

Неподвижные точки и рекуррентные схемы

Мальковский Н. В.

Санкт-Петербургский Академический Университет



Определение

Пусть X – произвольное пространство, $f : X \rightarrow X$. Неподвижной точкой отображения f на X называется $x^* \in X$, для которого выполняется

$$f(x^*) = x^*.$$

Примеры

1. $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^2$, две неподвижные точки: $\{0, 1\}$.
2. $g : \mathbb{R}^n \rightarrow \mathbb{R}$, g дифференцируема $f(x) = x - \alpha \nabla g(x)$. Любая точка минимума g является неподвижной точкой f .
3. $\frac{d}{dx} : C^{(1)}[a, b] \rightarrow C[a, b]$ – оператор дифференцирования. Неподвижные точки – функции вида Ce^x .
4. $f : \Delta_n \rightarrow \Delta_n$, где $\Delta_n = \{x \in \mathbb{R}^n \mid x \geq 0_n, \mathbf{1}^T x = 1\}$ – стандартный n -мерный симплекс, $f(x) = Px$, где P^T – стохастическая матрица.
Процесс

$$x_{k+1} = f(x_k)$$

принято называть *марковским процессом*. Компонента i вектора x_k представляют собой вероятность оказаться в состоянии i на итерации k . Матрица P задает вероятность перехода. Предельное состояние процесса является неподвижной точкой f .

Определение

Пусть (X, δ_X) , (Y, δ_Y) – метрические пространства. Отображение $f : X \rightarrow Y$ называется непрерывным по Липшицу с константой $L > 0$, если $\forall x, y \in X$ выполняется

$$\delta_Y(f(x), f(y)) \leq L\delta_X(x, y).$$

Замечание. Непрерывность по Липшицу можно ввести и для нормированного пространства как

$$\|f(x) - f(y)\|_Y \leq L\|x - y\|_X$$

Непрерывность по Липшицу

Теорема (О функции с непрерывным по Липшицу градиентом)

Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, ∇f непрерывен по Липшицу с константой L , тогда $\forall x, y$

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|y - x\|^2,$$

$$(\nabla f(y) - \nabla f(x))^T(y - x) \leq L\|y - x\|^2.$$

Док-во. Заметим, что

$$(\nabla f(y) - \nabla f(x))^T(y - x) \leq \|\nabla f(y) - \nabla f(x)\| \cdot \|y - x\| \leq L\|y - x\|^2.$$

Остальное получается по аналогии с сильной выпуклостью.

Сжимающее отображение

Определение

Пусть (X, δ) – метрическое пространство. Отображение $f : X \rightarrow X$ называется сжимающим, если оно непрерывно по Липшицу с константой $L < 1$.

Примеры

1. $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $f(x) = \alpha x + b$, $|\alpha| < 1$.
2. $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $f(x) = Ax + b$, A – симметричная вещественная матрица $\sigma(A) < 1$ ($\sigma(A) = \max\{|\lambda| \mid \exists v \neq 0_n, Av = \lambda v\}$).

Теорема Банаха

Теорема (Банаха о сжимающем отображении)

Пусть (X, δ) – непустое полное метрическое пространство (банахово пространство), f – сжимающее отображение на X с константой q , тогда

- Существует единственная неподвижная точка $x^* \in X$ функции на f на X .
- При любом $x_0 \in X$ рекуррентная последовательность $x_{k+1} = f(x_k)$ сходится к x^* .

Док-во. Для последовательности $x_{k+1} = f(x_k)$ имеет место

$$\delta(x_{k+1}, x_k) = \delta(f(x_k), f(x_{k-1})) \leq q\delta(x_k, x_{k-1}) \leq \dots \leq q^k \delta(x_1, x_0).$$

Таким образом, для любых $n < m$

$$\delta(x_n, x_m) \leq \sum_{i=n}^{m-1} \delta(x_{i+1}, x_i) \leq q^n \sum_{i=0}^{m-n-1} q^i \delta(x_1, x_0) \leq \frac{q^n}{1-q} \delta(x_1, x_0),$$

а значит x_k – фундаментальная последовательность (последовательность Коши). Так как X полно, то x_k имеет предел в X , обозначим его за x^* .

Теорема Банаха

Так как $x_k \xrightarrow{\delta} x^*$, то

$$\delta(f(x_k), f(x^*)) \leq q\delta(x_k, x^*) \rightarrow 0,$$

т.е. $f(x_k) \xrightarrow{\delta} f(x^*)$. Переходя к пределу в рекуррентном соотношении получаем

$$x^* = \lim_{k \rightarrow \infty} x_{k+1} = \lim_{k \rightarrow \infty} f(x_k) = f(x^*).$$

Осталось показать единственность x^* : пусть $f(y) = y$, тогда

$$\delta(x^*, y) = \delta(f(x^*), f(y)) \leq q\delta(x^*, y) \Rightarrow \delta(x^*, y) = 0 \Rightarrow x^* = y. \blacksquare$$

Замечание. Можно оценить скорость сходимости:

$$\delta(x_k, x^*) \leq \sum_{i=k}^{\infty} \delta(x_i, x_{i+1}) \leq q^k \sum_{i=0}^{\infty} q^i \delta(x_1, x_0) = \frac{q^k}{1-q} \delta(x_1, x_0)$$

Замечания

Замечание 1. Легко проверить, что для функции $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = \frac{x + \sqrt{x^2 + 1}}{2}$ выполняется

$$|f(x) - f(y)| < |x - y|,$$

но при этом у f нет неподвижных точек

$$f(x) - x = \frac{\sqrt{x^2 + 1} - x}{2} > \frac{|x| - x}{2} \geq 0.$$

Замечание 2. Для аффинных отображений $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $f(x) = Ax + b$ выполняется

$$\|f(x) - f(y)\| = \|Ax - Ay\| \leq \|A\| \cdot \|x - y\|.$$

Для симметричных матриц $\|A\| = \sigma(A)$, что неверно для несимметричных матриц, однако условие $\sigma(A) < 1$ всегда гарантирует сходимость последовательности

$$x_{k+1} = f(x_k)$$

Сходимость линейных рекуррентных процессов

Посмотрим подробнее на последовательность вида

$$x_{k+1} = Ax_k + b, \quad (1)$$

где $x_k, b \in \mathbb{R}^n$, $\mathbb{R}^{n \times n}$.

Заметим для начала что для жорданова блока J_i

$$J_i = \begin{pmatrix} \lambda_i & 1 & \dots & 0 & 0 \\ 0 & \lambda_i & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & \dots & 0 & \lambda_i \end{pmatrix}$$

выполняется

$$J_i^k = \begin{pmatrix} \lambda_i^k & C_k^1 \lambda_i^{k-1} & \dots & C_k^{n-2} \lambda_i^{k-n+2} & C_k^{n-1} \lambda_i^{k-n+1} \\ 0 & \lambda_i^k & \dots & C_k^{n-3} \lambda_i^{k-n+3} & C_k^{n-2} \lambda_i^{k-n+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_i^k & C_k^1 \lambda_i^{k-1} \\ 0 & 0 & \dots & 0 & \lambda_i^k \end{pmatrix}$$

Сходимость линейных рекуррентных процессов

Далее, при фиксированном n , C_k^n – полином степени n от k , рост которой сколь угодно мал по сравнению с λ_i^k . Опуская некоторые технические детали, $\forall 0 < \epsilon < 1 - |\lambda_i| \exists C > 0 : \forall k$ выполняется

$$C_k^n \lambda_i^k \leq C(|\lambda_i| + \epsilon)^k$$

и аналогично для J^k : $\forall 0 < \epsilon < 1 - |\lambda_i| \exists C : \forall k$ выполняется

$$\|J_i^k\| \leq C(|\lambda_i| + \epsilon)^k$$

Далее, если $A = P^{-1}JP$ – жорданова форма A , то

$$J^k = \begin{pmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_l \end{pmatrix}^k = \begin{pmatrix} J_1^k & 0 & \dots & 0 \\ 0 & J_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_l^k \end{pmatrix}$$

И наконец $A^k = P^{-1}JPP^{-1}JP \dots P^{-1}JP = P^{-1}J^kP$, что наконец дает $\forall 0 < \epsilon < 1 - |\lambda_i| \exists C : \forall k \|A^k\| \leq C(\sigma(A) + \epsilon)^k$. ■

Сходимость линейных рекуррентных процессов

Теперь вернемся к последовательности (1). По аналогии с теоремой Банаха докажем существование предела в случае $\sigma(A) < 1$:

$$\begin{aligned}\|x_{k+1} - x_k\| &= \|Ax_k + b - Ax_{k-1} - b\| = \|A(x_k - x_{k-1})\| = \dots \\ &= \|A^k(x_1 - x_0)\| \leq \|A^k\| \|x_1 - x_0\|.\end{aligned}$$

При $n < m$

$$\|x_m - x_n\| \leq \sum_{i=n}^{m-1} \|x_{i+1} - x_i\| \leq \|A^n\| \|x_1 - x_0\| \sum_{i=0}^{m-n-1} \|A^i\|$$

Из полученных результатов $\exists C_1 > 0$: $\sum_{i=0}^{\infty} \|A^i\| < C_1$. Пусть $\epsilon = \frac{1}{2}(1 - \sigma(A))$, тогда для некоторого $C_2 \forall n$

$$\|x_m - x_n\| \leq C_1 C_2 \left(\frac{\sigma(A) + 1}{2} \right)^n \|x_1 - x_0\| \xrightarrow{n \rightarrow \infty} 0,$$

то есть x_k – фундаментальная последовательность, а значит имеет предел, обозначим его x^* .

Сходимость линейных рекуррентных процессов

Очевидным образом

$$x^* = Ax^* + b.$$

Наконец

$$\|x_k - x^*\| = \|Ax_{k-1} + b - Ax^* - b\| = \dots = \|A^k(x_0 - x^*)\| \leq C(\sigma(A) + \epsilon)^k \|x_0 - x^*\|. \blacksquare$$

Сходимость нелинейных рекуррентных процессов

Пусть теперь f – произвольная функция, x^* – неподвижная точка f , рассмотрим процесс

$$x_{k+1} = f(x_k) \quad (2)$$

Если f дифференцируема в точке x^* , то поведение (2) схоже её линейному приближению $x_{k+1} = f(x^*) + \nabla f(x^*)^T (x_k - x^*)$.

Теорема

Если f дифференцируема в точке x^* , тогда для любого $0 < \epsilon < 1 - \sigma(\nabla f(x^*))$ найдутся такие $\delta > 0$ и C , что при $\|x_0 - x^*\| < \delta$ в процессе (2) выполняется

$$\|x_k - x^*\| \leq C(\sigma(\nabla f(x^*)) + \epsilon)^k$$

Док-во. Обозначим $A = \nabla f(x^*)$, $\sigma = \sigma(A)$. Из дифференцируемости f в x^* существует $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\frac{\alpha(x)}{\|x - x^*\|} \xrightarrow{x \rightarrow x^*} 0$

$$f(x) = f(x^*) + A(x - x^*) + \alpha(x)$$

Сходимость нелинейных рекуррентных процессов

$$\begin{aligned}\|x_k - x^*\| &= \|A(x_{k-1} - x^*) + \alpha(x_{k-1})\| = \dots \\ &\leq \|A^k\| \|x_0 - x^*\| + \sum_{i=1}^k \|A^{i-1}\| \|\alpha(x_{k-i})\|\end{aligned}$$

Из доказанного ранее $\exists C : \|A^k\| \leq C(\sigma + \epsilon)^k$. Зафиксируем некоторое $K > 0$, выберем δ так, чтобы для всех x_k $0 \leq k \leq K$ выполнялось

$|\alpha(x_k)| \leq \frac{\sigma + \epsilon}{C(C+1)K} \|x_k - x^*\|$. Докажем по индукции, что в этом случае при $0 \leq k \leq K$ выполняется $\|x_k - x^*\| \leq (C + 1)(\sigma + \epsilon)^k \|x_0 - x^*\|$:

$$\begin{aligned}\|x_k - x^*\| &\leq \|A^k\| \|x_0 - x^*\| + \sum_{i=1}^k \|A^{i-1}\| \|\alpha(x_{k-i})\| \\ &\leq C(\sigma + \epsilon)^k \|x_0 - x^*\| + \frac{1}{(C + 1)K} \sum_{i=1}^k (\sigma + \epsilon)^i \|x_{k-i} - x^*\| \\ &\leq \left(C(\sigma + \epsilon)^k + \frac{1}{K} \sum_{i=1}^k (\sigma + \epsilon)^k \right) \|x_0 - x^*\| \leq (C + 1)(\sigma + \epsilon)^k \|x_0 - x^*\|\end{aligned}$$

Сходимость нелинейных рекуррентных процессов

Выберем K (и вместе с ним δ) так, чтобы выполнялось $\gamma = (C + 1)(\sigma + \epsilon)^K < 1$, что дает нам

$$\|x_K - x^*\| \leq \gamma \|x_0 - x^*\| < \|x_0 - x^*\|.$$

Учитывая это можно повторить всю процедуру не изменяя выбранных констант начиная не из x_0 , а из x_K , что дает для $0 \leq k \leq K$

$$\|x_{K+k} - x^*\| \leq (C + 1)(\sigma + \epsilon)^k \|x_K - x^*\| \leq (C + 1)^2 \gamma (\sigma + \epsilon)^{K+k} \|x_0 - x^*\|$$

Повторяя нужное число раз, при $k = iK + j$

$$\begin{aligned} \|x_k - x^*\| &\leq (C + 1)^{i+1} (\sigma + \epsilon)^k \|x_0 - x^*\| \\ &\leq (C + 1) \left((C + 1) \left(\frac{\sigma + \epsilon}{\sigma + 2\epsilon} \right)^K \right)^i (\sigma + 2\epsilon)^k \end{aligned}$$

Наконец выберем K и вместе с ним δ так, чтобы $(C + 1) \left(\frac{\sigma + \epsilon}{\sigma + 2\epsilon} \right)^K < 1$. ■

Замечание 1. Схожие соображения используются для анализа локальной устойчивости динамических систем: если $x^* = f(x^*)$, то решение $x(t) = x^*$ системы дифференциальных

$$\dot{x} = f(x) - x$$

асимптотически устойчива если вещественные части всех собственных чисел матрицы $\nabla f(x^*) - I$ строго отрицательны. Систему (2) можно рассматривать как дискретизацию этой:

$$x_{k+1} = f(x_k) \Rightarrow x_{k+1} - x_k = f(x_k) - x_k$$

Квадратичная сходимость рекуррентных процессов

Теорема

Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $x^* = f(x^*)$, f дифференцируема на $S = \{\|x - x^*\| \leq \|x_0 - x^*\|\}$, ∇f удовлетворяет условию Липшица с константой L , $\nabla f(x^*) = 0$ и $q = \frac{L}{2}\|x_0 - x^*\| < 1$, то последовательность (2) удовлетворяет

$$\|x_k - x^*\| \leq \frac{2}{L}q^{2^k}$$

Док-во.

$$\|x_1 - x^*\| = \|f(x_0) - f(x^*)\| = \|f(x_0) - f(x^*) - \nabla f(x^*)^T(x_0 - x^*)\| \leq \frac{L}{2}\|x_0 - x^*\|^2$$

В частности $\|x_1 - x^*\| \leq q\|x_0 - x^*\|$, т.е. $x_1 \in S$, что позволяет повторить оценку для x_2, \dots

$$\begin{aligned} \|x_k - x^*\| &\leq \frac{L}{2}\|x_{k-1} - x^*\|^2 \\ &\leq \left(\frac{L}{2}\right)^{2^k - 1} \|x_0 - x^*\|^{2^k} = \frac{2}{L}q^{2^k} \quad \blacksquare \end{aligned}$$