aligned}
V_t^* &= \min_{u_t=0} z_t^\top Qz + u_{t-1}^\top Ru_{t-1} + V_T(x_T) \\
&= \min_{u_t=0} z_t^\top Qz + u_{t-1}^\top Ru_{t-1} + (Az + Bu_{t-1})^\top P_T(Az + Bu_{t-1}) \\
\\
V_{t-1}(z) &= z^\top(Q + A^\top P_T A - A^\top P_T B(R + B^\top P_T B)^{-1}B^\top P_T A)z = z^\top P_{T-1}z \\
P_{T-1} &= Q_A^\top P_T A - A^\top P_T B
\end{aligned} \tag{13}$$

where $u_T = k_T x_t, k_t = -(R + B^\top P_T B)^{-1}B^\top P_T B.$

## 1.6 Stochastic linear quadratic control over finite horizon

Consider the following discrete time dynamics:

$$x_{t+1} = Ax_t + Bu_t + w_t, u_t = u_t(x_t)$$

where $w_t$ is the process noise, or disturbance at time $t$. In particular, $w_t$ is independent and identically distributed (IID), of zero mean, and of known variance. This implies that $\mathbb{E}w_t = 0 = \mathbb{E}[w_t w_s^\top] = W\delta_{st}$, where $\delta_{st} = 1$, if $s = t$ and $\delta_{st} = 0$, if $s \neq 1$. $x_0$ is independent of the noise process $w_t$, $\mathbb{E}[x_0] = 0, \mathbb{E}[x_0 x_0^\top] = X_0$.

The difference now is that $x_{t+1}$ is a random variable, even if we know $x_t$ and $u_t$ and so the open loop approach can not work here.

Our cost function before was

$$J_{LQR}(u) = \sum_{t=0}^{N} x_t^\top Q x_t + u_t^\top R u_t + x_T^\top Q_T x_T$$

but in this context we can not solve for the optimal control input $u^*$ as now, it depends on the process noise $w_t$ and we do not have access to it. In our setting, we have the information on the statistics of $w_t$ and we can exploit this to get the average stochastic cost function:

$$J_{SLQR} = \mathbb{E}[J_{LQR}]$$

Another possible set up, is that $\mathbb{E}[WW^\top] = W$ is a measure of energy and can look at worst case performance:

$$J_{WC} = \max_w J_{LQR}.$$

Instead of including statistics, one can bound the infinity norm of noise, $||w||_\infty \leq k$ and input $||u||_\infty \leq k$, ($\mathcal{L}_1$ optimal control).

We choose $J(x,u)$ based on the known structure of the disturbances and performance requirements. In this class, we will look at the expected LQR cost:

$$\min_u \mathbb{E}[\sum_t x_t^\top Q x_t + u_t^\top R u_t + x_T^\top Q_T x_T]$$

*Interpretation* : Let $Q = C^\top C$, as before, and $R = \rho I, y = Cx_t$. The cost function is:

$$J(x,u) = \sum_t TrY_t + \rho TrU_t + TrY_T$$

i.e. we want to minimize the output and control effort variance.

We can not do the open loop strategy because of the dependence on $w_t$, and so $x_{t+1}$ becomes a random variable. So we use dynamic programming but now our cost to go we has an expected value $\mathbb{E}[\cdot]$ wrapped around.

$$V_t(z) = \mathbb{E}[\min_u g_t(x_t, u_t(x_t)) + V_{t+1}(x_{t+1})]$$

For $t = T, x_T = z$,

$$V_T(z) = \mathbb{E}_{w_T}[x_T^\top Q_T x_T | x_T = z] = z^\top Q_T z = z^\top P_T z$$

.

For $t = T-1, x_{T-1} = z, x_T = Az + Bu_{T-1} + w_{T-1}$,

$$\begin{aligned}
V_{T-1}(z) &= \min_{u_{T-1}} \mathbb{E}_{w_{T-1}}[x_{T-1}^\top Q x_{T-1} + u_{T-1}^\top R u_{T-1} V_T(x_T) | x_{T-1} = z] \\
&= \min_{u_{T-1}} z^\top Q z + u_{T-1}^\top R u_{T-1} \\
&\quad + \mathbb{E}[(Az + Bu_{T-1} + w_{T-1})^\top P_T (Az + Bu_{T-1} + w_{T-1})] \\
&= z^\top Q z + z^\top A^\top P_T A z + 2\mathbb{E}[z^\top A^\top P_T w_{T-1}] \\
&\quad + \min_{u_{T-1}} u_{T-1}^\top (R + B^\top P_T B) u_{T-1} \\
&\quad + 2u_{T-1}^\top B^\top P_T A z + 2\mathbb{E}[w_{T-1}^\top P_T B u_{T-1}] \\
&= z^\top (Q + A^\top P_T A) z + Tr(P_T W) + \\
&\quad + \min_{u_{T-1}} u_{T-1}^\top (R + B^\top P_T B) u_{T-1} + 2u_{T-1}^\top B^\top P_T A z
\end{aligned} \tag{14}$$

This implies that

$$
\begin{array}{rcl}
u_{T-1}(z) & = & K_{T-1}z, K_{T-1} = -(R + B^\top AB)^{-1}B^\top P_T A) \\
V_{T-1}(z) & = & z^\top(Q + A^\top P_T A - A^\top P_T B(R + B^\top P_T A))z + Tr(P_T W) \\
& =: & \underbrace{z^\top P_{T-1} z}_{\text{quadratic LQR}} + \underbrace{q_{T-1}}_{\text{running cost}}
\end{array} \tag{15}
$$

For the induction step, now assume that $V_t(z) = z^\top P_t z + q_t$ and do the same exercise to get:

$$
P_t = Q + A^\top P_{t+1} A + A^\top P_{t+1} B K_t, \tag{16}
$$

where $K_t = -(R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A, P_T = Q_T$.

$$
\begin{array}{ll}
q_t & = q_{t+1} + Tr(P_{t+2}W), q_T = 0 \\
u_t & = k_t x_t
\end{array} \tag{17}
$$

We get that the optimal policy is identical to the deterministic one and independent of $X_0$ and $W$. This is a special case of the certainty equivalency principle, which states that the result of an optimization is the same as for the corresponding deterministic problem where $w$ is replaced by $\mathbb{E}[w]$.

The optimal cost

$$
J^* = \mathbb{E}[v_0] = TrX_0P_0 + q_0 = TrX_0P_0 + \underbrace{\sum_{t=1}^{T} TrWP_t}_{\text{cost of not knowing W}} .
$$

*Interpretation:* Suppose $x_0$ is known and $W = 0$ i.e. $w_t = 0, \forall t$. This reduces to $x_0^\top P_0 x_0 = J^*_{LQR}$. So now if $x_0$ is unknown, $Tr(X_0P_0) = \mathbb{E}[J^*_{LQR}]$, for $W = 0$, i.e. the average cost incurred by the initial conditions $x_0$.

So now let's look at the behavior of $x_{t+1}$ as a stochastic process:

$$
\mathbb{E}x_0 = 0, \mathbb{E}x_0 x_0^\top = X_0
$$

$$
\mathbb{E}x_1 = \mathbb{E}((A + BK)x_0 + w_0 = 0), \mathbb{E}x_1 x_1^\top = (A + BK)X_0(A + BK)^\top + W
$$

$$
\mathbb{E}x_{t+1} = 0, \mathbb{E}x_{t+1} x_{t+1}^\top = (A + BK)X_t(A + BK)^\top + W
$$

$$
X_{t+1} = (A + BK)X_t(A + BK)^\top + W
$$

Suppose that we know $x_t = 0$: then this implies that $\mathbb{E}x_t x_t^\top = X_t = 0$, or from the equation that we just derived, that $X_{t+1} = W$. So we find ourselves in a very similar situation in trying to interpret $TrWP_{t+1}$: it is our average optimal cost-to-go starting at $t$, and having covariance $W$ in our "initial condition" $x_t$, and $w_t = \cdots = w_{T-1} = 0$.

Thus the cost is in fact a linear superposition of the average energy, or variance, transferred to the system by each element of the disturbance (i.e. $x_0, w_0, \ldots, w_{T-1}$).

**Example 1** *Let*

$$x_{t+1} = A_t x_t + B_t u_t + w_t,$$

*where now we allow* $(A_t,\, B_t)$ *to be random variables. We can apply the same methods, but now we take the expectation over* $A_t$ *and* $B_t$*, as well as* $w_t$*.*

$$\begin{aligned} V_t^* &= \min_{u_t} \mathbb{E}[x_t^\top Q x_t + u_t^\top R u_t \\ &\quad +(A_t x_t + B_t u_t + w_t)^\top P_{t+1}(Ax_t + B_t u_t + w_t)|x_t = z] \\ &= z^\top Q z + z^\top \mathbb{E}[A_t^\top P_{t+1} A_t] z + Tr(P_{t+1}W) \\ &\quad + \min_{u_t} u_t^\top (R + \mathbb{E}[B_t^\top P_{t+1} B_t] u_t + 2u_t^\top \mathbb{E}[B_t^\top P_{t+1} A_t] z \end{aligned} \tag{18}$$

*The optimal control is:*

$$u_t^* = -(R + \mathbb{E}[B_t^\top P_{t+1} B_t])^{-1} \mathbb{E}[B_t^\top P_{t+1} A_t] z$$

*and*

$$P_t = Q + \mathbb{E}[A_t^\top P A_t] - \mathbb{E}[A^\top P_{t+1} B_t](R + \mathbb{E}[B_t^\top P_{t+1} B_t])^{-1} \mathbb{E}[A^\top P_{t+1} A_t].$$

*We can use this idea to solve a networked control system problem where a controller communicates with an actuator across a lossy link. In particular, assume that with probability* $p$*, the control action gets through, and with probability* $1 - p$*, it does not, and that drops are independent of each other.*

*In that case, our stochastic state space representation is given by*

$$\begin{aligned} A_t &= A,\ \forall t \geq 0 \\ B_t &= \begin{cases} B & w.p.\ p \\ 0 & w.p.\ 1-p \end{cases} \end{aligned} \tag{19}$$

*and we can solve the LQR problem by noting that*

$$\mathbb{E}[A_t] = A$$
$$\mathbb{E}[B_t] = pB,$$

*and*

$$\mathbb{E}[B_t^\top P_{t+1} B_t] = pB^\top P_{t+1} B$$

.