

Метод Ньютона

Мальковский Н. В.

Санкт-Петербургский Академический Университет



Метод Ньютона и градиентный спуск

Градиентный спуск:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k).$$

Метод Ньютона и градиентный спуск

Градиентный спуск:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k).$$

Основная формальная модификация метода Ньютона:

$$\alpha_k \rightarrow \alpha(x) \in \mathbb{R}^{n \times n}$$

Метод Ньютона и градиентный спуск

Градиентный спуск:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k).$$

Основная формальная модификация метода Ньютона:

$$\alpha_k \rightarrow \alpha(x) \in \mathbb{R}^{n \times n}$$

Общая идея: на каждом шаге минимизируем квадратичное приближение f :

$$x_{k+1} = \operatorname{argmin}_x (f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2} (x - x_k) \nabla^2 f(x_k) (x - x_k))$$

Метод Ньютона и градиентный спуск

Градиентный спуск:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k).$$

Основная формальная модификация метода Ньютона:

$$\alpha_k \rightarrow \alpha(x) \in \mathbb{R}^{n \times n}$$

Общая идея: на каждом шаге минимизируем квадратичное приближение f :

$$x_{k+1} = \operatorname{argmin}_x (f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2} (x - x_k) \nabla^2 f(x_k) (x - x_k))$$

Условия оптимальности первого рода дают уравнение на x_{k+1} :

$$\nabla f(x_k) + \nabla^2 f(x_k) (x - x_k) = 0$$

Таким образом

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k) \quad (1)$$

За и против

Плюсы:

За и против

Плюсы:

- Сходится гораздо быстрее.

За и против

Плюсы:

- Сходится гораздо быстрее.

Минусы:

- Нужна возможность измерить гессиан $\nabla^2 f(x_k)$

За и против

Плюсы:

- Сходится гораздо быстрее.

Минусы:

- Нужна возможность измерить гессиан $\nabla^2 f(x_k)$
- Больше вычислений на итерацию: необходимо либо обращать гессиан, либо решать систему

$$\nabla^2 f(x_k)(x - x_k) = -\nabla f(x_k)$$

Метод Ньютона для уравнений

Рассмотрим систему уравнений

$$f(x) = 0_n,$$

где $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, f дифференцируема.

Метод Ньютона для уравнений

Рассмотрим систему уравнений

$$f(x) = 0_n,$$

где $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, f дифференцируема.

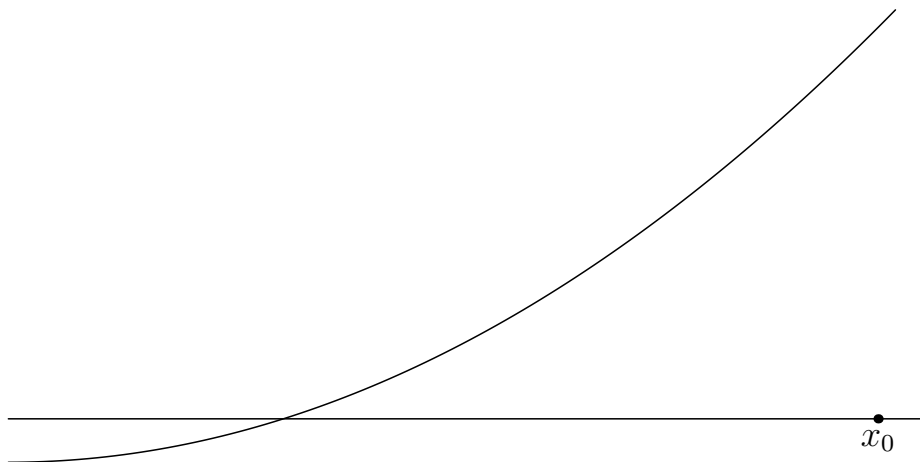
Аналогично задаче минимизации, метод Ньютона для решение системы уравнений заключается в замене исходной системы на её линейное приближение и последовательное итерирование. Решение линейной системы:

$$x_{k+1} : f(x_k) + \nabla f(x_k)(x_{k+1} - x_k) = 0 \Rightarrow$$

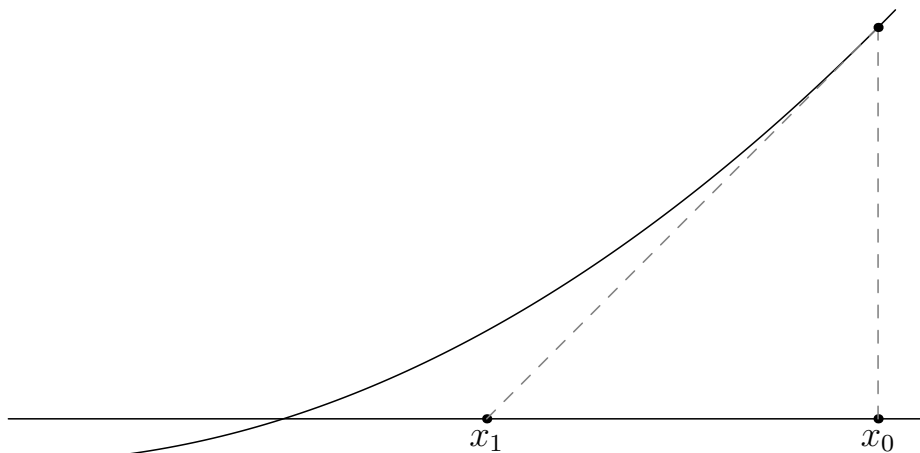
что дает

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

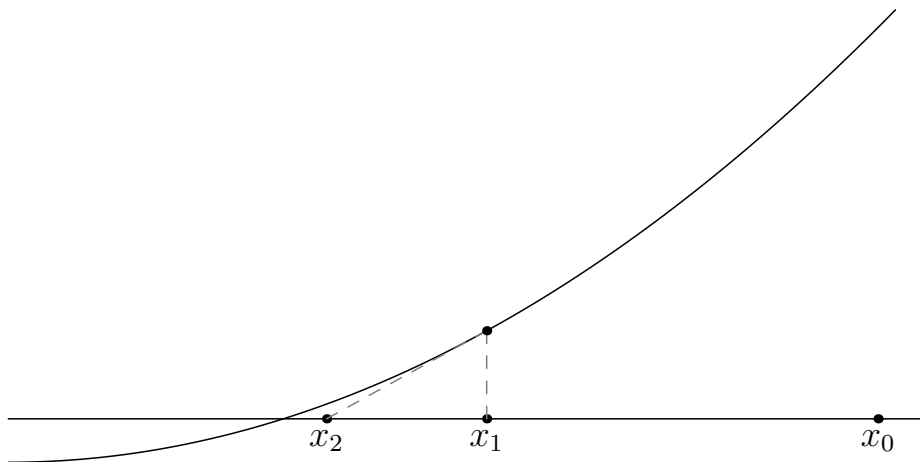
Метод Ньютона в одномерном случае



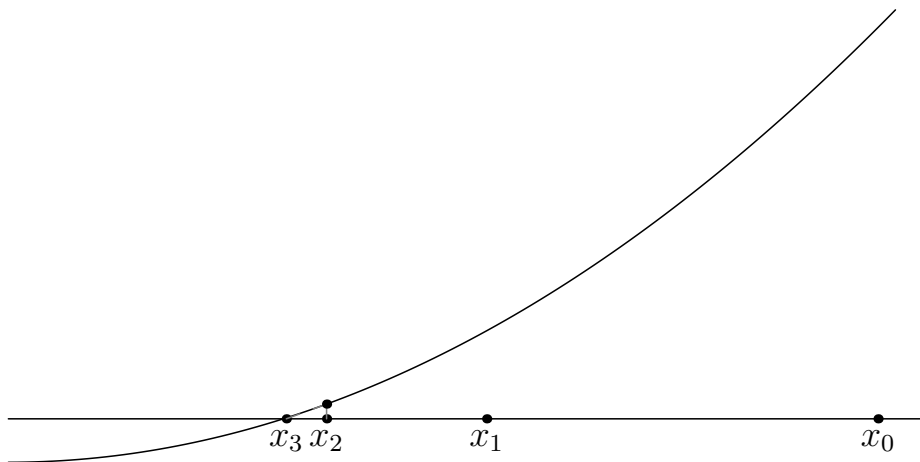
Метод Ньютона в одномерном случае



Метод Ньютона в одномерном случае



Метод Ньютона в одномерном случае



Пример: вычисление обратной функции

Предположим, что мы умеем вычислять $f(x)$ и $\nabla f(x)$, хотим научиться вычислять $f^{-1}(x)$.

Пример: вычисление обратной функции

Предположим, что мы умеем вычислять $f(x)$ и $\nabla f(x)$, хотим научиться вычислять $f^{-1}(x)$.

Пусть $f^{-1}(y^*) = x^*$, тогда $f(x^*) = y^*$, отсюда вычисление $f^{-1}(y^*)$ равносильно решению системы

$$f(x) - y^* = 0.$$

Пример: вычисление обратной функции

Предположим, что мы умеем вычислять $f(x)$ и $\nabla f(x)$, хотим научиться вычислять $f^{-1}(x)$.

Пусть $f^{-1}(y^*) = x^*$, тогда $f(x^*) = y^*$, отсюда вычисление $f^{-1}(y^*)$ равносильно решению системы

$$f(x) - y^* = 0.$$

Используя метод Ньютона получаем алгоритм

$$x_{k+1} = x_k - [\nabla f(x)]^{-1}(f(x_k) - y^*)$$

Пример: вычисления вещественного корня

Пусть $f(x) = x^2$, $f^{-1}(x) = \sqrt{x}$, тогда рассмотренная схема имеет вид

$$x_{k+1} = x_k - \frac{x_k^2 - y^*}{2x_k} = \frac{1}{2} \left(x_k + \frac{y^*}{x_k} \right)$$

Пример: вычисления вещественного корня

Пусть $f(x) = x^2$, $f^{-1}(x) = \sqrt{x}$, тогда рассмотренная схема имеет вид

$$x_{k+1} = x_k - \frac{x_k^2 - y^*}{2x_k} = \frac{1}{2} \left(x_k + \frac{y^*}{x_k} \right)$$

Функция $g(x) = \frac{1}{2} \left(x + \frac{y^*}{x} \right)$ имеет две неподвижные точки $\pm\sqrt{y^*}$, а значит последовательность $x_{k+1} = g(x_k)$ сходится либо к одной из них

Очевидным образом, если $x > 0$, то $g(x) > 0$ и наоборот. Следовательно, если $x_0 > 0$, то последовательность сходится к $\sqrt{y^*}$, а если $x_0 < 0$, то последовательность сходится к $-\sqrt{y^*}$.

Рассмотрим близкий к (2) метод:

$$x_{k+1} = x_k - [\nabla f(x^*)]^{-1} f(x_k) \quad (3)$$

Анализ сходимости

Рассмотрим близкий к (2) метод:

$$x_{k+1} = x_k - [\nabla f(x^*)]^{-1} f(x_k) \quad (3)$$

Этот метод представляет только теоретическую ценность, так как предполагает возможность вычисления $\nabla f(x^*)$, что практически не осуществимо без знания x^* .

Анализ сходимости

Рассмотрим близкий к (2) метод:

$$x_{k+1} = x_k - [\nabla f(x^*)]^{-1}f(x_k) \quad (3)$$

Этот метод представляет только теоретическую ценность, так как предполагает возможность вычисления $\nabla f(x^*)$, что практически не осуществимо без знания x^* .

Пусть $g(x) = x - [\nabla f(x^*)]^{-1}f(x)$, тогда (3) имеет вид

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

и при этом

$$\nabla g(x) = I - [\nabla f(x^*)]^{-1}\nabla f(x).$$

Анализ сходимости

Рассмотрим близкий к (2) метод:

$$x_{k+1} = x_k - [\nabla f(x^*)]^{-1}f(x_k) \quad (3)$$

Этот метод представляет только теоретическую ценность, так как предполагает возможность вычисления $\nabla f(x^*)$, что практически не осуществимо без знания x^* .

Пусть $g(x) = x - [\nabla f(x^*)]^{-1}f(x)$, тогда (3) имеет вид

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

и при этом

$$\nabla g(x) = I - [\nabla f(x^*)]^{-1}\nabla f(x).$$

Таким образом, $g(x^*) = 0$ и ∇g удовлетворяет условию Липшица, если ∇f удовлетворяет.

Анализ сходимости

Используя теорему о квадратичной сходимости для рекуррентных процессов получаем

Теорема

Если f дифференцируема на $S = \{x \mid \|x - x^*\| \leq \|x_0 - x^*\|\}$, $\det \nabla f(x^*) \neq 0$, ∇f удовлетворяет условию Липшица с константой L на S и

$q = \frac{1 + \|\nabla f(x^*)\|^{-1} L}{2} \|x_0 - x^*\| < 1$, то для последовательности (3) выполняется

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

Анализ сходимости

Используя теорему о квадратичной сходимости для рекуррентных процессов получаем

Теорема

Если f дифференцируема на $S = \{x \mid \|x - x^*\| \leq \|x_0 - x^*\|\}$, $\det \nabla f(x^*) \neq 0$, ∇f удовлетворяет условию Липшица с константой L на S и $q = \frac{1 + \|\nabla f(x^*)\|^{-1} \|L\|}{2} \|x_0 - x^*\| < 1$, то для последовательности (3) выполняется

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

Док-во Применяем теорему о квадратичной сходимости для $g(x) = x - [\nabla f(x^*)]^{-1} f(x)$ и учитываем, что

$$\begin{aligned} \|\nabla g(x) - \nabla g(y)\| &= \|[\nabla f(x^*)]^{-1} (\nabla f(x) - \nabla f(y)) + (x - y)\| \\ &\leq (1 + \|\nabla f(x^*)\|^{-1} \|L\|) \|x - y\| \quad \blacksquare \end{aligned}$$

Анализ сходимости

Возвращаемся к (1).

Теорема (О скорости сходимости метода Ньютона для задач оптимизации)

Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}$ дважды дифференцируема на $S = \left\{ \|x - x^*\| \leq \frac{m}{\gamma M} \right\}$ при некотором $\gamma \geq 3/2$, x^* – точка минимума f на S и $ml \leq \nabla^2 f(x^*)$ при $m > 0$, для $\nabla^2 f$ на S выполняется условие Липшица с константой M , $x_0 \in S$, тогда для последовательности (1) выполняется

$$\|x_{k+1} - x^*\| \leq \frac{M \|x_k - x^*\|^2}{2(m - M \|x_k - x^*\|)}$$

Анализ сходимости

Возвращаемся к (1).

Теорема (О скорости сходимости метода Ньютона для задач оптимизации)

Пусть $f : \mathbb{R}^n \rightarrow \mathbb{R}$ дважды дифференцируема на $S = \{ \|x - x^*\| \leq \frac{m}{\gamma M} \}$ при некотором $\gamma \geq 3/2$, x^* – точка минимума f на S и $mI \preceq \nabla^2 f(x^*)$ при $m > 0$, для $\nabla^2 f$ на S выполняется условие Липшица с константой M , $x_0 \in S$, тогда для последовательности (1) выполняется

$$\|x_{k+1} - x^*\| \leq \frac{M \|x_k - x^*\|^2}{2(m - M \|x_k - x^*\|)}$$

Док-во. Обозначим $G_k = \int_0^1 [\nabla f(x_k) - \nabla f(x^* + t(x_k - x^*))] dt$, тогда

$$\begin{aligned} x_{k+1} - x^* &= x_k - x^* - [\nabla f(x_k)]^{-1} f(x_k) \\ &= x_k - x^* - [\nabla f(x_k)]^{-1} \int_0^1 \nabla f(x^* + t(x_k - x^*)) (x_k - x^*) dt \\ &= [\nabla f(x_k)]^{-1} G_k (x - x^*). \end{aligned}$$

Анализ сходимости

Далее

$$\begin{aligned}\|G_k\| &= \left\| \int_0^1 [\nabla f(x_k) - \nabla f(x^* + t(x_k - x^*))] dt \right\| \\ &\leq \int_0^1 \|[\nabla f(x_k) - \nabla f(x^* + t(x_k - x^*))]\| dt \\ &\leq \int_0^1 M(1-t)\|x_k - x^*\| dt \\ &= \frac{M\|x_k - x^*\|}{2}\end{aligned}$$

Анализ сходимости

Далее

$$\begin{aligned}\|G_k\| &= \left\| \int_0^1 [\nabla f(x_k) - \nabla f(x^* + t(x_k - x^*))] dt \right\| \\ &\leq \int_0^1 \|[\nabla f(x_k) - \nabla f(x^* + t(x_k - x^*))]\| dt \\ &\leq \int_0^1 M(1-t)\|x_k - x^*\| dt \\ &= \frac{M\|x_k - x^*\|}{2}\end{aligned}$$

Вновь используя условие Липшица для ∇f получаем, что

$$-M\|x_k - x^*\|I \preceq \nabla f(x_k) - \nabla f(x^*) \preceq M\|x_k - x^*\|I,$$

Анализ сходимости

Далее

$$\begin{aligned}\|G_k\| &= \left\| \int_0^1 [\nabla f(x_k) - \nabla f(x^* + t(x_k - x^*))] dt \right\| \\ &\leq \int_0^1 \|[\nabla f(x_k) - \nabla f(x^* + t(x_k - x^*))]\| dt \\ &\leq \int_0^1 M(1-t)\|x_k - x^*\| dt \\ &= \frac{M\|x_k - x^*\|}{2}\end{aligned}$$

Вновь используя условие Липшица для ∇f получаем, что

$$-M\|x_k - x^*\|l \preceq \nabla f(x_k) - \nabla f(x^*) \preceq M\|x_k - x^*\|l,$$

что дает

$$(m - M\|x_k - x^*\|)l \preceq \nabla f(x^*) - M\|x_k - x^*\|l \preceq \nabla f(x_k)$$

Анализ сходимости

Далее

$$\begin{aligned}\|G_k\| &= \left\| \int_0^1 [\nabla f(x_k) - \nabla f(x^* + t(x_k - x^*))] dt \right\| \\ &\leq \int_0^1 \|[\nabla f(x_k) - \nabla f(x^* + t(x_k - x^*))]\| dt \\ &\leq \int_0^1 M(1-t)\|x_k - x^*\| dt \\ &= \frac{M\|x_k - x^*\|}{2}\end{aligned}$$

Вновь используя условие Липшица для ∇f получаем, что

$$-M\|x_k - x^*\|I \preceq \nabla f(x_k) - \nabla f(x^*) \preceq M\|x_k - x^*\|I,$$

что дает

$$(m - M\|x_k - x^*\|)I \preceq \nabla f(x^*) - M\|x_k - x^*\|I \preceq \nabla f(x_k)$$

Если $x_k \in S$, то $M\|x_k - x^*\| - m > 0$, а $\nabla f(x_k)$ обратима и при этом

$$\|[\nabla f(x_k)]^{-1}\| \leq (m - M\|x_k - x^*\|)^{-1}$$

Анализ сходимости

Итог:

$$\|x_{k+1} - x^*\| \leq \frac{M \|x_k - x^*\|^2}{2(m - M \|x_k - x^*\|)}$$

что дает оценку скорости сходимости.

Анализ сходимости

Итог:

$$\|x_{k+1} - x^*\| \leq \frac{M\|x_k - x^*\|^2}{2(m - M\|x_k - x^*\|)}$$

что дает оценку скорости сходимости.

Осталось показать, что $\|x_{k+1} - x^*\| \leq \|x_k - x^*\|$, чтобы гарантировать $x_k \in S$:

$$\begin{aligned} \frac{M\|x_k - x^*\|^2}{2(m - M\|x_k - x^*\|)} &\leq \|x_k - x^*\| \Leftrightarrow \\ M\|x_k - x^*\|^2 &\leq 2(m - M\|x_k - x^*\|)\|x_k - x^*\| \Leftrightarrow \\ 3M\|x_k - x^*\|^2 &\leq 2m\|x_k - x^*\|. \end{aligned}$$

Анализ сходимости

Итог:

$$\|x_{k+1} - x^*\| \leq \frac{M\|x_k - x^*\|^2}{2(m - M\|x_k - x^*\|)}$$

что дает оценку скорости сходимости.

Осталось показать, что $\|x_{k+1} - x^*\| \leq \|x_k - x^*\|$, чтобы гарантировать $x_k \in S$:

$$\begin{aligned} \frac{M\|x_k - x^*\|^2}{2(m - M\|x_k - x^*\|)} &\leq \|x_k - x^*\| \Leftrightarrow \\ M\|x_k - x^*\|^2 &\leq 2(m - M\|x_k - x^*\|)\|x_k - x^*\| \Leftrightarrow \\ 3M\|x_k - x^*\|^2 &\leq 2m\|x_k - x^*\|. \end{aligned}$$

Таким образом, если $\gamma \geq 3/2$, то $\|x_{k+1} - x^*\| \leq \|x_k - x^*\|$, $x_0 \in S$ и $x_k \in S$. ■

Ссылки на литературу

Нестеров Ю. Е. Методы выпуклой оптимизации // параграф 1.2.4

Vandenberghе L. Unconstrained optimization (slides)