Курс “Методы линейной алгебры в теории приближений”. Лекция 2. Неравенства для собственных и сингулярных чисел.

7 октября 2023 г.

Минимаксная формула Wielandt-a На прошлой лекции мы доказали две формулы:

Courant–Fischer: $\lambda_i(A) = \max\limits_{\dim V = i} \min\limits_{v \in V, |v|=1} v^* A v$.

Частичный след: $\lambda_1(A) + \ldots + \lambda_i(A) = \max\limits_{\dim V = i} \mathrm{Tr}(A|_V)$.

Мы докажем формулу Wielandt-a, которая обобщает эти два утверждения. В ней даётся представление для произвольной суммы $\lambda_{i_1}(A) + \ldots + \lambda_{i_k}(A)$. Зафиксируем числа $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$. Частичным флагом называется цепочка вложенных подпространств $V_1 \subset V_2 \subset \ldots \subset V_k$ таких что $\dim V_j = i_j$, $j = 1, \ldots, k$. Многообразием Шуберта $X(V_1, \ldots, V_k)$ называется семейство подпространств W размерности k, таких что $\dim(W \cap V_j) \geqslant j$, $j = 1, \ldots, k$. Формула имеет следующий вид:

$$\lambda_{i_1}(A) + \ldots + \lambda_{i_k}(A) = \max_{\substack{V_1 \subset \ldots \subset V_k \\ \dim V_j = i_j}} \min_{W \in X(V_1, \ldots, V_k)} \mathrm{Tr}(A|_W).$$

Для начала убедимся, что формула Wielandt-a обобщает две предыдущие. При $k = 1$ получаем, очевидно, Куранта–Фишера. При $i_1 = 1$, $i_2 = 2$, $\ldots$, $i_k = k$ подпространство W обязано совпадать с V_k и получается формула частичного следа.

Можно представлять себе многообразие Шуберта следующим образом. Поскольку $\dim(W \cap V_1) \geqslant 1$, найдётся ненулевой вектор $w_1 \in W \cap V_1$. Далее, $W \cap V_2 \cap \mathrm{span}\{w_1\}^\perp \ni w_2$, и т.д. Строится ортонормированный базис $\{w_1, \ldots, w_n\}$ пространства W со свойством $w_j \in V_j$. Наоборот, если есть ортонормированный базис с условием $w_j \in V_j$, то для $W :=$

$\operatorname{span}\{w_1, \ldots, w_k\}$ выполнены требуемые неравенства $\dim(W \cap V_j) \geqslant j$. При этом $\operatorname{Tr}(A|_W) = \sum_{j=1}^k w_j^* A w_j$.

Перейдём к доказательству.

Доказательство. Как обычно, рассматриваем диагонализующий базис $\{f_1, \ldots, f_n\}$. Пример пространств V_j строится легко: $V_j := \operatorname{span}\{f_1, \ldots, f_{i_j}\}$. Пусть $\{w_1, \ldots, w_k\}$ — произвольный ортонормированный базис с $w_j \in V_j$. Тогда $w_j^* A w_j \geqslant \lambda_{i_j}$. Суммируя по j, получаем, что $\min_w \geqslant \sum \lambda_{i_j}$.

Обратно: нужно для любого набора подпространств построить систему $\{w_j\}$, на которой след мал. Проводим индукцию по k. Случай $k = 1$ это Курант-Фишер. Пусть для $k - 1$ доказано, перейдём к k.

Сначала заметим, что можно считать $i_1 = 1$. Действительно, в противном случае перейдём от V_j к $V_j' := V_j \cap \{f_2, \ldots, f_n\}$. Тем самым размерности i_j уменьшатся на единицу, но и собственные значения оператора сдвинутся: вместо $(\lambda_1, \ldots, \lambda_n)$ теперь $(\lambda_2, \ldots, \lambda_n)$. Ясно, что так дело сводится к $i_1 = 1$.

Итак, $V_1 = \operatorname{span}\{w_1\}$. По предположению индукции, для $V_2 \subset \ldots \subset V_k$ есть минимизирующий базис $\{w_2, \ldots, w_k\}$, $w_j \in V_j$, $j = 2, \ldots, k$. Для него

$$\sum_{j=2}^k w_j^* A w_j \leqslant \lambda_{i_2}(A) + \ldots + \lambda_{i_k}(A).$$

Рассмотрим $W := \operatorname{span}\{w_1, \ldots, w_k\}$, он удовлетворяет условию $\dim(W \cap V_j) \geqslant j$, однако, теперь $\{w_j\}$ это не ортонормированный базис. Не страшно — заменим w_1 на элемент $u \in W \cap \operatorname{span}\{w_2, \ldots, w_k\}^\perp$. Тогда $u^* A u \leqslant \lambda_1(A) = \lambda_{i_1}(A)$. Складывая это с предыдущим неравенством, получим нужную оценку следа. $\square$

Неравенства Лидского и Wielandt–Hoffman Неравенство Лидского:

$$\lambda_{i_1}(A+B) + \ldots + \lambda_{i_k}(A+B) \leqslant \lambda_{i_1}(A) + \ldots + \lambda_{i_k}(A) + \lambda_1(B) + \ldots + \lambda_k(B).$$

Оно выводится из формулы Wielandt-а. Действительно, для оценки суммы сверху нужно для любого набора $\{V_j\}$ найти $W \in X(\{V_j\})$ с оценкой сверху на $\operatorname{Tr}(A + B|_W)$. Пусть W — подпространство, минимизирующее $\operatorname{Tr}(A|_W)$. Тогда $\operatorname{Tr}(A|_W) \leqslant \lambda_{i_1}(A) + \ldots + \lambda_{i_k}(A)$. Кроме того, всегда $\operatorname{Tr}(B|_W) \leqslant \lambda_1(B) + \ldots + \lambda_k(B)$. Складывая эти неравенства и пользуясь линейностью следа, получим требуемое.

Следствие неравенства Лидского: для любого набора неотрицательных коэффициентов $c_1, \ldots, c_n$ имеем

$$\sum_{k=1}^{n} c_k \lambda_k(A+B) \leqslant \sum_{k=1}^{n} c_k \lambda_k(A) + \sum_{k=1}^{n} c_k^* \lambda_k(B),$$

где $c_1^* \geqslant c_2^* \geqslant \ldots \geqslant c_n^*$ это невозрастающая перестановка чисел (c_k).

Доказательство. Применим трюк: выразим c_i в виде интеграла $\int_0^\infty \mathbf{1}\{c_i \geqslant \lambda\}\, d\lambda$. Подставим вместо c_i подинтегральное выражение для фиксированного λ. Пусть $c_1^* \geqslant c_2^* \geqslant \ldots \geqslant c_k^* \geqslant \lambda > c_{k+1}^* \geqslant \ldots \geqslant c_n^*$ и пусть индексы c_i, соответствующие c_j^*, $j \leqslant k$, это $i_1, \ldots, i_k$. Тогда $\mathbf{1}\{c_j \geqslant \lambda\} = 1$ если j принадлежит множеству $\{i_1, \ldots, i_k\}$ и 0 в противном случае. Поэтому в левой части получится $\sum_{j=1}^{k} \lambda_{i_j}(A+B)$. В правой части для A всё аналогично, а для B получится сумма первых k собственных чисел. Мы получили в точности неравенство Лидского.

Интегрируя по λ, получим нужное нам неравенство. □

Наконец, докажем неравенство Wielandt–Hoffman:

$$\|\vec{\lambda}(A+B) - \vec{\lambda}(A)\|_{\ell_p^n} \leqslant \|\vec{\lambda}(B)\|_{\ell_p^n}.$$

Запишем предыдущее неравенство для пар (A, B) и $A+B, -B$:

$$\sum_{i=1}^{n} c_i(\lambda_i(A+B) - \lambda_i(A)) \leqslant \sum_{i=1}^{n} c_i^*(\lambda_i(B)), \tag{1}$$

$$\sum_{i=1}^{n} c_i(\lambda_i(A) - \lambda_i(A+B)) \leqslant \sum_{i=1}^{n} c_i^*(\lambda_i(-B)) = \sum_{i=1}^{n} -c_i^* \lambda_{n-i+1}(B)). \tag{2}$$

Разобьём множество индексов на $I^+ := \{i\colon \lambda_i(A+B) \geqslant \lambda_i(B)\}$ и $I^- := \{i\colon \lambda_i(A+B) < \lambda_i(B)\}$. Пусть $|I^+| = s$. Возьмём коэффициенты (a_i), такие что $a_i \geqslant 0$ на I^+ и $a_i = 0$ на I^-. Используем неравенство (1) для a_i:

$$\sum_{i \in I^+} a_i |\lambda_i(A+B) - \lambda_i(A)| \leqslant \sum_{i \in I^+} a_i^* |\lambda_i(B)| = \sum_{i=1}^{s} a_i^* |\lambda_i(B)|.$$

Аналогично, возьмём коэффициенты (b_i) на I^- и используем (2):

$$\sum_{i \in I^-} b_i |\lambda_i(A+B) - \lambda_i(A)| \leqslant \sum_{i \in I^-} b_i^* |\lambda_{n-i+1}(B)| = \sum_{i=1}^{n-s} b_i^* |\lambda_{n-i+1}(B)| = \sum_{i=s+1}^{n} b_{n-i+1}^* |\lambda_i(B)|.$$

Объединим (a_i) и (b_i) в одну последовательность (c_i), тогда:

$$\sum_{i=1}^{n} c_i|\lambda_i(A+B)-\lambda_i(A)| \leqslant \sum_{i=1}^{n} c_i^{**}|\lambda_i(B)|,$$

где последовательность (c_i^{**}) получается так: сначала идёт неубывающая перестановка $\{c_i\colon i \in I^+\}$, потом неубывающая перестановка $\{c_i\colon i \in I^-\}$. Нам не важен порядок c_i^{**}; главное, что это перестановка c_i.

При $p = 1$ достаточно взять $c_i \equiv 1$; при $p = \infty$ взять единицу при максимальном $|\lambda_i(A+B)-\lambda_i(A)|$. В общем случае полагаем $c_i = |\lambda_i(A+B)-\lambda_i(A)|^{p-1}$ и применяем неравенство Гёльдера и всё получается (проверьте!).

Следствие: величина $\|\vec{\lambda}(A)\|_p$ удовлетворяет неравенству треугольника и потому является нормой. Она называется нормой Шаттена.

SVD. Вектор собственных чисел матрицы во многом её характеризует. Недостаток собственных чисел в том, что они определены только для квадратных матриц. В общем случае нужны *сингулярные числа.*

Теорема. (Сингулярное разложение, Singular Value Decomposition, SVD.) Пусть A — произвольная $n \times p$ комплексная матрица. Она представляется в виде $A = U\Sigma V^*$, где U — унитарная матрица $n \times n$; V — унитарная матрица $p \times p$; Σ — прямоугольная матрица, элементы вне диагонали равны нулю, на диагонали стоят неотрицательные числа $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_{\min(n,p)} \geqslant 0$. Эти числа называются *сингуряными числами* матрицы A и они определены однозначно.

Изобразим SVD для “вертикальной” матрицы ($n \geqslant p$):

$$\begin{pmatrix} . & . & . \\ . & . & . \\ . & A & . \\ . & . & . \\ . & . & . \end{pmatrix} = \begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ . & . & U & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} * \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \ldots & 0 \\ 0 & 0 & \sigma_p \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} . & . & . \\ . & V^* & . \\ . & . & . \end{pmatrix}$$

Поскольку элементы последних $n-p$ столбцов матрицы U умножаются на ноль, они не играют роли и можно изобразить усечённое представле-

ние:

$$\begin{pmatrix} . & . & . \\ . & . & . \\ . & A & . \\ . & . & . \\ . & . & . \end{pmatrix} = \begin{pmatrix} . & . & . \\ . & . & . \\ . & \tilde{U} & . \\ . & . & . \\ . & . & . \end{pmatrix} * \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \sigma_p \end{pmatrix} * \begin{pmatrix} . & . & . \\ . & V^* & . \\ . & . & . \end{pmatrix}$$

В операторных терминах (разберём случай $n \geqslant p$): существует ортнорромированный базис $v_1, \dots, v_p \in \mathbb{C}^p$, и ортонормированный базис $u_1, \dots, u_n \in \mathbb{C}^n$, такие что

$$Av_j = \sigma_j u_j, \ A^* u_j = \sigma_j v_j \quad j = 1, \dots, p,$$

и $A^* u_j = 0$ при $j > p$. Это ясно из матричного представления, u_j и v_j суть столбцы матриц U и V.

То есть, как устроено линейное отображение $A\colon \mathbb{C}^p \to \mathbb{C}^p$: сначала это растяжение вдоль осей v_i (т.е. $v_i \mapsto \sigma_i v_i$), потом поворот (при котором $v_i \mapsto u_i$).

Доказательство. Установим единственность σ_j: из представления $A = U\Sigma V^*$ следует, что $A^*A = V\Sigma\Sigma^* V^*$, матрица $\Sigma\Sigma^*$ диагональна с (σ_j^2) на диагонали, ясно что это собственные значения матрицы A^*A, а они определены однозначно.

Теперь ясно, как доказывать существование разложения (пусть $n \geqslant p$). Матрица A^*A эрмитова; пусть $\lambda_1, \dots, \lambda_p$ это её собственные числа и $v_1, \dots, v_p$ — ортонормированный базис из собственных векторов. Собственные числа неотрицательны:

$$\lambda_i \langle v_i, v_i \rangle = \langle A^* A v_i, v_i \rangle = \langle A v_i, A v_i \rangle \geqslant 0.$$

Положим $\sigma_i := \lambda_i^{1/2}$. Вектора v_i уже построены. Положим $u_j := \sigma_j^{-1} A v_j$ при $j \leqslant p$. Стандартным образом проверяется, что эти вектора ортонормированы, дополним их до ортобазиса. Собственно, мы и получили SVD. $\square$

Неравенства для сингулярных чисел Обсудим связь сингулярных и собственных чисел. Из доказательства видно, что

$$\sigma_i(A) = \sqrt{\lambda_i(A^*A)} = \sqrt{\lambda_i(AA^*)}, \quad i = 1, \dots, \min(n, p).$$

Если матрица A эрмитова, то, как нетрудно видеть, $\sigma_i(A) = |\lambda_i(A)|$. Таким образом, сингулярные числа теряют часть информации.

По произвольной матрице можно построить эрмитову так:

$$\begin{pmatrix} . & . & . \\ . & . & . \\ . & A & . \\ . & . & . \\ . & . & . \end{pmatrix} \mapsto \tilde{A} := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & . & . & . \\ 0 & 0 & 0 & 0 & 0 & . & . & . \\ 0 & 0 & 0 & 0 & 0 & . & A & . \\ 0 & 0 & 0 & 0 & 0 & . & . & . \\ 0 & 0 & 0 & 0 & 0 & . & . & . \\ . & . & . & . & . & 0 & 0 & 0 \\ . & . & A^* & . & . & 0 & 0 & 0 \\ . & . & . & . & . & 0 & 0 & 0 \end{pmatrix}$$

Нетрудно проверить, что собственные числа $\tilde{A}$ это $\pm\sigma_j(A)$ (и нули).

Courant–Fischer для сингулярных чисел:

$$\sigma_i(A) = \max_{\dim V = i} \min_{\substack{v \in V, \\ |v|=1}} |Av|.$$

Действительно, возведём в квадрат, слева получим $\lambda_i(A^*A)$, справа будет $|Av|^2 = \langle v, A^*Av\rangle$, используем Куранта–Фишера для A^*A.

Для всякого k сумма $\sigma_1(A) + \ldots + \sigma_k(A)$ является нормой. Называется норма Ки-Фаня (Ky-Fan). Эта норма индуцируется с выпуклого функционала $\lambda_1 + \ldots + \lambda_k$ при вложении $A \mapsto \tilde{A}$.

Величина $\|\vec{\sigma}(A)\|_p$, аналогично, является нормой и называется нормой Шаттена $\|A\|_{S_p}$. Она согласована с прежним определением для эрмитовых матриц. Для сингулярных чисел также справедливы неравенства Лидского:

$$\sigma_{i_1}(A+B) + \ldots + \sigma_{i_k}(A+B) \leqslant \sigma_{i_1}(A) + \ldots + \sigma_{i_k}(A) + \sigma_1(B) + \ldots + \sigma_k(B)$$

и неравенство Wielandt–Hoffman:

$$\|\vec{\sigma}(A+B) - \vec{\sigma}(A)\|_p \leqslant \|\vec{\sigma}(B)\|_p.$$

(Проверьте!)

Задача. На пространстве матриц определено скалярное произведение

$$\langle A, B\rangle := \sum_{i,j} \overline{A}_{i,j} B_{i,j} = \mathrm{Tr}(A^*B).$$

Доказать некоммутативное неравенство Гёльдера: для любого $1 \leqslant p \leqslant \infty$, $1/p + 1/p' = 1$,

$$|\operatorname{Tr}(A^* B)| \leqslant \|A\|_{S_p} \|B\|_{S_{p'}}.$$