Review:

Shortest Path LP's

1) Enumerate Paths $\mathcal{R}$ (set of routes)

$R \in \mathbb{R}^{|E| \times |\mathcal{R}|}$ $\quad r \in \mathcal{R}$ particular route

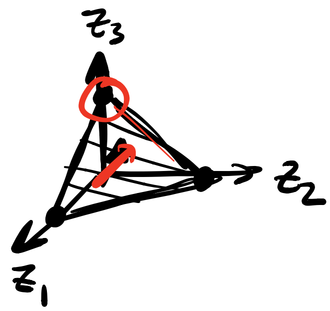
indicator matrix for which edges are in ea route

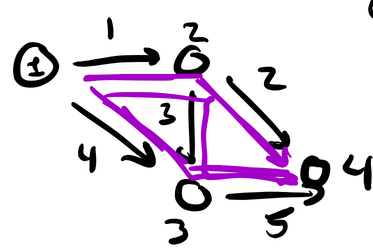
$$\begin{aligned} \min_{z \in \mathbb{R}^{|\mathcal{R}|}} \quad & l^T z \\ \text{s.t.} \quad & \mathbb{1}^T z = 1, \quad z \geq 0 \end{aligned}$$

$l^T = c^T R$ $\quad l_r$ cost of taking route r

$$l_r = \sum_{e \in r} c_e$$

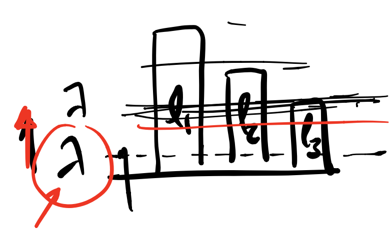
Pic





Dual

$$\begin{aligned} \max_{\lambda, u \in \mathbb{R}^{|\mathcal{R}|}} \quad & \lambda \\ \text{s.t.} \quad & \lambda \mathbb{1}^T = l^T - u^T, \quad u \geq 0 \end{aligned}$$



$\Rightarrow \boxed{\lambda \mathbb{1}^T \leq l^T}$ $\qquad \lambda \mathbb{1}^T = l^T - u^T$

$\lambda = l_r - u_r, \; u_r \geq 0 \iff \lambda \leq l_r$

2) Edge Formulation $\qquad x^* = \mathbb{R} z^*$

$$\min_{x \in \mathbb{R}^{|E|}} \; c^T x$$

$$\text{s.t.} \quad \underbrace{E}_{\text{incidence matrix}} x = \underbrace{b}_{\text{source sink vector}} \qquad x \geq 0$$

$$b = \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \begin{array}{l} \leftarrow \text{origin} \\ \\ \\ \\ \leftarrow \text{dest} \end{array}$$

Dual

$$\max_{\substack{v \in \mathbb{R}^{|S|} \\ \mu \in \mathbb{R}^{|E|}}} \; -v^T b$$

$$-v^T b = \underbrace{v_o}_{\text{cost to go at origin}} - \underbrace{v_d}_{\text{cost to go at dest.}}$$

$$\text{s.t.} \quad -v^T E = \underbrace{c^T - \mu^T}_{}, \quad \mu \geq 0$$

$$v^T E \leq c^T$$

V: value function
"cost-to-go" from ea. node
to destination

μ_e: inefficiency of taking
edge e

$$E = E_i - E_o$$

$$v^T E = v^T E_i - v^T E_o$$

$$v^T E = c^T - \mu^T$$

for ea. edge e… $\quad i \xrightarrow{e} j$

$$-(v_j - v_i) = c_e - \mu_e$$

$$v_i = c_e - \cancel{\mu_e} + v_j \longrightarrow v_i \le c_e + v_j$$

v_i should be a lower bound on the cost to go at node i

$$\boxed{x \ge 0} \quad \boxed{x < 0}$$

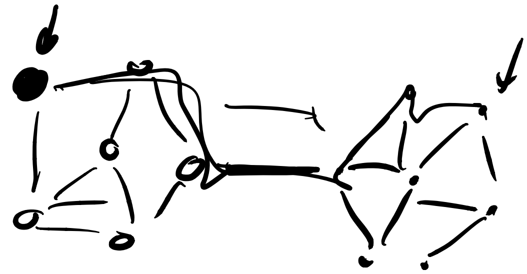
$$\left(\min_x \max_{v,\mu \ge 0} c^T x + v^T(Ex - b) - \mu^T x \leftarrow \quad -v^T b \right)$$

$$\max_v \min_x$$

$$c^T + v^T E - \mu^T = 0$$

$$Ex = b$$

↳ always an extra eqn.



$$\mathbb{1}^T (Ex = b)$$

$$\mathbb{1}^T E x = \mathbb{1}^T b$$

$$0x = 0$$

$[E|b]$ not full row rank ← extra row.

Fix problem:

$$U^{-1} U (Ex = b)$$

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} I & 0 \\ \mathbb{1}^T & 1 \end{bmatrix} \qquad U^{-1} = \begin{bmatrix} I & 0 \\ -\mathbb{1}^T & 1 \end{bmatrix}$$

$$U^{-1}(UEx = Ub)$$

$$\begin{bmatrix} I & 0 \\ & \mathbb{1}^T \end{bmatrix} E x = \begin{bmatrix} I & 0 \\ & \mathbb{1}^T \end{bmatrix} b$$

$$\begin{bmatrix} [I \; 0] E \\ \underline{\mathbb{1}^T E} \end{bmatrix} x = \begin{bmatrix} [I \; 0] E \\ 0 \end{bmatrix} x = \begin{bmatrix} [I \; 0] b \\ \underline{\mathbb{1}^T b} \end{bmatrix} = \begin{bmatrix} [I \; 0] b \\ 0 \end{bmatrix}$$

isolated the redundant constraint

$$\rightarrow \underbrace{[I \; 0] E}_{\text{all but last row of } E} x = \underbrace{[I \; 0] b}_{\text{all but last row of } b}$$

$$v^T(Ex - b)$$

$$v^T U^{-1}(\underline{U E x - U b})$$

$$v^T U^{-1}\left(\begin{bmatrix} [I \; 0] E \\ 0 \end{bmatrix} x - \begin{bmatrix} [I \; 0] b \\ 0 \end{bmatrix}\right)$$

$$v^T \underline{\begin{bmatrix} I & 0 \\ -\mathbb{1}^T & 1 \end{bmatrix}} \left(\begin{bmatrix} [I \; 0] E \\ 0 \end{bmatrix} x - \begin{bmatrix} [I \; 0] b \\ 0 \end{bmatrix}\right) \leftarrow$$

performing a coord transform on dual variable v

$\star$

$$v'^T = v^T \begin{bmatrix} I & 0 \\ -\mathbb{1}^T & 1 \end{bmatrix} =$$

$$\underline{v'^T} = [\underline{v_0 \cdots v_s} \cdots v_d] \begin{bmatrix} I & 0 \\ -\mathbb{1}^T & 1 \end{bmatrix} = [v_0 - v_d \quad v_s - v_d \quad v_d]$$

$$[\underline{v_0 - v_d} \cdots v_s - v_d \cdots \cancel{v_d}] \left(\begin{bmatrix} [I \; 0] E \\ \cancel{0} \end{bmatrix} x - \begin{bmatrix} [I \; 0] b \\ \cancel{0} \end{bmatrix}\right)$$

$$[v_0 - v_d \cdots v_{|S|-1} - v_d]\left([I \; 0] E x - [I \; 0] b\right)$$

$$\max_{v} \; -v^T b$$

$$\text{s.t.} \; -v^T E = c^T - \mu^T, \quad \mu \geq 0$$

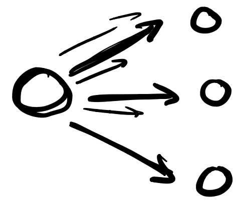
$$\begin{cases} \max_{v'} \; -v'^T \mathcal{U} b \Rightarrow v'_0 \\ \text{s.t.} \; -v'^T \mathcal{U} E = c^T - \mu^T, \quad \mu \geq 0 \end{cases}$$

→ maximize the cost-to-go from the origin

MARKOV DECISIONS PROCESSES (MDP)

flow problems w/ stochastic transitions

$G = (S, \mathcal{E})$ ←



Actions:

→ A_s: actions at state s

for $a \in A_s \implies \text{Prob}(s' | s, a) = \text{Prob}(s' | a)$ Transition kernel

probability of trans to state s' when action a is taken in state s

$$A = \bigcup_s A_s$$

↳ all actions

ea. action $a \in A$ implies being in a particular state s

→ actually represents a state action pair

- ea. action is only available from 1 state

Incidence Matrices

- $A \in \mathbb{R}^{|S| \times |A|}$

indicator matrix for which actions are available from which states

$$A_{sa} = \begin{cases} 1 & \text{if } a \in A_s \\ 0 & \text{otherwise} \end{cases}$$

- $P \in \mathbb{R}^{|S| \times |A|}$ Transitional kernel matrix

$P_{s'a} = \text{Prob}(s'|a)$

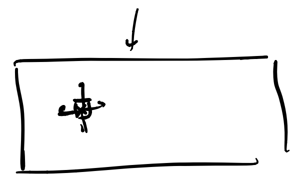
- $W \in \mathbb{R}^{|E| \times |A|}$ Trans kernel ...

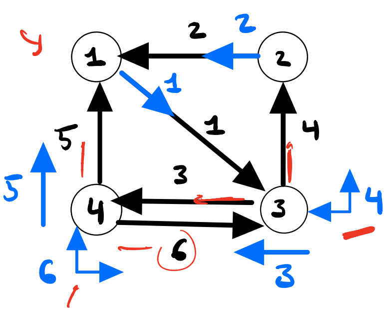
$$W_{ea} = \text{Prob}(e|a,s)$$

$$= \text{Prob}(e|a)$$

if edge e: $s \longrightarrow s'$

$\text{Prob}(e|a) = \text{Prob}(s'|a,s)$





$\mathcal{A}$: actions

$\mathcal{E}$: edges

Ex.

$$E_i = \begin{array}{c} \\ \\ 4 \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccc} & & 6 & & & \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$E_o = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$A = \begin{array}{c} \\ |S| \end{array} \begin{array}{c} |A| \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

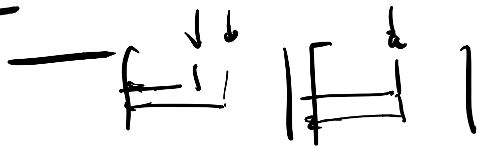
$$P = \begin{array}{c} \\ |S| \end{array} \begin{array}{c} |A| \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 1 & 0.5 & 0 & 0 \end{bmatrix} \end{array}$$

$$W = \begin{array}{c} \\ |E| \end{array} \begin{array}{c} |A| \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \boxed{0.5} & 0 & 0 \\ 0 & 0 & 0 & \boxed{0.5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.5 \end{bmatrix} \end{array}$$

$\left.\begin{array}{c} P \\ W \end{array}\right\}$ transition information

Relationships

$$\begin{array}{c c} & |A| \\ |S| & \left[\; P \;\right] \end{array} = \begin{array}{c c} & |E| \\ |S| & \left[\; E_i \;\right] \end{array} \begin{array}{c c} & |A| \\ |E| & \left[\; W \;\right] \end{array} \qquad A = E_o W$$



Before :

non stochastic mass conservation

$$x \in \mathbb{R}^{|E|} \to \boxed{E_i x = E_o x} \qquad Ex = 0$$

Now : $y \in \mathbb{R}^{|A|}$ stochastic flow vector

probability distribution over the actions

y_a : mass taking action a

$\varphi \in \mathbb{R}^{|S|}$ probability distribution over states

φ_s : mass in state s

for a particular distribution y over actions we can compute the corresponding state dist

$$\varphi = Ay \leftarrow$$ summing probabilities on all actions in ea state

steady stochastic flow .

$$\boxed{Py = \varphi = Ay} \Leftarrow$$

where that mass dist transitions to

mass distribution over states

$x = Wy$ dist on actions

corresponding dist on edges

$$Py = Ay \qquad P = E_i W \quad A = E_o W$$

$$\underline{E_i} W y = \underline{E_o W} y \Rightarrow (\underline{\underline{E_i - E_o}}) \underbrace{Wy}_{x} = 0$$

<u>Mass</u> <u>conservation</u>:

everything here is "column stochastic"

entries in ea. column sum to 1

$$\mathbb{1}^T_{|S|} E_i = \mathbb{1}^T_{|S|} E_o = \mathbb{1}^T_{|E|}$$

$$\mathbb{1}^T_{|S|} P = \mathbb{1}^T_{|S|} A = \mathbb{1}^T_{|A|}$$

$$\underline{\mathbb{1}^T_{|E|} W = \mathbb{1}^T_{|A|}}$$

$$\underline{\mathbb{1}^T_{|A|} y = 1}, \quad \mathbb{1}^T_{|S|} \varphi = 1, \quad \mathbb{1}^T_{|E|} x = 1$$

$$\varphi = Ay \qquad \underline{\mathbb{1}^T_{|S|}} \varphi = \mathbb{1}^T_{|S|} A y = \underline{\mathbb{1}^T_{|A|} y} = 1$$

Rewards $r \in \mathbb{R}^{|A|}$ r_a : reward for taking action a

FINITE HORIZON MDP LP: $x^TQx + u^TRu$

PRIMAL

$$\max_{y(t),\ t=0,\dots,T} \sum_{t=0}^{T-1} r(t)^T y(t) + g^T A y(T)$$

$y(t)$: mass dist over actions at time t

$r(t)$: rewards at time t

g_s: final reward at state s

$x^+ = Ax + Bu$

s.t.

$$A y(0) = \varphi(0)$$

initial state distribution (given)

mass cons. over time

$$A y(t+1) = \boxed{P} y(t) \quad t = 0, \dots, T-1$$

$$y(t) \geq 0$$

$$\varphi(t+1) = P y(t)$$

Dual Variables

: $A y(0) = \varphi(0) \Rightarrow v(0) \in \mathbb{R}^{|S|}$

$A y(t+1) = P y(t) \Rightarrow v(t+1) \in \mathbb{R}^{|S|}$

$y(t) \geq 0 \Rightarrow \mu(t) \in \mathbb{R}_+^{|A|}$

Lagrangian

$$\mathcal{L}(y, v, \mu) = \sum_{t=0}^{T-1} r(t)^T y(t) + g^T A y(T)$$

$$- v(0)^T (A y(0) - \varphi(0)) - \sum_{t=0}^{T-1} v(t+1)^T (A y(t+1) - P y(t)) + \sum_{t=0}^{T} \mu(t)^T y(t)$$

$$\max_{y} \min_{v,\mu \geq 0} \mathcal{L} \leq \min_{v,\mu \geq 0} \max_{y} \mathcal{L}$$

$$\frac{\partial \mathcal{L}}{\partial y} = 0$$

$$\frac{\partial \mathcal{L}}{\partial y(t)} = r(t)^T + v(t+1)^T P - v(t)^T A + \mu(t)^T = 0$$

$t = 0, \dots, T-1$

$$\frac{\partial \mathcal{L}}{\partial y(T)} = g^T A - v(T)^T A + \mu(T)^T = 0$$

Dual.

$$\min_{v,\mu} \quad v(0)^T \varphi(0) \rightarrow$$

"expected total reward if our initial mass dist is $\varphi(0)$"

s.t. $v(T)^T A = g^T A + \mu(T)^T$, $\mu(T) \geq 0$

$v(t)^T A = r(t)^T + v(t+1)^T P + \mu(t)^T$ $\quad t = 0, \dots, T-1$

$\mu(t) \geq 0$

Bellman Eqn.

elementwise

$$v_s(t) = r_a(t) + v(t+1)^T P[:, a] + \mu_a(t)$$

value at next time

$$V_s(t) = r_a(t) + \underbrace{v(t+1)^T P[:,a]}_{} + \mu_a(t)$$

$V_s(t)$: value function at state s time t

$r_a(t)$: reward for taking action a

$q_a(t+1)$ = "expected reward to go for taking action a"

$\mu_a(t)$: inefficiency of action a

$$v(t+1)^T P := q(t+1)^T$$

"Q-value from Q learning"

$$q(t+1) \in \mathbb{R}^{|A|}$$

$$V_s(t) \geq r_a(t) + v(t+1)^T P[:,a]$$

$$V_s(T) = g_s + \mu_a(t)$$

$$V_s(T) \geq g_s$$

$$v^T A = q^T$$

RL / Q learning

$$v^T A = r^T + q^T + \mu^T$$

Infinite Horizon - Average Reward MDP LP

assume steady state...

$y \in \mathbb{R}^{|A|}$: steady state distribution

$r \in \mathbb{R}^{|A|}$: reward vector

$$\boxed{\begin{aligned} &\max_{y \in \mathbb{R}^{|A|}} \ r^T y \rightarrow \text{average expected reward} \\ &\text{s.t.} \quad Ay = Py, \ \mathbb{1}^T y = 1, \ y \geq 0 \end{aligned}}$$

$$E_0 W_y = E_i W_y$$

$$(E_i - E_0) W_y = 0$$

Dual variables:

$Ay = Py$	$v \in \mathbb{R}^{\|S\|}$
$\mathbb{1}^T y = 1$	$\lambda \in \mathbb{R}$
$y \geq 0$	$\mu \in \mathbb{R}_+^{\|A\|}$

$$\min_{v, \lambda, \mu} \ \lambda$$

$$\text{s.t.} \quad \lambda \mathbb{1}^T + v^T A = r^T + v^T P + \mu^T$$

$$\mu \geq 0$$

steady state Bellman eqn

λ: overall average cost

v: "value function" tells you how much ea. r_a differs from λ

μ_a: inefficiency of ea. action

$\mu_a y_a = 0$

Note: special conditions on P required for this average reward formulation. * see below

$\Rightarrow$ every choice of actions results in a steady state distribution

Connections w/ Markov Chains

normally solving an MDP selecting a "policy"

policy: mixed strategy at ea. state
(feedback)

pick probability distribution over actions

$$\pi_s \in \mathbb{R}^{|A_s|} \qquad \mathbb{1}^T \pi_s = 1 \qquad \pi_s \geq 0 \qquad (A\pi = I)$$

$$(\pi_s)_a = \text{Prob}(a|s) \qquad \varphi = Ay$$

$$(\pi_s)_a = \frac{y_a}{\varphi_s} = \frac{y_a}{\sum_{a \in A_s} y_a}$$

$\Pi \in \mathbb{R}^{|A| \times |S|}$ $\quad \Pi = \begin{bmatrix} \underline{\pi_1} & & 0 \\ & \ddots & \\ 0 & & \pi_{|S|} \end{bmatrix}$

state distribution φ:

$$\underset{\downarrow}{y} = \Pi \underline{\varphi} = \begin{bmatrix} \pi_1 & & \\ & \ddots & \\ & & \pi_{|S|} \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_{|S|} \end{bmatrix} = \begin{bmatrix} \pi_1 \varphi_1 \\ \vdots \\ \pi_{|S|} \varphi_{|S|} \end{bmatrix}$$

Markov Chain:

$$M \in \mathbb{R}^{|S| \times |S|} \qquad M_{s's} = \text{Prob}(s' \mid s)$$

steady state dist:

$$\varphi = M\varphi \implies \varphi \text{ steady state state dist.}$$

when you pick π or Π $\rightarrow$ selecting Markov chain

$$M = P_{|A|} \, \Pi_{|S|} \qquad I = \underline{A\Pi}$$

$$|S| \begin{bmatrix} P \end{bmatrix}_{|A|} \begin{bmatrix} \quad \end{bmatrix}$$

$$\underline{Ay = Py} \implies \underline{A\Pi\varphi} = \underline{P\Pi\varphi} \implies \underline{\varphi = M\varphi}$$

$\varphi = M\varphi \Rightarrow \varphi$ is an right eigenvector of M with eigenvalue 1

how do we know φ exists...

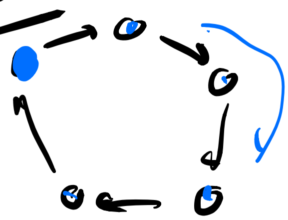
$\mathbb{1}^T M = \mathbb{1}^T \longrightarrow \mathbb{1}^T$ is a left eigenvector of M

★ For any Π:

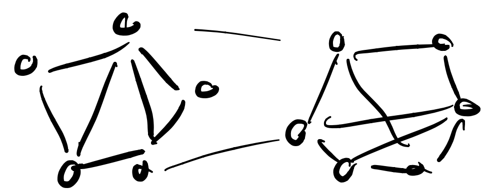
the resulting Markov matrix $M = P\Pi$ is aperiodic & irreducible ☆

$\rightarrow$ unique solution to $\varphi = M\varphi$

Periodic



Not irreducible



Perron Frobenius Thm

$$y \iff \varphi, \Pi$$

$$- \; y \implies \varphi = Ay$$

$$(\Pi_s)_a = y_a / \varphi_s$$

$$\Pi = dg(y) A^T dg(\varphi)^{-1}$$

$$dg(z) = \begin{bmatrix} z_1 & & 0 \\ & \ddots & \\ 0 & & z_n \end{bmatrix}$$

$$\Longleftarrow$$

$$y = \Pi \varphi$$

Discounted Infinite Horizon MDP LP

γ: discount factor $0 \le \gamma \le 1$

$$\max_y \quad r^T y$$
$$\text{s.t.} \quad Ay = \gamma P y + (1-\gamma)\varphi(0) \quad \cancel{\mathbb{1}^T y = 1} \quad y \ge 0$$

$$Ay = Py \quad \mathbb{1}^T y = 1 \quad y \ge 0$$

$$\min_v \quad (1-\gamma) v^T \varphi(0)$$ expected discounted reward

$$\text{s.t.} \quad v^T A \ge \underbrace{r^T + \gamma v^T P}_{}$$

discounted Bellman equation

always solvable for any P.