B9120 Dynamic Programming Lecture 2 - 02/03/2020

Inventory Control and Linear-Quadratic Control

Lecturer: Daniel Russo *Scribe: Utkarsh Patange, Shangzhou Xia*

Outline

- Optimality of base-stock policies in inventory control
- Optimality of linear policies in linear quadratic control

1 The template for structural DP arguments

In this section of the course, many arguments follow the same template:

- We first recognize that the terminal cost function J_N^* has a nice property (e.g. convexity, or monotonicty).
- We then argue that this property implies that the policy μ_{N-1}^* has some nice structure (e.g. a threshold policy is opimal).
- To extend this by induction, we show that if a cost function J satisfies the property, then the next cost function
$$J^+(x) = \min_{u \in U(x)} \mathbb{E}\left[g(x,u,w) + J(f(x,u,w))\right]$$
that is generated by a step of the DP algorithm will also satisfy this property.

In the sequel we apply such arguments in the studying inventory control and linear systems with quadratic cost ("LQ" control problems). In these problems, the convexity of the cost-to-go functions play a key role.

2 Operations that preserve convexity

We note some operations that preserve convexity which will be useful in showing the convexity of cost-to-go functions. For a detailed treatment, please refer to the book on Convex Optimization [Boyd and Vandenberghe].

- Non-negative weighted sums:
  - If $f_1, \ldots, f_m : \mathcal{D} \to \mathbb{R}$ are convex and $w_!, \ldots, w_m \geq 0$, then $w_1 f_1 + \cdots + w_m f_m$ is convex.
  - For some $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, the expectation $g : \mathcal{X} \to \mathbb{R}$ defined as
  $$g(x) = \int f(x,y) w(y)\, \mathrm{d}y$$
  is convex if $w(y) \geq 0$ and the mapping $x \mapsto f(x,y)$ is convex for all $y \in \mathcal{Y}$.
- Composition with an affine map: $g(x) = f(Ax+b)$ is convex if f is convex.
- Point wise supremum: $g(x) = \sup_{y \in \mathcal{Y}} f(x,y)$ is convex if $x \mapsto f(x,y)$ is convex for all $y \in \mathcal{Y}$.

3 Inventory Control Problems

Problem formulation Inventory control problems are classical finite-horizon optimization problems that can be solved using dynamic programming approaches. Assume that the firm begins with some inventory x_0 and, in each priod k, decides upon ordering replenishment $u_k \geq 0$ after observing a stochastic demand the inventory position $x_k \geq 0$ in the current time period $k \in \{0, 1, 2, \ldots, N-1\}$. The evolution of the inventory follows the recursive relation

$$x_{k+1} = x_k + u_k - w_k, \qquad \forall\, k = \{0, 1, \ldots, N-1\}$$

and unfulfilled orders are allowed to be backlogged ($x_k < 0$) until replenishment products fulfill them. The firm's objective is to minimize the overall expected cost:

$$\mathbb{E}_{\{w_k\}}\left[\sum_{k=0}^{N-1} g_k(x_k, u_k, w_k) + g_N(x_N)\right],$$

where expectation is taken with respect to the random demands $\{w_k\}$ which is assumed to be i.i.d. Per-stage cost functions take the form

$$\begin{aligned} g_k(x_k, u_k, w_k) &= cu_k + r(x_k + u_k - w_k), \qquad \forall\, k = \{0, 1, \ldots, N-1\} \\ g_N(x_N) &= 0, \end{aligned}$$

where we define:

$$r(x) = p \max\{0, -x\} + h \max\{0, x\},$$

which denotes the cost due to backlogging or holding every period. *Throughout, we assume that $p > c$ so as to exclude the trivial setting in which it is optimal to never order inventory and let all demand go unfulfilled.* Another property that will be useful to our analysis is that of *coercivity* of a function, defined as:

Definition 1. *A function $r : \mathbb{R} \to \mathbb{R}$ is coercive if $r(x) \to \infty$ as $|x| \to \infty$.*

It is easy to see that the function, $r(\cdot)$, is both convex (sum of convex functions) and coercive (the costs are infinite for infinite inventory or infinite backlogging). We now present the main result for this inventory control problem which shows that base-stock policies are optimal. Such a result is quite interesting. For a continuous state space (like in this problem), we can characterize the optimal policy just using a bunch of scalars which can be easily stored on a computer.

Proposition 2. *(Base-stock policies are optimal) There exists an optimal policy $\pi^* = (\mu_0^*, \ldots, \mu_{N-1}^*)$ such that,*

$$\mu_k^*(x_k) = (S_k - x_k)^+ = \begin{cases} S_k - x_k, & x_k \leq S_k \\ 0, & \textit{otherwise}, \end{cases}$$

for some scalars $S_0, S_1, \ldots, S_{N-1}$.

Proof. To prove this claim, we parameterize the problem using target inventory positions, $y_k := x_k + u_k$ for each k. Our goal is to show that $y_k = \max\{S_k, x_k\}$ is optimal which implies the base-stock policy stated above is optimal. For $y \geq x$, define

$$Q_k^*(x, y) = \mathbb{E}\left[c(y - x) + r(y - w_k) + J_{k+1}^*(y - w_k)\right] = \underbrace{\mathbb{E}\left[cy + r(y - w_k) + J_{k+1}^*(y - w_k)\right]}_{G_k(y)} - cx.$$

Then, the DP algorithm yields for each k that

$$J_k^*(x) = \min_{y \geq x} \left[G_k(y) - cx\right]. \tag{1}$$

The constraint that $y \geq x$ is due to the fact that we cannot order negative inventory. We define

$$S_k^* \in \arg\min_{y\in\mathbb{R}} G_k(y)$$

to be the ideal inventory position in period k. Be warned, however, that the **existence** of such a scalar remains in question, so we proceed to prove that $G_k(\cdot)$ is **convex** and **coercive**, which implies that an unconstrained minimizer, S_k^*, exists.

- Claim: G_k is convex.

  First note $G_{N-1}(\cdot)$ is convex as $J_N^*(\cdot) = 0$ as well as $r(\cdot)$ are both convex. More precisely, $G_{N-1}(y)$ is a sum of a linear term cy, the expectation $E_{w_{N-1}}[r(y - w_{N-1})]$ which is convex (expectation preserves convexity and $r(y - w_{N-1})$ is convex in y for all w_{N-1}) and $J_N^*(y - w_{N-1})$ which is 0. Further note that for any convex function $F(\cdot)$, the function $g(x) = \min_{y\geq x} F(y)$ is also convex. (draw a picture) Hence,

  $$J_{N-1}^*(x) = \min_{y\geq x} G_{N-1}(y) - cx$$

  is also convex. We use backward induction to complete the argument for all k periods.

- Claim: G_k is coercive.

  Since the expected future cost function $J_{k+1}^*(\cdot)$ is bounded below by zero[1], we have that

  $$\begin{aligned} G_k(y) &\geq \mathbb{E}[cy + r(y - w_k)] \\ &= \mathbb{E}[c(y^+ - y^-) + p(y - w_k)^- + h(y - w_k)^+] \\ &\geq \mathbb{E}[-cy^- + py^- + h(y - w_k)^+] \qquad (y^+ \geq 0 \text{ and } w_k \geq 0) \\ &= (p - c)y^- + h\mathbb{E}[(y - w_k)^+] \end{aligned}$$

  As $p > c$ and $h > 0$, the lower bound is clearly coercive. The first term, $(p - c)y^- \to \infty$ as $y \to -\infty$. The second term, $E[(y - w_k)^+] \to \infty$ as $y \to \infty$. Hence, $G_k(y) \to \infty$ as $y \to \pm\infty$.

Therefore, we have showed the existence of an unconstrained minimizer S_k^* of G_k, so the DP algorithm guarantees the optimality of the base-stock policy as:

$$\arg\min_{y\geq x} G_k(y) = \begin{cases} S_k^*, & \text{if } S_k^* \geq x \\ x_k, & \text{otherwise} \end{cases}$$

where recall that $S_k^* := \arg\min_{y\in\mathbb{R}} G_k(y)$ □

3.1 On the extension to fixed costs

A large operations literature treats more realistic generalizations of the inventory control problem, such as the incorporation of fixed order costs. In this case, the per-stage cost would be

$$g_k(x, u, w) = \begin{cases} r(x + u - w), & u = 0 \\ K + cu + r(x + u - w), & u > 0 \end{cases}.$$

In this problem, base-stock policies constructed above are no longer optimal. This makes intuitive sense: if the inventory level x_k is already close to the base-stock level S_k^* then replenishment would not offer enough benefit to justify the fixed cost.

[1] Per-period costs are non-negative which imply cumulative costs have to be non-negative as well.

It turns out that the optimal policy takes the form of a multiperiod (s, S) policy. This policy takes the form

$$\mu_k^*(x_k) = \begin{cases} 0, & \text{if } s_k^* \geq x \\ S_k - x_k, & \text{if } x_k < s_k \end{cases}$$

where $S_k = \arg\min_y G_k(y)$ and $s_k = \min\{y | G_k(y) = K + G_k(S_k)\}$ for each k.

Showing the optimality of such policies is more difficult than showing the optimality of base-stock policies, because the function $G_k(\cdot)$ is generally no longer convex. Pioneering work by Scarf observed that these functions satisfy a relaxed notion of convexity he called K–convexity. Volume 1, section 3.2 of Bertsekas' textbook contains an inductive argument showing that G_k is $K-$convex. That is, G_k satisfies the following property:

$$K + G_k(z + y) \geq G_k(y) + z\left(\frac{G_k(y) - G_k(y-b)}{b}\right), \qquad \text{for all } z \geq 0, b > 0, y.$$

Remark: *I find it helpful to to think of this as a relaxation of the fundamental property that a convex function lies above all its tangents. When first grappling with this formula, it may be helpful to suppose G_k is differentiable at y and take $b \to 0$ so that $\frac{G_k(y)-G_k(y-b)}{b} \to G_k'(y)$. Then, you can roughly think of this condition as saying that G_k never lies below its tangent by more than the fixed amount $K > 0$. A more precise understanding requires working through the analysis, as this condition is really engineered to make the proof work.*

4 LQ Control

See also Bertsekas, *Dynamic Programming and Optimal Control* Vol. 1 Section 3.1

Problem Statement Here, we consider the special case of linear system when the cost is quadratic. We have,

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k & k &= \{0, 1, \ldots, N-1\} \\ g(x_k, u_k) &= x_k^\top Q x_k + u^\top R u & k &= \{0, 1, \ldots, N-1\} \end{aligned}$$

Here, we assume that $x_k \in \mathbb{R}^n, u_k \in \mathbb{R}^m$ and the cost matrices Q, R are symmetric and positive definite (i.e., $Q \in \mathbb{S}^n_{++}, R \in \mathbb{S}^m_{++}$. We also assume the noise $\{w_k\}$ to be i.i.d with zero mean (i.e., $\mathbb{E}(w_k) = 0$) and finite second moment. Note, we only need to assume the noise to be independent, however assuming the noise to be i.i.d simplifies notation. Similarly, the matrices A, B, R, Q can all be period-dependent but we will assume these to be fixed for simplicity.

Proposition 3. *The optimal policy $\pi^* = \left(\mu_0^*, \ldots, \mu_{N-1}^*\right)$ takes the form $\mu_k^*(x) = L_k x$, where the driving matrix $L_k \in \mathbb{R}^{m \times n}$ doesn't depend on the distribution of w_k's.*

Since L_k does not depend on noise distribution, the optimal controller is the same as that in a problem with deterministic dynamics (i.e. $w_k = 0$). This surprising property is called *certainty equivalence* and it makes it possible for L_k's to be computed by "recursive linear algebra".

Proof. Let us set up some notation. We define the optimal cost-to-go functions as:

$$\begin{aligned} J_N^*(x) &= x^\top K_N x + c_N \\ J_{N-1}^*(x) &= \min_u \left[g(x, u) + \mathbb{E}_w \left[J_N^*(f(x, u, w))\right]\right] \end{aligned}$$

where $K_N = Q, c_N = 0$ and $f(x,u,w) = Ax + Bu + w$. To simplify notation, let us define:

$$h(x,u) = g(x,u) + \mathbb{E}_w \left[J_N^*(f(x,u,w)) \right].$$

Our aim is to prove that the given policy is optimal by the principle of mathematical induction. We prove this in 3 steps.

Step 1: <u>$h(x,u)$ is convex quadratic in (x,u).</u>

Why? This is because

- g is convex quadratic
- $J_N^*(x)$ is convex quadratic
- $(x,u,w) \mapsto f(x,u,w)$ is an affine function, and the composition of a convex quadratic function with an affine function is convex quadratic as well.

To see this more explicitly, note that the function $x \to J_N^*(x)$ is a convex function. $(x,u,w) \mapsto f(x,u,w)$ is an affine function. Thus, $\forall w\ (x,u) \to J_N^*(f(x,u,w))$ is a convex function, and taking expectation of this w.r.t w preserves convexity. Since, $g(x,u)$ is also convex, $h(x,u)$ is just a sum of two convex functions, and hence is convex in turn.

To see that it is also quadratic, we write its full expansion. We have,

$$\begin{aligned} h(x,u) &= u^\top Ru + x^\top Qx + \mathbb{E}\left[(Ax + Bu + w)^\top K_N (Ax + Bu + w) \right] + c_N \\ &= u^\top (R + B^\top K_N B)u + x^\top (Q + A^\top K_N A)x + 2x^\top A^\top K_N Bu + \mathbb{E}\left[w^\top Qw \right] + c_N \end{aligned}$$

which, clearly, is quadratic in (x,u).

Step 2: <u>The minimizer $x \longrightarrow \arg\min_{u \in R^m} h(x,u)$ is a linear function.</u>

To see this, we apply the first order conditions for minimality. At the point of minimality, the first derivative $\nabla_u h(x,u)$ should vanish. We have,

$$\begin{aligned} \nabla_u h(x,u) &= 2\left(R + B^\top K_N B\right)u + 2B^\top K_N Ax \\ \nabla_u h(x,u) = 0 &\Rightarrow u = -(R + B^\top K_N B)^{-1} B^\top K_N Ax \\ \therefore \mu_{N-1}^*(x) &= L_{N-1}^* x \qquad \text{where } L_{N-1}^* = -(R + B^\top K_N B)^{-1} B^\top K_N A \end{aligned}$$

Note here, that indeed, L_{N-1}^* doesn't depend on the distribution of w_k's.

Step 3: <u>Induction step</u>

To complete the proof, note that $J_{N-1}^*(x) = \min_u h(x,u) = h(x, L_{N-1}^* x)$ is a composition of a convex quadratic h with a linear function $x \mapsto (x, L_{N-1}^* x)$, and thus $J_{N-1}^*(x)$ is also convex quadratic. More explicitly, we can write

$$J_{N-1}^*(x) = \min_u h(x,u) = h(x, L_{N-1}^* x) = x^\top K_{N-1} x + c_{N-1}$$

$$\begin{aligned} \text{Where} \quad K_{N-1} &= {L_{N-1}^*}^\top (R + B^\top K_N B) L_{N-1}^* + Q + A^\top K_N A + 2A^\top K_N B L_{N-1}^* \\ &= A^\top K_N B (R + B^\top K_N B)^{-1} B^\top K_N A - 2A^\top K_N B (R + B^\top K_N B)^{-1} B^\top K_N A + Q + A^\top K_N A \\ &= A^\top (K_N - K_N B (B^\top K_N B + R)^{-1} B^\top K_N) A + Q \\ c_{N-1} &= c_N + \mathbb{E}\left(w^\top Qw \right) \end{aligned}$$

The above recurrence relation is called the Riccati equation. We note here, that K_{N-1} is symmetric whenever K_N, Q and R are symmetric. K_{N-1} is positive semi-definite, since whave just concluded that J_{N-1}^* is convex. (This can also be verified directly) These are all the ingredients required to take the induction further and prove that $\mu_{N-2}^*(x)$ is also linear in x with the same arguments as above. Thus, with the principle of mathematical induction, we have managed to prove that the optimal policy $\pi^* = (\mu_0^*, \ldots, \mu_{N-1}^*)$ takes the form $\mu_k^*(x) = L_k x$ as required. □