value iteration

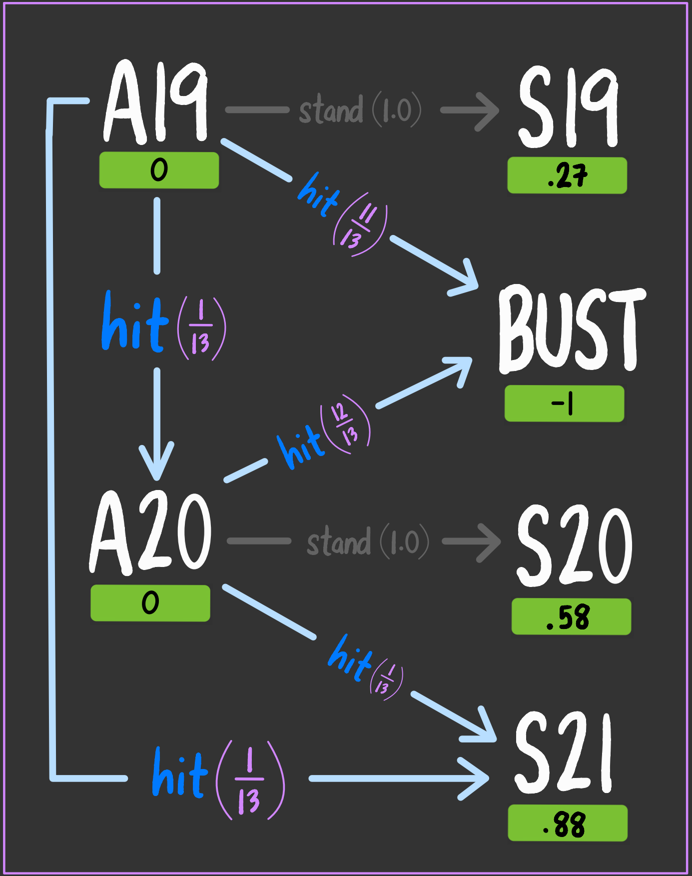
31 oct 2022

CSCI 373

last time, we defined the expected utility of being in state q, given policy π

$$U^{\pi}(q) = \sum_{path} U(q \xrightarrow{\pi(q)} path)\, P(q \xrightarrow{\pi(q)} path)$$

$U^{\pi}(A19) = -.84$

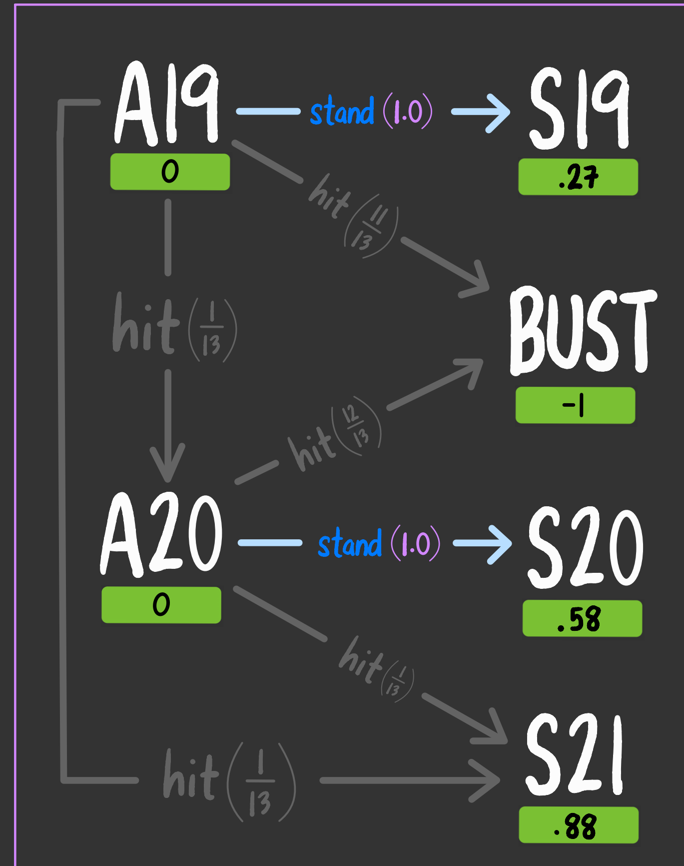


markov decision process

in order to make decisions, we'd like to know the policy that maximizes our expected utility

$$\pi^* = \arg\max_{\pi} U^{\pi}(q)$$

$$\pi^* = \{A19 \mapsto \text{stand},\ A20 \mapsto \text{stand}\}$$

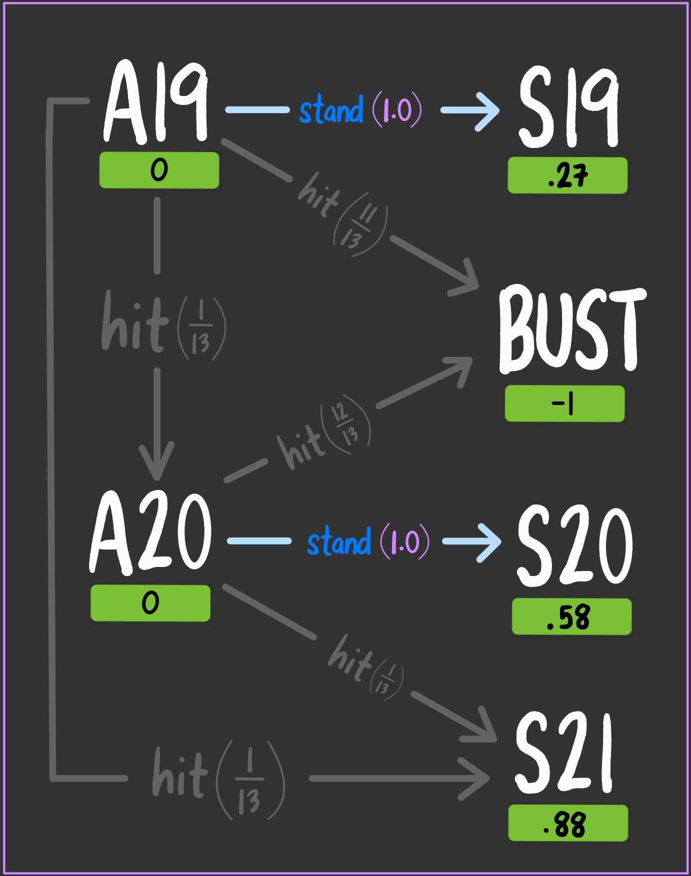


markov decision process

but first we'll focus on how to compute the **maximum expected utility**

$$U(q) = \max_{\pi} U^{\pi}(q)$$

$$U(A19) = .27$$

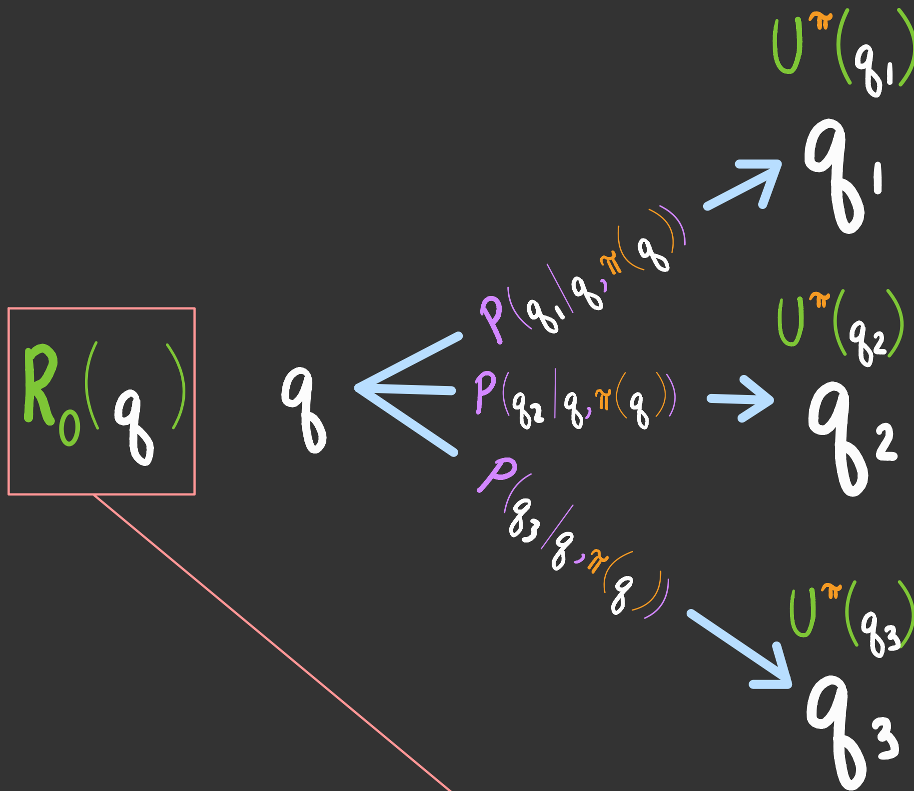


markov decision process

want: maximum expected utility $U(q) = \max\limits_{\pi} U^{\pi}(q)$

last time, we observed:

$$\overbrace{U^{\pi}(q)}^{\text{the expected utility of policy } \pi \text{ in state } q} = \underbrace{R_0(q)}_{\text{the immediate reward of being in state } q} + \overbrace{\gamma}^{\text{discounting factor}} \sum_{q' \in Q} \underbrace{P(q' \mid q, \pi(q))}_{\text{the probability of getting from state } q \text{ to state } q' \text{ using the action advised by the policy}} \overbrace{U^{\pi}(q')}^{\text{the expected utility of policy } \pi \text{ in state } q'}$$



$$+\gamma P(q' \mid q, \pi(q))\, U^\pi(q')$$

$$+\gamma P(q' \mid q, \pi(q))\, U^\pi(q')$$

$$+\gamma P(q' \mid q, \pi(q))\, U^\pi(q')$$

$$U^\pi(q) = \boxed{R_0(q)} + \boxed{\gamma \sum_{q' \in Q} P(q' \mid q, \pi(q))\, U^\pi(q')}$$

$$U(q) = \max_{\pi} U^{\pi}(q)$$

want: maximum expected utility $U(q) = \max_{\pi} U^{\pi}(q)$

$$U(q) = \max_{\pi} U^{\pi}(q)$$

$$= \max_{\pi} R_0(q) + \gamma \sum_{q' \in Q} P(q' \mid q, \pi(q)) U^{\pi}(q')$$

last time, we observed:

$$U^{\pi}(q) = R_0(q) + \gamma \sum_{q' \in Q} P(q' \mid q, \pi(q)) U^{\pi}(q')$$

$$U(q) = \max_{\pi} U^{\pi}(q)$$

$$= \max_{\pi} R_0(q) + \gamma \sum_{q' \in Q} P(q' \mid q, \pi(q))\, U^{\pi}(q')$$

$$= R_0(q) + \gamma \max_{\pi} \sum_{q' \in Q} P(q' \mid q, \pi(q))\, U^{\pi}(q')$$

always nonnegative

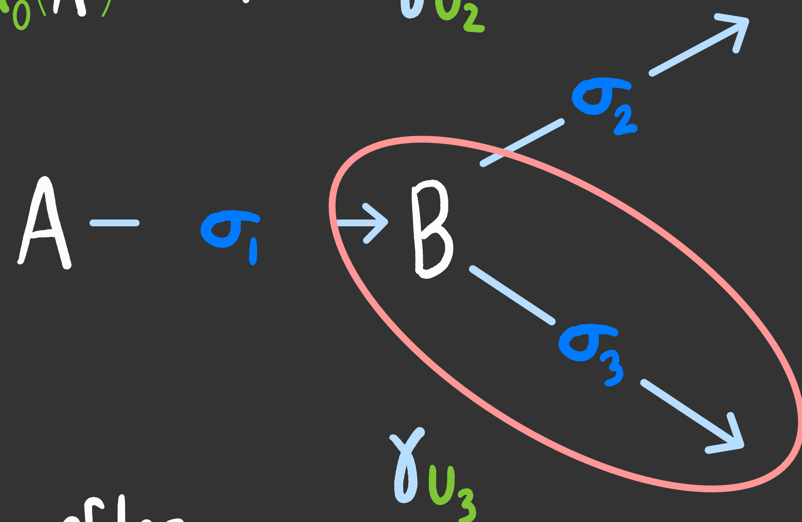
$$\max_{\pi} U^{\pi}(A) = R_0(A) + \gamma U_2$$

$$\begin{tikzcd} & & \cdot \\ A \arrow[r, "\sigma_1"] & B \arrow[ur, "\sigma_2"] \arrow[dr, "\sigma_3"'] & \\ & & \cdot \end{tikzcd}$$

suppose:

in state A, after going to state B, the best action is σ_2

$$\max_{\pi} U^{\pi}(A) = R_0(A) + \gamma U_2$$



this means $U_2 > U_3$

suppose:

in state A, after going to state B, the best action is σ_2

$$\max_{\pi} U^{\pi}(A) = R_0(A) + \gamma U_2$$

$$\begin{tikzcd}
 & & {} \\
A \arrow[r, "\sigma_1"] & B \arrow[ur, "\sigma_2"] \arrow[dr, "\sigma_3"'] & \\
 & & {}
\end{tikzcd}$$

this means

$U_2 > U_3$

suppose:

in state A, after going to state B, the best action is σ_2

so the best action in state B regardless of our origin is σ_2

$$
\begin{aligned}
U(q) &= \max_{\pi} U^{\pi}(q) \\
&= \max_{\pi} \; R_0(q) + \gamma \sum_{q' \in Q} P(q' \mid q, \pi(q))\, U^{\pi}(q') \\
&= R_0(q) + \gamma \max_{\pi} \sum_{q' \in Q} P(q' \mid q, \pi(q))\, U^{\pi}(q') \\
&= R_0(q) + \gamma \max_{\pi} \sum_{q' \in Q} P(q' \mid q, \pi(q)) \max_{\pi'} U^{\pi'}(q') \\
&= R_0(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q')
\end{aligned}
$$

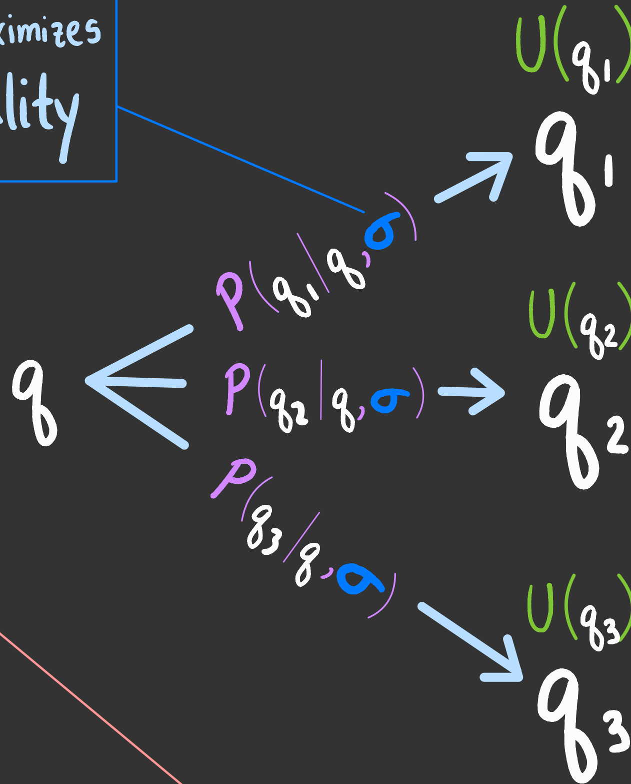
we obtain something called a

bellman equation

$$\underbrace{U(q)}_{\text{the maximum expected utility of state } q} = \underbrace{R(q)}_{\text{the immediate reward of being in state } q} + \underbrace{\gamma}_{\text{discounting factor}} \max_{\sigma} \sum_{q' \in Q} \underbrace{P(q' \mid q, \sigma)}_{\text{the probability of getting from state } q \text{ to state } q' \text{ if we do action } \sigma} \underbrace{U(q')}_{\text{the maximum expected utility of state } q'}$$

the action that maximizes **expected utility**

$R_0(q)$

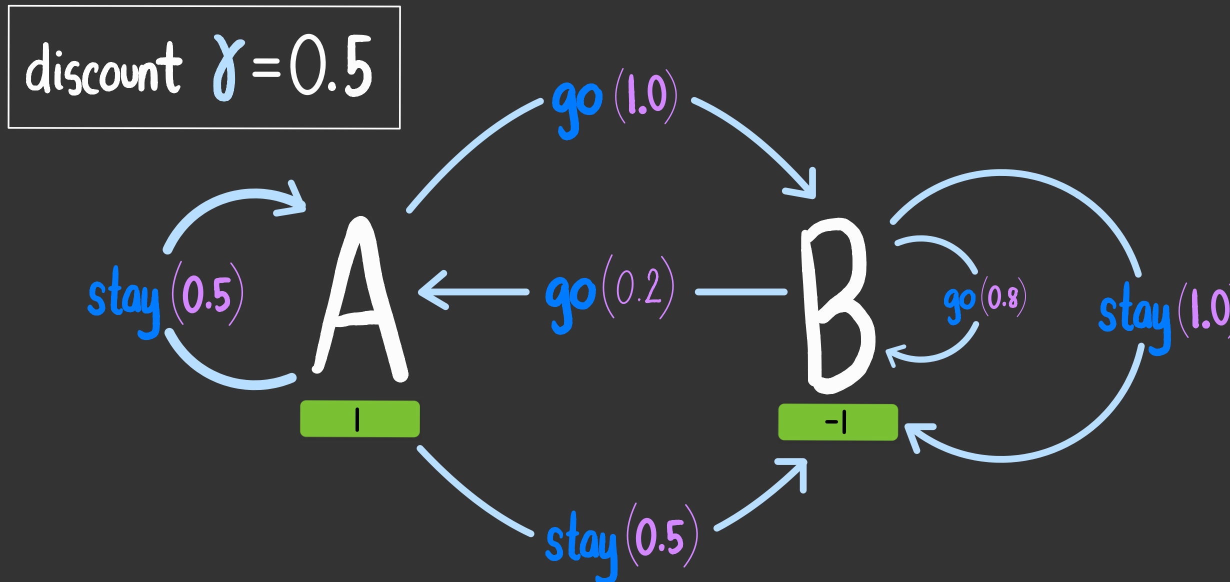


$$+ \gamma P(q_1 | q, \sigma) U(q_1)$$

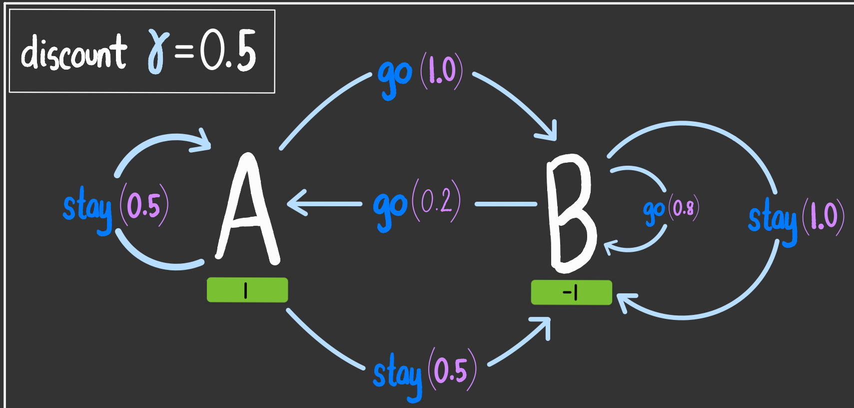
$$+ \gamma P(q_2 | q, \sigma) U(q_2)$$

$$+ \gamma P(q_3 | q, \sigma) U(q_3)$$

$$U(q) = \boxed{R_0(q)} + \boxed{\max_{\sigma} \gamma \sum_{q' \in Q} P(q' | q, \sigma) U(q')}$$



Consider the bellman equations for the above markov decision process

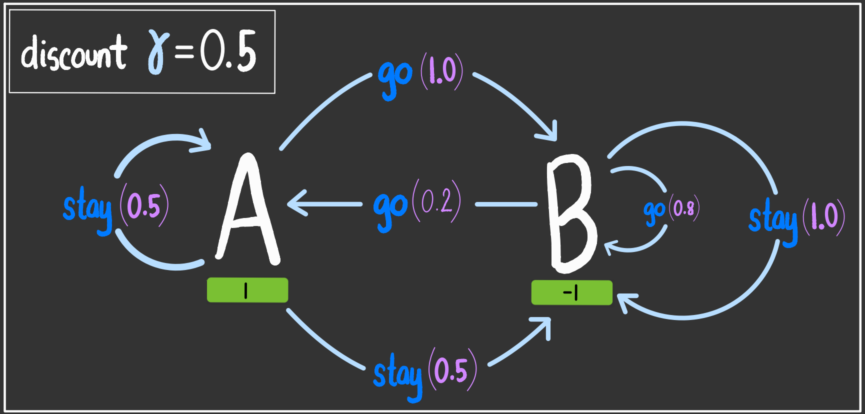


bellman equation:

$$U(q) = R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q')$$

$$U(A) = \ ?$$

$$U(B) = \ ?$$

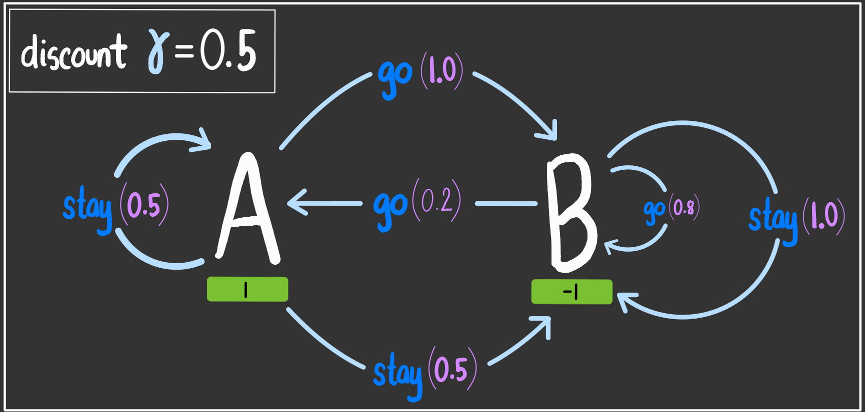


bellman equation:

$$U(q) = R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q')$$

$$U(A) = R(A) + \gamma \max \begin{Bmatrix} P(A \mid A, \text{stay})\, U(A) + P(B \mid A, \text{stay})\, U(B) \\ P(A \mid A, \text{go})\, U(A) + P(B \mid A, \text{go})\, U(B) \end{Bmatrix}$$

$$U(B) = R(B) + \gamma \max \begin{Bmatrix} P(A \mid B, \text{stay})\, U(A) + P(B \mid B, \text{stay})\, U(B) \\ P(A \mid B, \text{go})\, U(A) + P(B \mid B, \text{go})\, U(B) \end{Bmatrix}$$

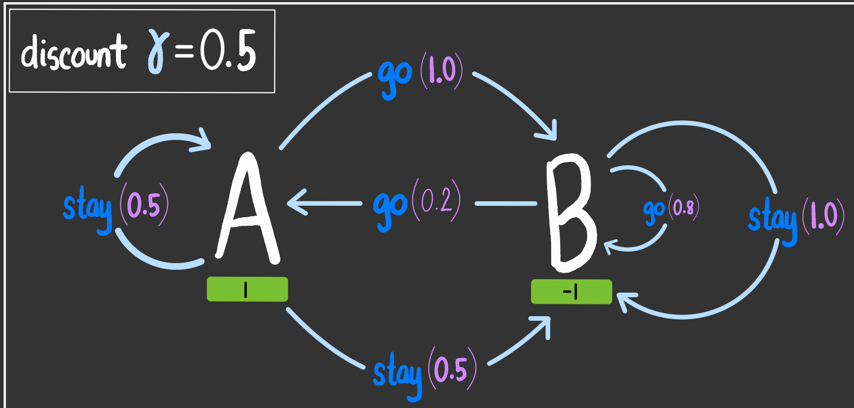


bellman equation:

$$U(q) = R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q')$$

$$U(A) = 1 + 0.5 \max \begin{Bmatrix} 0.5\, U(A) + 0.5\, U(B) \\ 0.0\, U(A) + 1.0\, U(B) \end{Bmatrix}$$

$$U(B) = -1 + 0.5 \max \begin{Bmatrix} 0.0\, U(A) + 1.0\, U(B) \\ 0.2\, U(A) + 0.8\, U(B) \end{Bmatrix}$$



bellman equation:

$$U(q) = R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q')$$

$$U(A) = 1 + 0.5 \max \{ U(B),\ 0.5\, U(A) + 0.5\, U(B) \}$$

$$U(B) = -1 + 0.5 \max \{ U(B),\ 0.2\, U(A) + 0.8\, U(B) \}$$

this is a system of

two equations with two unknowns

$$U(A) = 1 + 0.5 \max\left\{U(B),\ 0.5\,U(A) + 0.5\,U(B)\right\}$$

$$U(B) = -1 + 0.5 \max\left\{U(B),\ 0.2\,U(A) + 0.8\,U(B)\right\}$$

this is a system of

two equations with two unknowns

but they aren't linear equations

$$U(A) = 1 + 0.5 \max\{U(B),\ 0.5\,U(A) + 0.5\,U(B)\}$$

$$U(B) = -1 + 0.5 \max\{U(B),\ 0.2\,U(A) + 0.8\,U(B)\}$$

an iterative strategy

- guess values for the unknowns
  - u_0^A is a guess for $U(A)$
  - u_0^B is a guess for $U(B)$
- compute new guesses using the equations
- repeat

$$U(A) = 1 + 0.5 \max \left\{ \overset{u_0^B}{\cancel{U(B)}},\ 0.5\, \overset{u_0^A}{\cancel{U(A)}} + 0.5\, \overset{u_0^B}{\cancel{U(B)}} \right\}$$

$$U(B) = -1 + 0.5 \max \left\{ \overset{u_0^B}{\cancel{U(B)}},\ 0.2\, \overset{u_0^A}{\cancel{U(A)}} + 0.8\, \overset{u_0^B}{\cancel{U(B)}} \right\}$$

an iterative strategy

- guess values for the unknowns
  - u_0^A is a guess for $U(A)$
  - u_0^B is a guess for $U(B)$
- compute new guesses using the equations
- repeat

$$U(A) = 1 + 0.5 \max\left\{ u_0^B ,\ 0.5\, u_0^A + 0.5\, u_0^B \right\}$$

$$U(B) = -1 + 0.5 \max\left\{ u_0^B ,\ 0.2\, u_0^A + 0.8\, u_0^B \right\}$$

an iterative strategy

- guess values for the unknowns
  - u_0^A is a guess for $U(A)$
  - u_0^B is a guess for $U(B)$
- compute new guesses using the equations
- repeat

$$\overset{u_1^A}{\cancel{U(A)}} \leftarrow 1 + 0.5 \max\left\{ u_0^B \, , \, 0.5\, u_0^A + 0.5\, u_0^B \right\}$$

$$\overset{u_1^B}{\cancel{U(B)}} \leftarrow -1 + 0.5 \max\left\{ u_0^B \, , \, 0.2\, u_0^A + 0.8\, u_0^B \right\}$$

an iterative strategy

- guess values for the unknowns
  - u_0^A is a guess for $U(A)$
  - u_0^B is a guess for $U(B)$
- compute new guesses using the equations
- repeat

$$u_1^A \leftarrow 1 + 0.5 \max \left\{ u_0^B \,,\; 0.5\, u_0^A + 0.5\, u_0^B \right\}$$

$$u_1^B \leftarrow -1 + 0.5 \max \left\{ u_0^B \,,\; 0.2\, u_0^A + 0.8\, u_0^B \right\}$$

an iterative strategy

- guess values for the unknowns
  - U_0^A is a guess for $U(A)$
  - U_0^B is a guess for $U(B)$
- compute new guesses using the equations
- repeat

$$U_{t+1}^A \leftarrow 1 + 0.5 \max \left\{ U_t^B \ , \ 0.5\, U_t^A + 0.5\, U_t^B \right\}$$

$$U_{t+1}^B \leftarrow -1 + 0.5 \max \left\{ U_t^B \ , \ 0.2\, U_t^A + 0.8\, U_t^B \right\}$$

to the laptop!

DISCOUNT =	0.5

t	U(A)	U(B)
0	1.2500	-1.1300
1	1.6250	-1.3270
2	1.8125	-1.3683
3	1.9063	-1.3661
4	1.9531	-1.3558
5	1.9766	-1.3470
6	1.9883	-1.3411
7	1.9941	-1.3376
8	1.9971	-1.3356
9	1.9985	-1.3345
10	1.9993	-1.3340
11	1.9996	-1.3337
12	1.9998	-1.3335
13	1.9999	-1.3334
14	2.0000	-1.3334
15	2.0000	-1.3334

value iteration

- guess values for the unknowns

  for each state q, U_0^q is a guess for $U(q)$

- for $t = 1$ to T:

  compute new guesses using the bellman equations, i.e.

  for each state q, $U_{t+1}^q \leftarrow R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U_t^{q'}$

Claim: if $\gamma < 1$, then value iteration converges to the maximum expected utilities, i.e. $\lim_{t \to \infty} U_t^q = U(q)$

Claim: if $\gamma < 1$, then value iteration converges to the maximum expected utilities, i.e. $\lim_{t \to \infty} U_t^q = U(q)$

1. assume a solution to the bellman equations exists, i.e there exists a function $U : Q \to \mathbb{R}$ such that

$$U(q) = R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q) \quad \text{for every state } q$$

this is the solution to the bellman equations $\longrightarrow$ $\begin{bmatrix} U(q_0) \\ \\ U(q_n) \end{bmatrix}$

Claim: if $\gamma < 1$, then value iteration converges to the maximum expected utilities, i.e. $\lim_{t \to \infty} U_t^q = U(q)$

2. measure how far our guesses at iteration t are from the solution using the following distance function:

$$\text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right) = \max_{q \in Q} \left| U_t^q - U(q) \right|$$

$$\mathrm{dist}\left(\begin{array}{c} \text{guesses at} \\ \text{time } t \\ \begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix} \end{array}, \begin{array}{c} \text{true} \\ \text{utilities} \\ \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \end{array}\right) = \max_{q \in Q} \left| U_t^q - U(q) \right|$$

e.g.

$$\mathrm{dist}\left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix}\right) = ?$$

$$\operatorname{dist}\left(\begin{array}{cc} \text{guesses at time } t & \text{true utilities} \\ \begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, & \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \end{array} \right) = \max_{q \in Q} \left| U_t^q - U(q) \right|$$

e.g.

$$\operatorname{dist}\left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix} \right) = \max \left\{ |2-5|, |4-3| \right\} = 3$$

guesses at time t   true utilities

$$\mathrm{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right) = \max \begin{bmatrix} \left| U_t^{q_0} - U(q_0) \right| \\ \vdots \\ \left| U_t^{q_n} - U(q_n) \right| \end{bmatrix}$$

equals zero iff our guesses are all correct

e.g.

$$\mathrm{dist}\left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix} \right) = \max \left\{ |2-2|, |4-4| \right\} = 0$$

Claim: if $\gamma < 1$, then value iteration converges to the maximum expected utilities, i.e. $\lim_{t \to \infty} U_t^q = U(q)$

suppose we can show that every step of value iteration brings our guesses **closer** to the maximum expected utilities

$$\text{dist}\left(\begin{bmatrix} U_{t+1}^{q_0} \\ \vdots \\ U_{t+1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right) < K \, \text{dist}\left(\begin{bmatrix} U_{t}^{q_0} \\ \vdots \\ U_{t}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right)$$

for some fraction $K \in [0, 1)$

suppose for some fraction $K \in [0, 1)$:

$$\mathrm{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right) \leqslant K \, \mathrm{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right)$$

suppose for some fraction $K \in [0, 1)$:

$$\mathrm{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant K\,\mathrm{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

then :

$$\mathrm{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant K^t\,\mathrm{dist}\left(\begin{bmatrix} U_0^{q_0} \\ \vdots \\ U_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

$$\boxed{\mathrm{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant K\,\mathrm{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant K \cdot K\,\mathrm{dist}\left(\begin{bmatrix} U_{t-2}^{q_0} \\ \vdots \\ U_{t-2}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant \cdots}$$

suppose for some fraction $K \in [0, 1)$:

$$\text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq K \,\text{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

then :

$$\text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq K^t \,\text{dist}\left(\begin{bmatrix} U_0^{q_0} \\ \vdots \\ U_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

$$\lim_{t \to \infty} \text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq \lim_{t \to \infty} K^t \,\text{dist}\left(\begin{bmatrix} U_0^{q_0} \\ \vdots \\ U_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

suppose for some fraction $K \in [0, 1)$:

$$\text{dist}\left(\begin{bmatrix} u_t^{q_0} \\ \vdots \\ u_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant K \,\text{dist}\left(\begin{bmatrix} u_{t-1}^{q_0} \\ \vdots \\ u_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

then :

$$\text{dist}\left(\begin{bmatrix} u_t^{q_0} \\ \vdots \\ u_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant K^t \,\text{dist}\left(\begin{bmatrix} u_0^{q_0} \\ \vdots \\ u_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

$$\lim_{t \to \infty} \text{dist}\left(\begin{bmatrix} u_t^{q_0} \\ \vdots \\ u_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant \underbrace{\text{dist}\left(\begin{bmatrix} u_0^{q_0} \\ \vdots \\ u_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)}_{\text{doesn't depend on } t} \lim_{t \to \infty} K^t$$

suppose for some fraction $K \in [0, 1)$:

$$\operatorname{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq K \operatorname{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

then :

$$\operatorname{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq K^t \operatorname{dist}\left(\begin{bmatrix} U_0^{q_0} \\ \vdots \\ U_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

$$\lim_{t \to \infty} \operatorname{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq \operatorname{dist}\left(\begin{bmatrix} U_0^{q_0} \\ \vdots \\ U_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \lim_{t \to \infty} K^t$$

suppose for some fraction $K \in [0, 1)$:

$$\text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq K \, \text{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

then:

$$\text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq K^t \, \text{dist}\left(\begin{bmatrix} U_0^{q_0} \\ \vdots \\ U_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

$$\lim_{t \to \infty} \text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq 0$$

suppose for some fraction $K \in [0, 1)$:

$$\mathrm{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right) \leq K \,\mathrm{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right)$$

then :

$$\mathrm{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right) \leq K^t \,\mathrm{dist}\left(\begin{bmatrix} U_0^{q_0} \\ \vdots \\ U_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right)$$

$$\lim_{t \to \infty} \underbrace{\mathrm{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix} \right)}_{\text{always nonnegative } \left(\max_{q \in Q} \left| U_t^q - U(q) \right| \right)} \leq 0$$

suppose for some fraction $K \in [0,1)$:

$$\mathrm{dist}\left(\begin{bmatrix} u_t^{q_0} \\ \vdots \\ u_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant K\, \mathrm{dist}\left(\begin{bmatrix} u_{t-1}^{q_0} \\ \vdots \\ u_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

then :

$$\mathrm{dist}\left(\begin{bmatrix} u_t^{q_0} \\ \vdots \\ u_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant K^t\, \mathrm{dist}\left(\begin{bmatrix} u_0^{q_0} \\ \vdots \\ u_0^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

$$\lim_{t\to\infty} \mathrm{dist}\left(\begin{bmatrix} u_t^{q_0} \\ \vdots \\ u_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) = 0$$

$$\text{if } \operatorname{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leqslant K \operatorname{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right), \quad \text{then: } \lim_{t\to\infty} \operatorname{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) = 0$$

for some fraction $K \in [0, 1)$

if we can show that every step of value iteration brings our guesses closer to the maximum expected utilities

then value iteration converges to the maximum expected utilities if $\text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq K \,\text{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$,

for some fraction $K \in [0, 1)$

then: $\lim_{t \to \infty} \text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) = 0$

if we can show that every step of value iteration brings our guesses closer to the maximum expected utilities

so let's show this

then value iteration converges to the maximum expected utilities want to show:

$$\text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq K \, \text{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right)$$

for some fraction $K \in [0, 1)$

$$\text{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) = \max_{q \in Q} \left| U_t^q - U(q) \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| =$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| =$$

$$U_{t+1}^{q} = R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U_{t}^{q'}$$

new guess at iteration t+1

$$U(q) = R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U(q')$$

bellman equation

$$\boxed{\text{want to show}: \max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_t^{q} - U(q) \right| \quad \text{for some fraction } K \in [0, 1)}$$

$$\left| U_{t+1}^{q} - U(q) \right| = \left| \left(R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' | q, \sigma) U_t^{q'} \right) - \left(R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' | q, \sigma) U(q') \right) \right|$$

want to show:
$$\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leq K \max_{q \in Q} \left| U_t^{q} - U(q) \right|$$
for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| = \left| \left(\cancel{R(q)} + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U_t^{q'} \right) - \left(\cancel{R(q)} + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q') \right) \right|$$

$$= \left| \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U_t^{q'} - \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q') \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| = \left| \left(\cancel{R(q)} + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U_{t}^{q'} \right) - \left(\cancel{R(q)} + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U(q') \right) \right|$$

$$= \left| \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U_{t}^{q'} - \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U(q') \right|$$

$$= \gamma \left| \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U_{t}^{q'} - \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U(q') \right|$$

always nonnegative

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leq K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| = \gamma \left| \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U_{t}^{q'} - \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U(q') \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| = \gamma \left| \max_{\sigma} \underbrace{\sum_{q' \in Q} P(q' \mid q, \sigma) U_{t}^{q'}}_{f(\sigma)} - \max_{\sigma} \underbrace{\sum_{q' \in Q} P(q' \mid q, \sigma) U(q')}_{g(\sigma)} \right|$$

want to show: $\max\limits_{q \in Q} \left| U^q_{t+1} - U(q) \right| \leq K \max\limits_{q \in Q} \left| U^q_t - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U^q_{t+1} - U(q) \right| = \gamma \left| \max_{\sigma} f(\sigma) - \max_{\sigma} g(\sigma) \right|$$

$$f(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma) U^{q'}_t$$

$$g(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma) U(q')$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| = \gamma \left| \max_{\sigma} f(\sigma) - \max_{\sigma} g(\sigma) \right|$$

$$f(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma) U_{t}^{q'}$$
$$g(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma) U(q')$$

$$\left| \max_{\sigma} f(\sigma) - \max_{\sigma} g(\sigma) \right| \;\lessgtr?\; \max_{\sigma} \left| f(\sigma) - g(\sigma) \right|$$

how are these related?

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leq K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

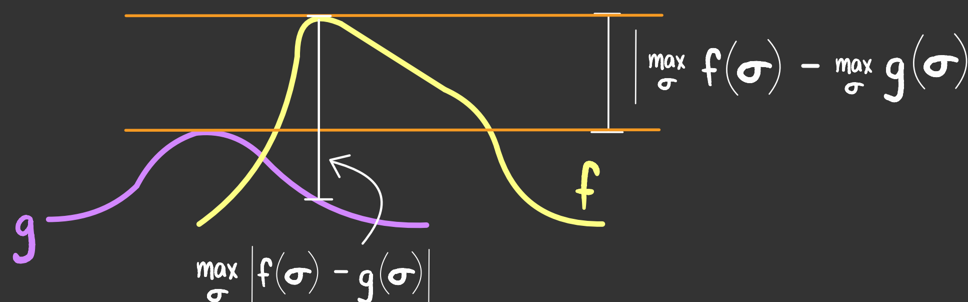
$$\left| U_{t+1}^{q} - U(q) \right| = \gamma \left| \max_{\sigma} f(\sigma) - \max_{\sigma} g(\sigma) \right|$$

$$f(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma) U_{t}^{q'}$$

$$g(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma) U(q')$$

$$\left| \max_{\sigma} f(\sigma) - \max_{\sigma} g(\sigma) \right|$$

$$\leq \max_{\sigma} \left| f(\sigma) - g(\sigma) \right|$$



want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leq K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$f(\sigma) = \sum_{q' \in Q} P(q' | q, \sigma) U_{t}^{q'}$$

$$g(\sigma) = \sum_{q' \in Q} P(q' | q, \sigma) U(q')$$

$$\left| U_{t+1}^{q} - U(q) \right| = \gamma \left| \max_{\sigma} f(\sigma) - \max_{\sigma} g(\sigma) \right|$$

$$\leq \gamma \max_{\sigma} \left| f(\sigma) - g(\sigma) \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leq K \max_{q \in Q} \left| U_t^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$f(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma) U_t^{q}$$

$$g(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma) U(q)$$

$$\begin{aligned}
\left| U_{t+1}^{q} - U(q) \right| &= \gamma \left| \max_{\sigma} f(\sigma) - \max_{\sigma} g(\sigma) \right| \\
&\leq \gamma \max_{\sigma} \left| f(\sigma) - g(\sigma) \right| \\
&= \gamma \max_{\sigma} \left| \sum_{q' \in Q} P(q' \mid q, \sigma) U_t^{q'} - \sum_{q' \in Q} P(q' \mid q, \sigma) U(q') \right|
\end{aligned}$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$f(\sigma) = \sum_{q' \in Q} P(q' | q, \sigma) U_{t}^{q}$$

$$g(\sigma) = \sum_{q' \in Q} P(q' | q, \sigma) U(q)$$

$$\left| U_{t+1}^{q} - U(q) \right| = \gamma \left| \max_{\sigma} f(\sigma) - \max_{\sigma} g(\sigma) \right|$$

$$\leqslant \gamma \max_{\sigma} \left| f(\sigma) - g(\sigma) \right|$$

$$= \gamma \max_{\sigma} \left| \sum_{q' \in Q} P(q' | q, \sigma) U_{t}^{q'} - \sum_{q' \in Q} P(q' | q, \sigma) U(q') \right|$$

$$= \gamma \max_{\sigma} \left| \sum_{q' \in Q} P(q' | q, \sigma) U_{t}^{q'} - P(q' | q, \sigma) U(q') \right|$$

want to show: $\max_{q \in Q} \left| U^q_{t+1} - U(q) \right| \leqslant K \max_{q \in Q} \left| U^q_t - U(q) \right|$ for some fraction $K \in [0, 1)$

$$f(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma)\, U^q_t$$
$$g(\sigma) = \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q)$$

$$\begin{aligned}
\left| U^q_{t+1} - U(q) \right| &= \gamma \left| \max_{\sigma} f(\sigma) - \max_{\sigma} g(\sigma) \right| \\
&\leqslant \gamma \max_{\sigma} \left| f(\sigma) - g(\sigma) \right| \\
&= \gamma \max_{\sigma} \left| \sum_{q' \in Q} P(q' \mid q, \sigma)\, U^{q'}_t - \sum_{q' \in Q} P(q' \mid q, \sigma)\, U(q') \right| \\
&= \gamma \max_{\sigma} \left| \sum_{q' \in Q} P(q' \mid q, \sigma) \left(U^{q'}_t - U(q') \right) \right|
\end{aligned}$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \left| \sum_{q' \in Q} P(q' \mid q, \sigma) \left(U_{t}^{q'} - U(q') \right) \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \left| \sum_{q' \in Q} \underbrace{P(q' \mid q, \sigma) \left(U_{t}^{q'} - U(q') \right)}_{f(q')} \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \left| \sum_{q' \in Q} f(q') \right|$$

$$\boxed{f(q') = P(q' \mid q, \sigma)\left(U_{t}^{q'} - U(q')\right)}$$

$$\text{want to show:} \quad \max_{q \in Q} \left| U^q_{t+1} - U(q) \right| \leq K \max_{q \in Q} \left| U^q_t - U(q) \right| \quad \text{for some fraction } K \in [0, 1)$$

$$\left| U^q_{t+1} - U(q) \right| \leq \gamma \max_{\sigma} \left| \sum_{q' \in Q} f(q') \right|$$

$$\boxed{f(q') = P(q' \mid q, \sigma)\left(U^{q'}_t - U(q')\right)}$$

$$\left| \sum_{q' \in Q} f(q') \right| \; \leq ? \geq \; \sum_{q' \in Q} \left| f(q') \right|$$

how are these related?

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \left| \sum_{q' \in Q} f(q') \right|$$

$$\boxed{f(q') = P(q' \mid q, \sigma)\left(U_{t}^{q'} - U(q')\right)}$$

$$\left| \sum_{q' \in Q} f(q') \right| \leqslant \sum_{q' \in Q} \left| f(q') \right|$$

e.g.

$$|2 + (-3) + 4| \leqslant |2| + |-3| + |4|$$

$$\boxed{\text{want to show:} \quad \max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right| \quad \text{for some fraction } K \in [0, 1)}$$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \left| \sum_{q' \in Q} f(q') \right|$$

$$\boxed{f(q') = P(q' \mid q, \sigma)\left(U_t^{q'} - U(q')\right)}$$

$$\leqslant \gamma \max_{\sigma} \sum_{q' \in Q} \left| f(q') \right|$$

$$= \gamma \max_{\sigma} \sum_{q' \in Q} \left| P(q' \mid q, \sigma)\left(U_t^{q'} - U(q')\right) \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \sum_{q' \in Q} \left| P(q' \mid q, \sigma) \left(U_{t}^{q'} - U(q') \right) \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_t^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \sum_{q' \in Q} \left| P(q' | q, \sigma) \left(U_t^{q'} - U(q') \right) \right|$$

$$= \gamma \max_{\sigma} \sum_{q' \in Q} \underbrace{P(q' | q, \sigma)}_{\text{always nonnegative}} \left| U_t^{q'} - U(q') \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \sum_{q' \in Q} \left| P(q' | q, \sigma) \left(U_{t}^{q'} - U(q') \right) \right|$$

$$= \gamma \max_{\sigma} \underbrace{\sum_{q' \in Q} P(q' | q, \sigma) \left| U_{t}^{q'} - U(q') \right|}$$

the expected value of

$\left| U_{t}^{q'} - U(q') \right|$

with respect to $P(q' | q, \sigma)$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \sum_{q' \in Q} \left| P(q' | q, \sigma) \left(U_{t}^{q'} - U(q') \right) \right|$$

$$= \gamma \max_{\sigma} \underbrace{\sum_{q' \in Q} P(q' | q, \sigma) \left| U_{t}^{q'} - U(q') \right|}$$

the expected value of $\left| U_{t}^{q'} - U(q') \right|$ with respect to $P(q' | q, \sigma)$ $\leqslant$ the maximum value of $\left| U_{t}^{q'} - U(q') \right|$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{\sigma} \sum_{q' \in Q} \left| P(q' | q, \sigma) \left(U_{t}^{q'} - U(q') \right) \right|$$

$$= \gamma \max_{\sigma} \underbrace{\sum_{q' \in Q} P(q' | q, \sigma) \left| U_{t}^{q'} - U(q') \right|}$$

the expected value of

$\left| U_{t}^{q'} - U(q') \right|$ $\leqslant$ $\max_{q'} \left| U_{t}^{q'} - U(q') \right|$

with respect to $P(q' | q, \sigma)$

want to show: $\max_{q\in Q} \left|U^{q}_{t+1} - U(q)\right| \leqslant K \max_{q \in Q} \left|U^{q}_{t} - U(q)\right|$ for some fraction $K \in [0, 1)$

$$\left|U^{q}_{t+1} - U(q)\right| \leqslant \gamma \max_{\sigma} \sum_{q' \in Q} \left| P(q' \mid q, \sigma)\left(U^{q'}_{t} - U(q')\right)\right|$$

$$= \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) \left|U^{q'}_{t} - U(q')\right|$$

$$\leqslant \gamma \max_{\sigma} \max_{q'} \left|U^{q'}_{t} - U(q')\right|$$

want to show: $\max_{q \in Q} \left| U^{q}_{t+1} - U(q) \right| \leqslant K \max_{q \in Q} \left| U^{q}_{t} - U(q) \right|$ for some fraction $K \in [0, 1)$

$$\left| U^{q}_{t+1} - U(q) \right| \leqslant \gamma \max_{\sigma} \sum_{q' \in Q} \left| P(q' \mid q, \sigma) \left(U^{q'}_{t} - U(q') \right) \right|$$

$$= \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) \left| U^{q'}_{t} - U(q') \right|$$

$$\leqslant \gamma \max_{\sigma} \max_{q'} \left| U^{q'}_{t} - U(q') \right|$$

$$\leqslant \gamma \max_{q'} \left| U^{q'}_{t} - U(q') \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

for every state $q \in Q$: $$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{q' \in Q} \left| U_{t}^{q'} - U(q') \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

for every state $q \in Q$: $$\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{q' \in Q} \left| U_{t}^{q'} - U(q') \right|$$

therefore: $$\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{q' \in Q} \left| U_{t}^{q'} - U(q') \right|$$

want to show: $\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant K \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|$ for some fraction $K \in [0, 1)$

for every state $q \in Q$: $\left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{q' \in Q} \left| U_{t}^{q'} - U(q') \right|$

therefore: $\boxed{\max_{q \in Q} \left| U_{t+1}^{q} - U(q) \right| \leqslant \gamma \max_{q \in Q} \left| U_{t}^{q} - U(q) \right|}$

$$\text{if}\ \operatorname{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) \leq \cancel{k}\,\gamma\, \operatorname{dist}\left(\begin{bmatrix} U_{t-1}^{q_0} \\ \vdots \\ U_{t-1}^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right),\quad \text{then:}\ \lim_{t\to\infty} \operatorname{dist}\left(\begin{bmatrix} U_t^{q_0} \\ \vdots \\ U_t^{q_n} \end{bmatrix}, \begin{bmatrix} U(q_0) \\ \vdots \\ U(q_n) \end{bmatrix}\right) = 0$$

for ~~some fraction K~~ discount γ $\in [0, 1)$

~~if we can show~~ we showed that every step of value iteration brings our guesses closer to the maximum expected utilities

if discount $\gamma < 1$

~~then~~ so value iteration converges to the maximum expected utilities

t	U(A)	U(B)
0	1.2500	-1.1300
1	1.6250	-1.3270
2	1.8125	-1.3683
3	1.9063	-1.3661
4	1.9531	-1.3558
5	1.9766	-1.3470
6	1.9883	-1.3411
7	1.9941	-1.3376
8	1.9971	-1.3356
9	1.9985	-1.3345
10	1.9993	-1.3340
11	1.9996	-1.3337
12	1.9998	-1.3335
13	1.9999	-1.3334
14	2.0000	-1.3334
15	2.0000	-1.3334

value iteration

- guess values for the unknowns

  for each state q, U_0^q is a guess for $U(q)$

- for $t = 1$ to T:

  compute new guesses using the bellman equations, i.e.

  for each state q, $U_{t+1}^q = R(q) + \gamma \max_{\sigma} \sum_{q' \in Q} P(q' \mid q, \sigma) U_t^{q'}$

if $\gamma < 1$, then value iteration converges to the maximum expected utilities, i.e. $\lim_{t \to \infty} U_t^q = U(q)$

proof

1 measure the distance between $U : Q \to \mathbb{R}$ and our guesses U_t^q at iteration t using:

$$\max_{q \in Q} \left| U_{t+1}^q - U(q) \right|$$

2 show:

$$\max_{q \in Q} \left| U_{t+1}^q - U(q) \right|$$

$$\leq \gamma \max_{q \in Q} \left| U_t^q - U(q) \right|$$

3 this implies:

$$\lim_{t \to \infty} U_t^q = U(q)$$