- One very large case of exact solutions arises when the value function at each stage inherits some structure owing to the same structure being present at the next stage. We will see 3 important examples.

- Linear value function and bang-bang control

Eg (Sequential investment with consumption)

- An investor has income X_t dollars at start of each day t (initial X_0)
- She consumes an amount A_t, and invests $X_t - A_t$
- The investment on day t gives return $\theta_t \sim F_t$ per dollar on each day till end of horizon
- Objective - Maximize total consumption $\sum_{t=0}^{T-1} A_t$

- Income on day $t+1$: $X_{t+1} = X_t + \theta_t (X_t - A_t)$
- Define $V_T(x) = 0 \quad \forall x$ (can not consume after day T)

$$V_t(X_t) = \max_{a \in [0, X_t]} \left[a + \mathbb{E}[V_{t+1}(X_t + \theta_t(X_t - A_t))] \right]$$

- $$V_{T-1}(x) = \max_{a \in [0,x]} \left[a + \mathbb{E}[V_T(x + \theta_{T-1}(x-a))] \right] = x$$

- Now suppose $V_t(x) = \alpha_t x \quad \forall x$

$$\Rightarrow V_{t-1}(x) = \max_{a \in [0,x]} \left[a + \mathbb{E}[\alpha_t (x + \theta_t(x-a))] \right]$$

$$= \underbrace{\begin{cases} (\alpha_t + \alpha_t \mathbb{E}[\theta_t])\, x & ; \alpha_t \mathbb{E}[\theta_t] \geq 1 \\ (\alpha_t + 1)\, x & ; \alpha_t \mathbb{E}[\theta_t] < 1 \end{cases}}_{\alpha_{t-1}}$$

Thus $\underbrace{A_t \text{ is either } 0 \text{ or } X_t}_{\text{bang-bang control}} \iff \underbrace{V_t \text{ is } \alpha_t X_t}_{\text{linear value fn}}$

- ## Linear quadratic regulator (LQR)

- We want to control a trajectory X_t via controls A_t
- Linear dynamics — $X_{t+1} = A_t X_t + B_t A_t + Z_t$

known matrices   ⫫ noise, $E[Z_t]=0$

Quadratic costs — $C_t = X_t^T Q_t X_t + A_t^T R_t A_t$

known, psd matrices

Finally at time T, we have termination cost $G_T = X_T^T Q_T X_t$

- $V_t(x) = \min_{a \in \mathbb{R}^d} \left[x^T Q_t x + a^T R_t a + \mathbb{E}\left[V_{t+1}(A_t x + B_t a + Z_t)\right]\right]$

- Consider the system in 1-d, with $t = T-1$

$$V_{T-1}(x) = \min_a \mathbb{E}\left[Q_{T-1} x^2 + R_{T-1} a^2 + Q_T (A_{T-1} x + B_{T-1} a + Z_{T-1})^2\right]$$

0 as $E[Z_t]=0$

$$= \min_a \mathbb{E}\Big[(R_{T-1} + Q_T B_{T-1}^2) a^2 + 2 Q_T B_{T-1}(A_{T-1} x + \cancel{Z_{T-1}}) a + Q_T Z_{T-1}^2$$
$$+ (Q_{T-1} + Q_T A_{T-1}^2) x^2 + 2 Q_T A_{T-1} x \underbrace{\cancel{Z_{T-1}}}_{0}\Big]$$

$$\Rightarrow a^*_{T-1}(x) = -\frac{Q_T B_{T-1}(A_{T-1} x)}{(R_{T-1} + Q_T B_{T-1}^2)}$$

and $V_{T-1}(x) = \left(\dfrac{Q_T^2 B_{T-1}^2 A_{T-1}^2}{(R_{T-1} + Q_T B_{T-1}^2)} + Q_{T-1} + Q_T A_{T-1}^2\right) x^2 + Q_T \mathbb{E}[Z_{T-1}^2]$

Similarly suppose $V_t(x) = K_t x^2 + C_t$, then we again have $a^*_{t-1}(X_{t-1}) = L_{t-1} X_{t-1}$ and $V_{t-1}(X_{t-1}) = K_{t-1} X_{t-1}^2 + C_{t-1}$!

- This extends if Q_t, R_t are psd

does not effect controls

$a^*_t(X_t) = L_t X_t$, $V_t(X_t) = X_t^T K_t X_t + C_t$, where

(discrete-time) Riccati Equations
$$\begin{cases} L_t = -(B_t^T K_{t+1} B_t + R_t)^{-1} B_t^T K_{t+1} A_t \\ K_t = A_t^T (K_{t+1} - K_{t+1} B_t (B_t^T K_{t+1} B_t + R_t)^{-1} B_t^T K_{t+1}) A_t + Q_t \end{cases}$$

Convex value fns and threshold policies

- In the previous examples, we saw 2 general closed-form solns
  1) linear value fn $\Rightarrow$ bang-bang control + linear value fn
  2) Quadratic value fn $\Rightarrow$ linear control + quadratic value fn

  Finally we will see a third conservation law

Eg (The Newsvendor)

- $X_t \equiv$ Stock at start of day t, A_t = stock ordered in day t

  $D_t \equiv$ (unknown) demand in day t $\sim F_t$

  $\Rightarrow X_{t+1} = X_t + A_t - D_t$

  Here we allow X_t to be negative (backorders)

- Costs - Holding cost $h(X_t + A_t - D_t)^+$

  Backorder cost $b(D_t - A_t - X_t)^+$

  Ordering cost $c(A_t)$

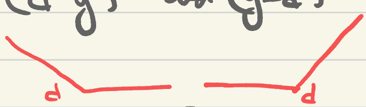
- Terminal condn - $V_T(x) = 0 \;\forall x$
- HJB eqn

$$V_t(x) = \min_{a \geq 0}\Big[ca + \overbrace{\mathbb{E}\big[h(x+a-D_t)^+ + b(D_t-a-x)^+\big]}^{H(x+a)} + \mathbb{E}\big[V_{t+1}((x+a-D_t))\big]\Big]$$

$$\Rightarrow V_t(x) + cx = \min_{y \geq x}\Big[cy + H(y) + \mathbb{E}\big[V_{t+1}((y-D_t))\big]\Big]$$

- $$V_{T-1}(x) = -cx + \min_{y \geq x}\Big[cy + \mathbb{E}\big[h(y-D_{T-1})^+ + b(D_{T-1}-y)^+\big]\Big]$$

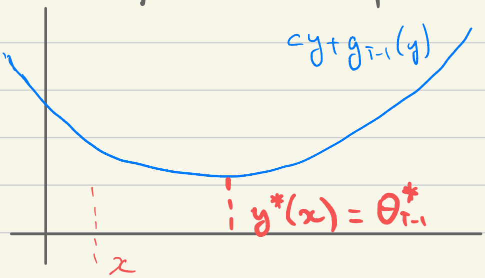
now observe 1) $(d-y)^+$ and $(y-d)^+$ are both convex

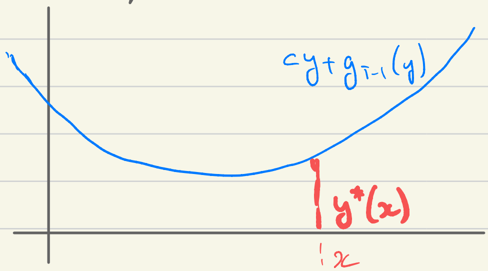


2) $\mathbb{E}[f(x,Z)]$ is convex in x if f is convex

$$V_{T-1}(x) = -cx + \min_{y \geq x}\Big[cy + \underbrace{g_{T-1}(y)}_{\text{convex fn of } y}\Big]$$

To find the optimal action, observe





$$\Rightarrow \quad y^*(x) = \begin{cases} \theta^*_{T-1} & ; x \le \theta^*_{T-1} \\ x & ; x > \theta^*_{T-1} \end{cases}, \text{ with } \theta^*_{T-1} = \min_{y \in \mathbb{R}} \left[cy + g_{T-1}(y) \right]$$

- Moreover, suppose $V_t(x)$ is convex in x

$$V_{t-1}(X_{t-1}) = -cx + \min_{y \geq X_{t-1}} \left[cy + H_{t-1}(y) + \mathbb{E}[V_t(y - D_{t-1})] \right]$$

Now we can repeat the same argument: cy is linear, $H_{t-1}(y) = \mathbb{E}[h(y-D_{t-1})^+ + b(D_{t-1}-y)^+]$ is convex in y and $\mathbb{E}[V_t(y-D_{t-1})]$ is a convex combination of convex fns $\Rightarrow$ convex

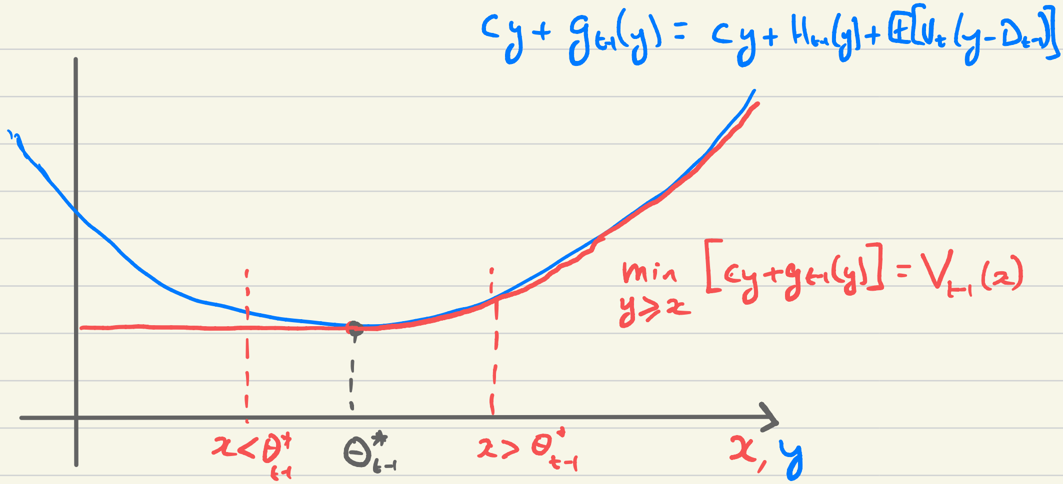
$$\Rightarrow \quad y^*_{t-1}(X_{t-1}) = \begin{cases} \theta^*_{t-1} & ; X_{t-1} \le \theta^*_{t-1} \\ X_{t-1} & ; X_{t-1} > \theta^*_{t-1} \end{cases}$$

$$\Rightarrow \quad A^*_{t-1}(X_{t-1}) = \begin{cases} \theta^*_{t-1} - X_{t-1} & ; X_{t-1} \le \theta^*_{t-1} \\ 0 & ; X_{t-1} \geq \theta^*_{t-1} \end{cases} \quad \text{(order upto policy)}$$

where $\theta^*_{t-1} = \operatorname{argmin}_{y \in \mathbb{R}} \left[cy + H_{t-1}(y) + \mathbb{E}[V_t(y - D_{t-1})] \right]$

Thus convex value fn $\Rightarrow$ opt policy is threshold-type

But is $V_{t-1}(x)$ still convex?



Imagine increasing x ; then $V_{t-1}(x)$ behaves as above

$\Rightarrow \quad V_{t-1}(x) = \min[cx + g_{t-1}(x), c\theta^*_{t-1} + g_{t-1}(\theta^*_{t-1})]$

Clearly this is convex! Thus we can continue the argument inductively.

Thus we have seen 3 examples of value fn structure $\Rightarrow$ policy structure $\Rightarrow$ value fn structure.

1) 'linear dynamics + linear cost' $\Leftrightarrow$ 'bang-bang control'
2) 'linear dynamics + quadratic cost' $\Leftrightarrow$ 'linear control'
3) 'linear dynamics + convex cost' $\Leftrightarrow$ 'threshold policies'

Note 1: These are both very general, but also need care with checking assumptions. For example, for LQR, we had 0-mean noise and no constraints; for inventory control, we assumed backorders were allowed...

Note 2: From an optimization perspective, what is happening here is that these value fn class - policy class pairs are the solution class of the corresponding dual LP (ie, the HJB eqns) and primal LP (ie., the state-action freq LP).