Методы оптимизации, ФКН ВШЭ, зима 2017

Семинар 1: скорости сходимости и матричные вычисления.

10 января 2017 г.

1 Скорости сходимости

В теории оптимизации мы хотим решить следующую задачу:

$$f^* = \min_{x \in E} f(x). \tag{1}$$

Функция $f : \mathbb{R}^n \to \mathbb{R}$ предполагается непрерывной и достаточно гладкой (для поиска оптимума мы будем использовать её производные), E — некоторое подмножество в $\mathbb{R}^n$.

Решить задачу (1) — значит найти оптимальное значение функции f^* и точку x^*, в которой это значение достигается:

$$x^* = \operatorname*{argmin}_{x \in E} f(x), \qquad f^* = f(x^*).$$

Но, к сожалению, довольно часто мы не можем решить задачу (1) точно, поэтому приходится довольствоваться *приближённым решением*, которое близко к оптимальному.

Наши методы обычно будут стартовать с некоторой начальной точки x_0 и итерационно строить последовательности:

$$\begin{array}{cccccc} x_0, & x_1, & x_2, & \ldots, & x_k, & \ldots \\ f(x_0), & f(x_1), & f(x_2), & \ldots, & f(x_k), & \ldots \end{array}$$

От любого хорошего метода мы хотим, чтобы последовательность $f(x_k)$ хотя бы сходилась к оптимуму f^* при стремлении числа итераций к бесконечности:

$$r_k \stackrel{\text{def}}{=} f(x_k) - f^*, \qquad \lim_{k \to +\infty} r_k = 0.$$

Но для того, чтобы сравнивать методы между собой, этого недостаточно. Введём понятие скорости сходимости.

Определение 1 *Последовательность $\{r_k \geq 0\}_{k \geq 0}$ называется линейно сходящейся с параметром $q \in (0,1)$, если найдется $C > 0$, что для всех $k \geqslant 0$ выполнено:*

$$r_k \leqslant Cq^k. \tag{2}$$

Нижняя граница тех q, для которых последовательность является линейно сходящейся с параметром q, называется *константой линейной сходимости.*

Определение 2 *Последовательность $\{r_k \geq 0\}_{k \geq 0}$ называется сверхлинейно (суперлинейно) сходящейся, если она является линейно сходящейся с любым параметром $q \in (0,1)$.*

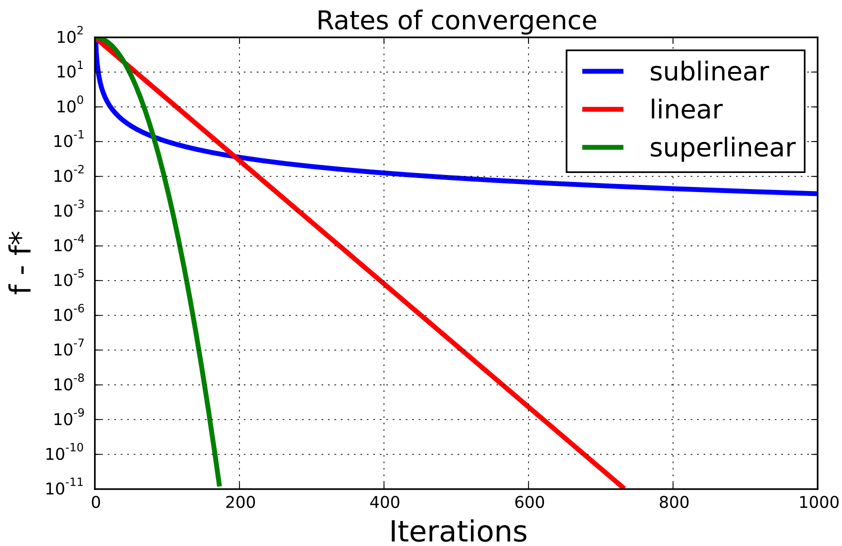
Замечание: Для сверхлинейно сходящейся последовательности константа линейной сходимости равна нулю.

Определение 3 *Последовательность $\{r_k \geq 0\}_{k \geq 0}$, сходящаяся к нулю, но не являющаяся линейно сходящейся (ни для одного $q \in (0,1)$) называется сублинейно сходящейся.*

Таким образом мы определили три вида сходимости:

- сублинейная сходимость

- линейная сходимость
- сверхлинейная сходимость



Удобно устанавливать скорость сходимости с помощью следующего теста:

Утверждение 1 *Пусть $r_k > 0$ для любого k и $r_k \to 0$.*

1. *Если $\limsup\limits_{k\to\infty} \frac{r_{k+1}}{r_k} = q < 1$, то $\{r_k\}$ имеет линейную сходимость с константой q.*

2. *Если $\lim\limits_{k\to\infty} \frac{r_{k+1}}{r_k} = 0$, то $\{r_k\}$ имеет сверхлинейную сходимость.*

3. *Если $\lim\limits_{k\to\infty} \frac{r_{k+1}}{r_k} = 1$, то $\{r_k\}$ имеет сублинейную сходимость.*

4. *В остальных случаях данный тест ничего не говорит о скорости сходимости $\{r_k\}$.*

Важно: скорость сходимости — асимптотическое понятие, константа C в определении может быть очень большой (и зависеть от выбранного q). Сходимость последовательности, начинающаяся с некоторого номера, влечёт ту же скорость сходимости и с самого начала, нужно лишь увеличить C должным образом.

Особый случай сверхлинейной сходимости — последовательность r_k, удовлетворяющая условию:

$$r_{k+1} \leq C r_k^2 \qquad \text{для любого } k \geqslant 0.$$

Такая сходимость называется *квадратичной*.

2 Матричные вычисления

Под вектором $x \in \mathbb{R}^m$ мы будем понимать обычно вектор-столбец. Матрицу $A \in \mathbb{R}^{n\times m}$ можно умножить на такой вектор слева:

$$y = Ax, \qquad y \in \mathbb{R}^n.$$

Операция транспонирования матрицы:

$$A \in \mathbb{R}^{n\times m} \quad \Rightarrow \quad A^T \in \mathbb{R}^{m\times n}, A_{ij}^T = A_{ji}.$$

Если мы хотим умножить матрицу на вектор справа, то вектор необходимо транспонировать:

$$u = v^T A, \qquad v \in \mathbb{R}^n,\ u \in \mathbb{R}^{1\times m}.$$

Также заметим, что для двух векторов одинакового размера $a, b \in \mathbb{R}^n$:

$$a^T b \ = \ \sum_{i=1}^{n} a_i b_i \ \in \mathbb{R} \quad \text{— число,}$$

$$ab^T \ = \ \Big(a_i b_j\Big)_{i,j=1}^{n,n} \ \in \mathbb{R}^{n \times n} \quad \text{— матрица.}$$

Для квадратной матрицы $A \in \mathbb{R}^{n \times n}$ определена операция взятия следа:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} A_{ii},$$

спектральная норма и норма Фробениуса:

$$||A|| = \max_{x \in \mathbb{R}^n, x \neq 0} \frac{x^T A x}{x^T x}, \qquad ||A||_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}^2 = \mathrm{Tr}(A^T A).$$

Некоторые свойства:

- $(AB)^{-1} = B^{-1} A^{-1}$;
- $(AB)^T = B^T A^T$;
- $(A^{-1})^T = (A^T)^{-1} = A^{-T}$;
- $\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB)$;
- $\mathrm{Tr}(A^T) = \mathrm{Tr}(A)$;
- Если $x \in \mathbb{R}$, то $\mathrm{Tr}(x) = x$.

2.1 Собственные числа, спектральное разложение

$A \in \mathbb{R}^{n \times n}$ — квадратная матрица.

Ненулевой вектор $u \in \mathbb{R}^n$ называется *собственным вектором* матрицы A, отвечающим *собственному значению* $\lambda \in \mathbb{C}$, если:

$$Au = \lambda u.$$

Утверждение 2 *Любая матрица имеет n собственных значений, с учётом кратности.*

Утверждение 3 *Если матрица симметричная: $A = A^T$, то все её собственные значения вещественны: $\lambda_i(A) \in \mathbb{R}$.*

Для симметричной матрицы будем нумеровать собственные значения в отсортированном порядке:

$$\lambda_1(A) \ \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_n(A).$$

$\lambda(A) = [\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A)] \in \mathbb{R}^n$ — вектор из всех собственных значений матрицы A с учётом кратности.

Множество всех симметричных матриц размера n будем обозначать $\mathbb{S}^n = \{A \in \mathbb{R}^{n \times n} \,|\, A = A^T\}$.

Утверждение 4 *Если матрица симметричная, то существует ортонормированный базис из её собственных векторов: $u_1, \ldots, u_n \in \mathbb{R}^n$:*

$$\langle u_i, u_j \rangle = u_i^T u_j = [i = j], \qquad Au_i = \lambda_i(A) u_i.$$

Соберём вектора $u_1, \ldots, u_n$ в матрицу U по столбцам.
Ортонормированность базиса означает, что $UU^T = U^T U = I$ — матрица U ортогональная, $U^{-1} = U^T$.
Получаем *спектральное разложение* матрицы, $\Lambda = \mathrm{Diag}(\lambda_1(A), \ldots, \lambda_n(A))$:

$$AU = U\Lambda \quad \Rightarrow \quad A = U \Lambda U^T.$$

2.2 Положительно определённые матрицы

Одним из важных классов симметричных матриц являются положительно определённые матрицы:

Определение 4 *Матрица $A \in \mathbb{S}^n$ называется положительно определённой ($A \succ 0$), если для любого $x \in \mathbb{R}^n, x \neq 0$ выполнено:*

$$x^T A x > 0.$$

Если для любого $x \in \mathbb{R}^n$ выполнено $x^T A x \geqslant 0$, то матрица называется *положительно полуопределённой* (или *неотрицательно определённой*): $A \succeq 0$.

Обозначим:

$$\mathbb{S}^n_+ = \{A \in \mathbb{S}^n \mid A \succeq 0\} \text{ — множество всех положительное полуопределённых матриц,}$$
$$\mathbb{S}^n_{++} = \{A \in \mathbb{S}^n \mid A \succ 0\} \text{ — множество всех положительно определённых матриц.}$$

Ясно, что $\mathbb{S}^n_{++} \subset \mathbb{S}^n_+ \subset \mathbb{S}^n$.

2.3 Линейные уравнения

Для произвольной матрицы размерность линейной оболочки её строк равна размерности линейной оболочки её столбцов.

Это число называется *раангом матрицы:* $\mathrm{rank}(A)$.

Утверждение 5 *Для матрицы $A \in \mathbb{R}^{n\times n}$ существует обратная $\Leftrightarrow$ $rank(A) = n$ $\Leftrightarrow$ $\Leftrightarrow$ $det(A) = \prod_{i=1}^n \lambda_i(A) \neq 0$.*

Для положительно определённой матрицы $A \succ 0$ всегда существует обратная A^{-1}.

Как соотносятся собственные значения A и A^{-1}?

Утверждение 6 $rank(uv^T) = 1$.

Утверждение 7 *Ранг матрицы равен минимальному числу одноранговых матриц, в представлении матрицы через их сумму:*

$$A = \sum_{i=1}^{rank(A)} u_i v_i^T.$$

Если мы решаем систему $Ax = b$, то либо $\mathrm{rank}(A) = n$ и тогда $x = A^{-1}b$, либо $\mathrm{rank}(A) < n$ и мы либо не имеем решений (если b не лежит в образе A), либо имеем целое пространство решений размерности $n-\mathrm{rank}(A)$.

Если решение существует, то всё множество решений представимо в виде $x + \ker(A)$, где x — произвольное решение, $\ker(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$.