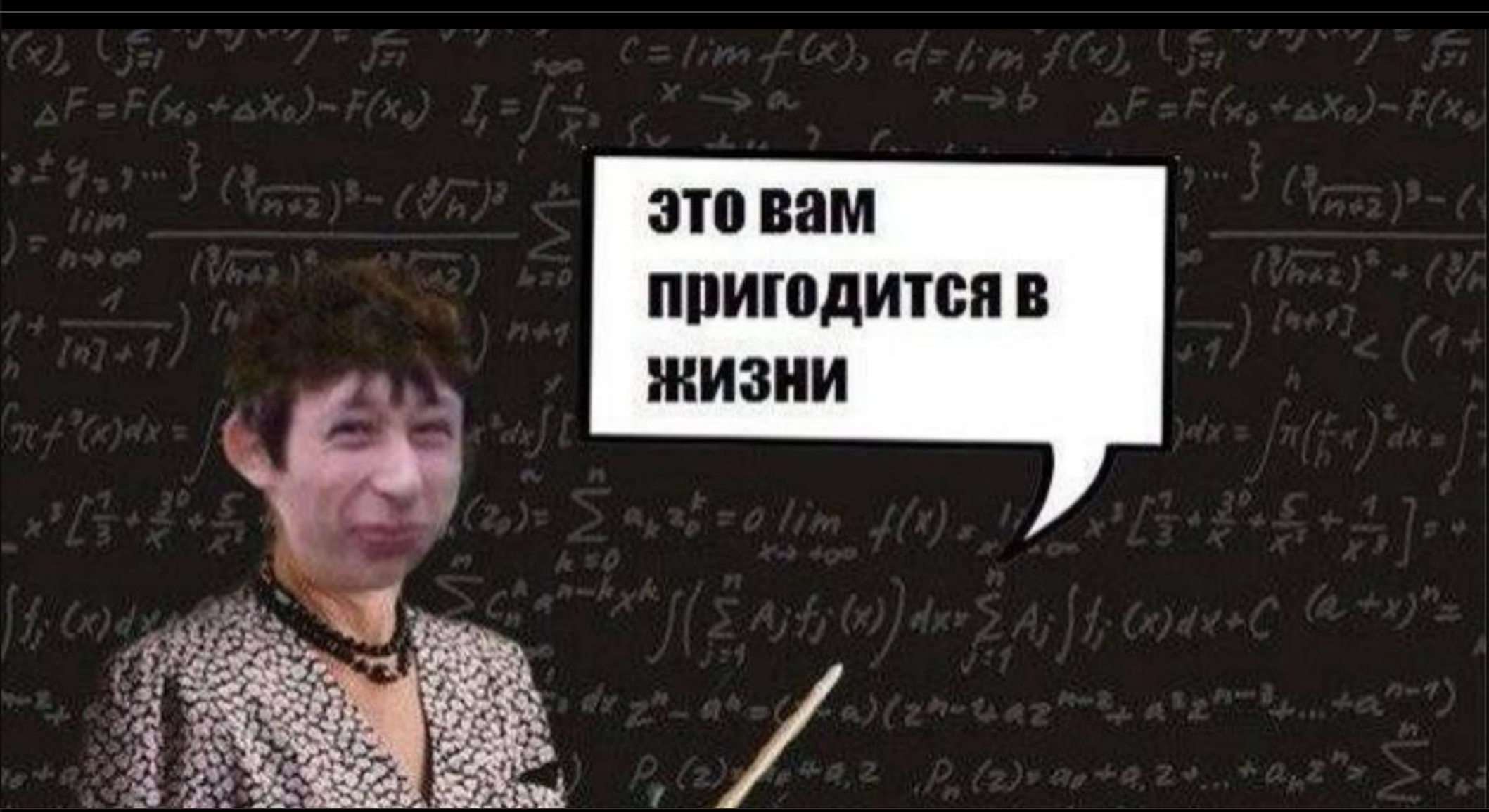
RL: Доказательство существования и единственности решения уравнений Беллмана


это вам
пригодится в
жизни

Утверждение

- В конечном Марковском процессе принятия решений (MDP) оптимальная политика определяется как политика, которая максимизирует ценность всех состояний одновременно

Уравнения Беллмана

$$V_\pi(s_t) = E_\pi\left(r_t + \gamma V(s_{t+1})\right)$$

$$Q_\pi(s_t) = E_\pi\left(r_t + \gamma Q(s_{t+1}, a_{t+1})\right)$$

- Связь между функцией ценности и ценности действия:

$$V_\pi(s_t) = E_\pi\left(Q(s_t, a_t)\right)$$

Оптимальная политика

- Политика является оптимальной тогда и только тогда, когда:

$$V_{\pi^*}(s) \geq V_{\pi}(s) \qquad (1)$$

- Для любого состояния s и любой другой политики

Теорема

- Если существует $s^* \in S$ и $a^* \in A(s^*)$ такие, что $Q_\pi(s^*, a^*) \geq V_\pi(s^*)$, тогда для политики $\pi'(a|s) = \pi(a|s)\left(1 - \delta_{s,s^*}\right) + \delta_{a,a^*}\delta_{s,s^*}$ мы имеем

$$V_{\pi'}(s) \geq V_\pi(s) \quad \forall s \in S$$

Доказательство: для любого $s \in S$

$$\begin{aligned} V_\pi(s) &= E_\pi(Q_\pi(s, A_t)|S_t = s) \leq \\ &\leq E_{\pi'}(Q_\pi(s, A_t)|S_t = s) \\ &= E_{\pi'}\left(R_{t+1} + \gamma V_\pi(S_{t+1})|S_t = s\right) \\ &\leq E_{\pi'}(R_{t+1} + \gamma R_{t+2} + \gamma^2 V_\pi(S_{t+2})|S_t = s) \\ &\leq E_{\pi'}(R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \cdots |S_t = s) \\ &= V_{\pi'}(s) \end{aligned}$$

- Теорема утверждает, что всякий раз, когда есть пара состояния и действия (s*, a*) со значением больше, чем значение состояния s*, относительно политики π, то существует другая политика π', которая лучше или равна (в терминах значений состояния) π во всех состояниях. В результате, если существует оптимальная политика π*, ее значения должны удовлетворять

$$V_{\pi^*}(s) = \max_{a \in A(s)} Q_{\pi^*}(s, a)$$

Уравнения оптимальности Беллмана

$$V_{\pi^*}(s) = \max_{a \in A(s)} Q_{\pi^*}(s, a)$$

$$V_{\pi^*}(s) = \max_{a \in A} E\left(R_{t+1} + \gamma V_{\pi^*}(S_{t+1}) | S_t = s, A_t = a\right)$$

$$= \max_{a \in A} \left(\sum_r r p(r|s, a) + \gamma \sum_{s'} V_{\pi^*}(s') p(s'|s, a) \right)$$

Существование и единственность

Чтобы показать, что существует оптимальная политика, необходимо доказать следующие два утверждения:

1. Система уравнений оптимальности Беллмана имеет решения, и
2. Одно из его решений имеет значения больше или равные значениям других решений во всех состояниях.

Оператор оптимальности Беллмана

$$V = \{V(s): s \in S\} \in R^{|S|}$$

- Оператор оптимальности Беллмана

$B: R^{|S|} \to R^{|S|}$ такой, что

$$BV(s) = \max_{a \in A(s)} E(R_{t+1} + \gamma V(S_{t+1}) | S_t = s, A_t = a)$$

$$V_{\pi^*} = BV_{\pi^*}$$

Сжимающие отображения и теорема Банах о неподвижной точек

Теорема:

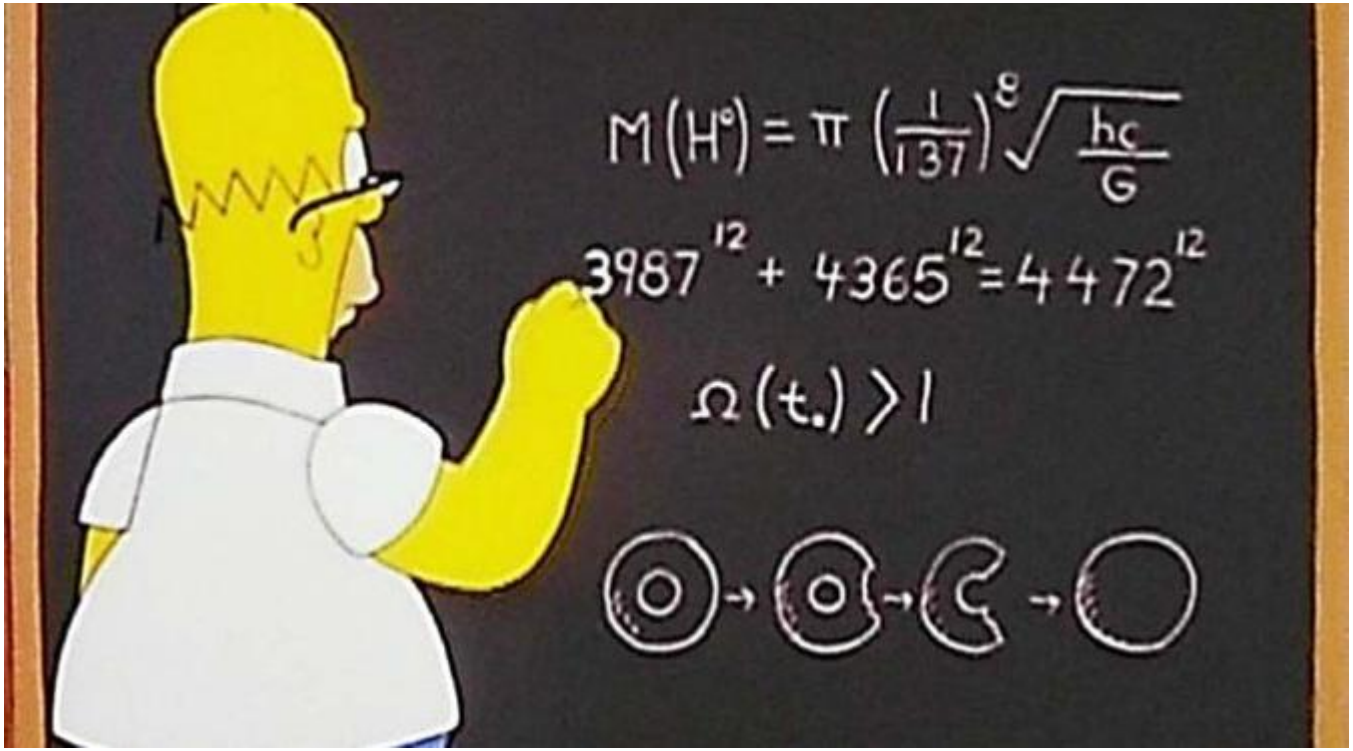
Пусть (X, d) - непустое полное метрическое пространство

Пусть $B: X \to X$ сжимающее отображение на *X*, т.е. существует число $0 \leq \alpha < 1$, такое что

$d(Bx, By) \leq \alpha d(x, y)$, для всех *x* и *y* из *X*

Тогда у отображения B существует, и притом единственная, неподвижная точка x^* из *X* (т.е. $Bx^* = x^*$)

Доказательство



Оператор оптимальности Беллмана – это сжимающее отображение в бесконечной норме

Бесконечная норма:

$$\|V - V'\|_\infty = \max_{s \in S} |V(s) - V(s')|$$

Лемма: Пусть задано ограниченное множество A, тогда для любых значений функций f, таких, что $f_1 : A \to R$ и $f_2 : A \to R$ справедливо:

$$\left| \max_{a \in A} f_1(a) - \max_{a \in A} f_2(a) \right| \leq \max_{a \in A} |f_1(a) - f_2(a)|$$

Теорема

На метрическом пространстве $\left(R^{|S|}, \|\cdot\|_\infty\right)$, оператор оптимальности Беллмана $B: R^{|S|} \to R^{|S|}$ сжимающий, т.е.

$$\exists \alpha \in [0,1): \|BV - BV'\|_\infty \leq \alpha \|V - V'\|_\infty \;\; \forall V, V' \in R^{|S|}$$

Доказательство

$$\|BV - BV'\|_\infty = \max_{s \in S} |BV(s) - BV'(s)|$$

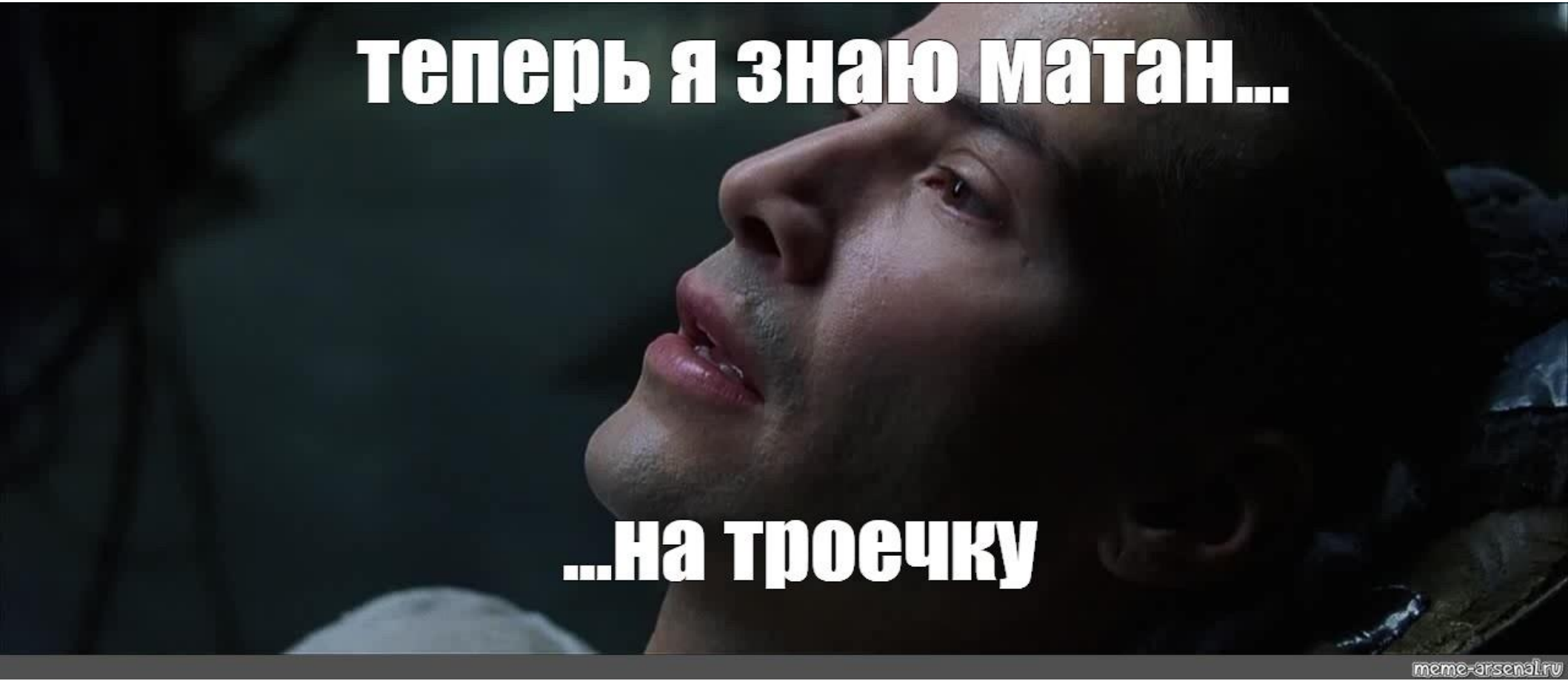
$$= \max_{s \in S} \left| \max_{a \in A(s)} E[R_{t+1} + \gamma V(s_{t+1})] - \max_{a \in A(s)} E[R_{t+1} + \gamma V'(s_{t+1})] \right|$$

$$\leq \max_{s \in S} \max_{a \in A(s)} |E[R_{t+1} + \gamma V(s_{t+1})] - E[R_{t+1} + \gamma V'(s_{t+1})]|$$

$$= \max_{s \in S} \max_{a \in A(s)} |\gamma E[V(s_{t+1}) - V'(s_{t+1})]|$$

$$\leq \max_{s \in S} \max_{a \in A(s)} \gamma E|[V(s_{t+1}) - V'(s_{t+1})]|$$

$$\leq \max_{s \in S} \max_{a \in A(s)} \gamma E|[V - V']| = \gamma \|V - V'\|_\infty$$


теперь я знаю матан...
...на троечку
meme-arsenal.ru