EE365: Value

Value function

- suppose you will receive a reward $g(x_1)$ depending on the state at $t = 1$
- how much should you pay at time $t = 0$ be in state i?

Define the *value* of state i, given by v_i, to be

$$v_i = \mathbf{E}\big(g(x_1) \mid x_0 = i\big)$$

(the term 'value' makes more sense when g_t is a reward, not a cost)

Value function

we have

$$\begin{aligned} v_i &= \mathbf{E}\big(g(x_1) \mid x_0 = i\big) \\ &= \sum_{j \in \mathcal{X}} \mathbf{Prob}(x_1 = j \mid x_0 = i) g_j \\ &= (Pg)_i \end{aligned}$$

- $v = Pg$ is the current value of reward g at the next time step (costs)
- *left* multiplication by P maps future reward back one step
- v_i is a *weighted average* of value of g at children of i
- recall *right* multiplication of π_t by P maps distribution forwards one step

Value propagation

suppose we iterate

$$\begin{aligned} v_0 &= g \\ v_{k+1} &= P v_k \qquad \text{for } k = 0, 1, \ldots \end{aligned}$$

- $(v_k)_i$ is the value of starting at state $x_0 = i$ if we are rewarded at time $t = k$
- $(v_k)_i = \mathbf{E}\big(g(x_k) \mid x_0 = i\big)$
- subscripts k, t denote times or iterations, so v_k is a vector
- subscripts i, j denote components, so v_i is the i'th component of v

Terminal costs

$$J = \lim_{t\to\infty} \mathbf{E}(g(x_t))$$

- we are rewarded when the state is absorbed
- we can evaluate J by *distribution propagation*

$$\begin{aligned} J &= \big(\lim_{t\to\infty} \pi_0 P^t\big) g \\ &= \pi_{\mathsf{ss}} g \end{aligned}$$

- π_{ss} gives probability distribution for where the state is absorbed

Terminal cost by value iteration

$$J = \lim_{t \to \infty} \mathbf{E}(g(x_t))$$

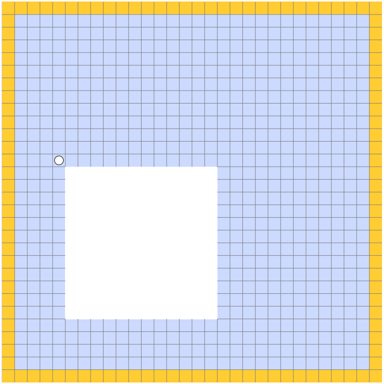
- alternatively, can evaluate J by *value iteration*

$$\begin{aligned} J &= \pi_0 \big(\lim_{t \to \infty} P^t g \big) \\ &= \pi_0 v_{\mathsf{ss}} \end{aligned}$$

- initialize $v_0 = g$ and iterate $v_{k+1} = P v_k$
- converges to steady state value $v_{\mathsf{ss}} = \lim_{k \to \infty} P^k g$ (if P^k converges)
- $(v_{\mathsf{ss}})_i$ gives value of starting in state $x_0 = i$

Example: random walk

- random walk on a 2-dimensional 30×30 grid, with square obstacle
- outer boundaries are absorbing
- boundary costs are 1, 2, 6, 10

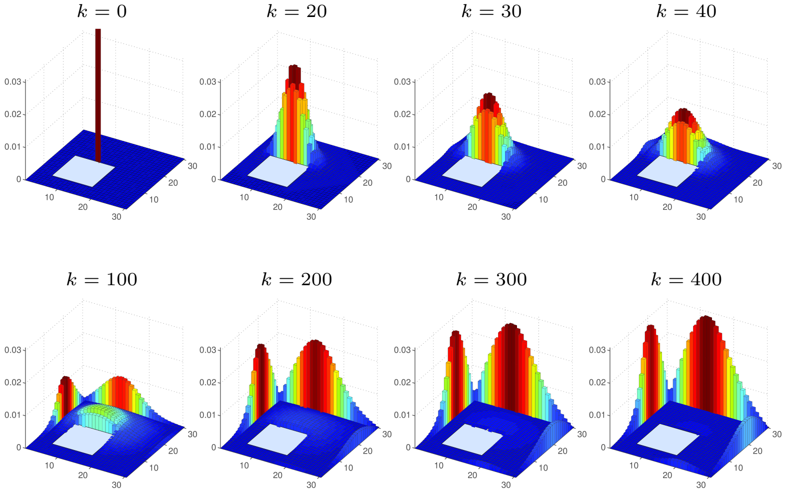


Transition probabilities

2 different cases:

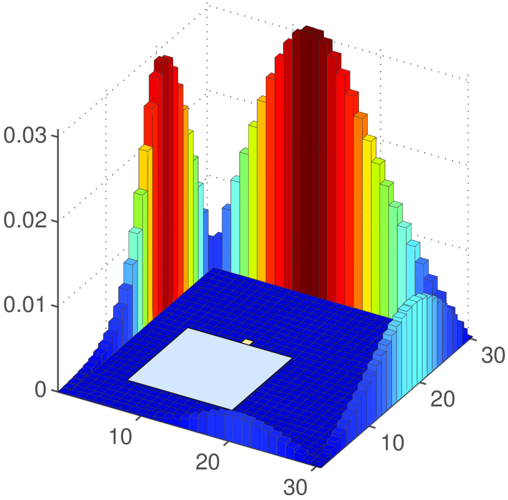
probability of staying at current state: $1/10$

Distribution propagation

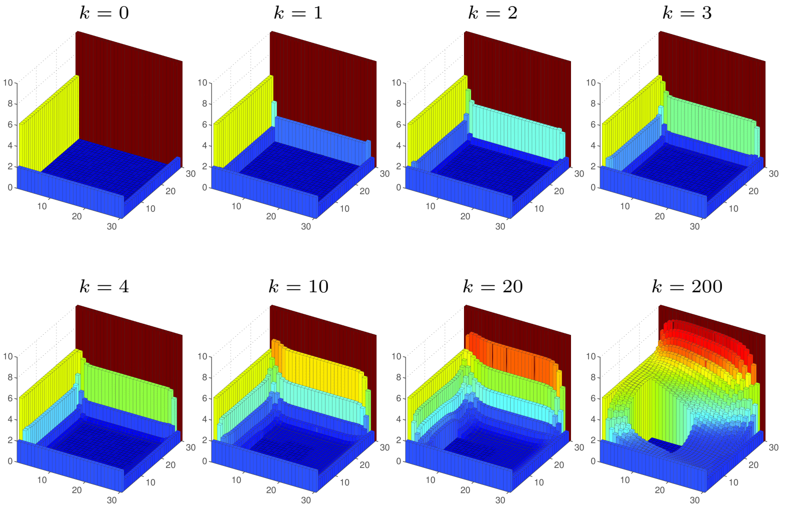


Steady state distribution

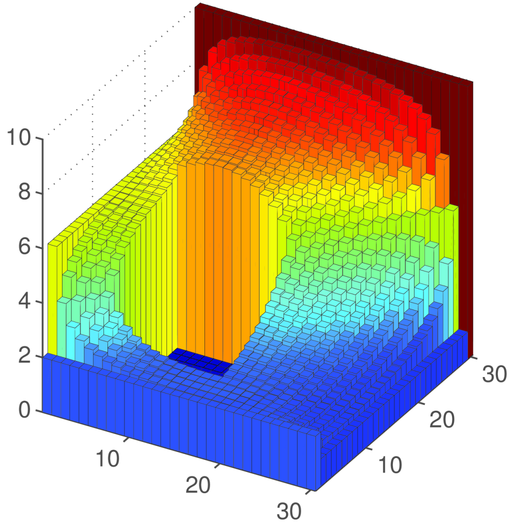
for the initial state $i = (12, 18)$, the ith row of L is below



Value function



Steady state value function



since $v_{\mathsf{ss}} = Pv_{\mathsf{ss}}$, the value function takes its max and min at the absorbing states

Harmonic functions

- suppose all closed classes are absorbing states, so $P_{22} = I$
- cost g is nonzero only on absorbing states
- steady-state value function v is unique solution to

$$\begin{aligned} v &= Pv \\ v_i &= g_i \qquad \text{if } i \text{ is absorbing} \end{aligned}$$

- the matrix $I - P$ is called the *discrete Laplacian*
- a function v satisfying $v = Pv$ is called a *harmonic function*
- called a *Dirichlet boundary value problem*