Курс “Методы линейной алгебры в теории приближений”. Лекция 5. Гамма-2 норма и аппроксимативный ранг

28 октября 2023 г.

Факторизационная γ_2-норма. Пусть A — вещественная $m \times n$ матрица.

Определение. Гамма-2 нормой матрицы A (обозначение: $\gamma_2(A)$) называется минимальное число C, такое что A можно представить в виде $A_{i,j} = \langle x_i, y_j \rangle$, где $x_1, \ldots, x_m, y_1, \ldots, y_n$ лежат в некотором евклидовом пространстве и $|x_i| \cdot |y_j| \leqslant C$ при всех i, j.

Несколько замечаний. Во-первых, прослеживается аналогия с обычным рангом матрицы, только в нём ограничивается не длина, а размерность векторов. Далее, можно считать, что $x_i, y_j \in \mathbb{R}^{m+n}$ (рассматриваем подпространство, натянутое на эти вектора). Поэтому, в частности, минимум C достигается.

Обобщение. Пусть $T\colon X \to Y$ оператор между нормированными пространствами. Определим факторизационную $\Gamma_2(X, Y)$-норму этого оператора. Рассмотрим факторизацию вида $T = RS$, где $S\colon X \to H$, $R\colon H \to Y$, и H — гильбертово пространство. Ясно, что $\|T\|_{X\to Y} \leqslant \|R\|_{H\to Y} \cdot \|S\|_{X\to H}$. Точная нижняя грань произведений $\|R\| \cdot \|S\|$ и называется Γ_2-нормой T.

Утверждение. $\|T\|_{\Gamma_2(X,Y)} \leqslant \|T\|_{X\to Y} \cdot \sqrt{\operatorname{rank} T}$.

Доказательство. Мы воспользуемся известной теоремой Джона. Она гласит, что для любого выпуклого тела $K \subset \mathbb{R}^n$ его эллипсоид максимального объёма $\mathcal{E}$ (т.е. эллипсоид $\mathcal{E} \subset K$ имеющий максимальный объём) обладает свойством $K \subset n\mathcal{E}$. Нам потребуется случай центрально–

симметричного K, в котором результат сильнее:

$$K \subset n^{1/2}\mathcal{E}.$$

Удобно вести разговор на языке нормированных пространств. Пусть V это n-мерное нормированное пространство (с шаром K). Тогда найдётся оператор $u\colon V \to \ell_2^n$ (можно считать u тождественным, а евклидову норму взять с шаром $\mathcal{E}$), такое что

$$\|u\| \leqslant n^{1/2}, \quad \|u^{-1}\| \leqslant 1. \tag{1}$$

Итак, рассмотрим $L := \mathrm{Im} T \subset Y$. Это n-мерное пространство ($n = \operatorname{rank} T$). Возьмём $u\colon L \to \ell_2^n$ из (1) и факторизуем

$$T = u^{-1}(uT), \quad \|u^{-1}\| \cdot \|uT\| \leqslant \|T\| \cdot \|u\| \cdot \|u^{-1}\| \leqslant n^{1/2}\|T\|.$$

□

Теперь поймём, почему $\Gamma_2(X, Y)$ это обобщение $\gamma_2(A)$. Действительно, посмотрим на A как на оператор $\ell_1^n \to \ell_\infty^m$. Факторизация это представление A в виде $A = BC$. При этом, как мы знаем, $\|C\|_{\ell_1^n \to \ell_2}$ это максимальная евклидова длина столбцов C, а $\|B\|_{\ell_2 \to \ell_\infty^m}$ это максимальная длина строк B. Значит, факторизация даёт представление вида $A_{i,j} = \langle B_i, C^j \rangle$ с соответствующим ограничением на длины, поэтому

$$\gamma_2(A) = \|A\|_{\Gamma_2(\ell_1, \ell_\infty)}.$$

Следствие. $\gamma_2(A) \leqslant \sqrt{\operatorname{rank} K} \cdot \max |A_{i,j}|$.

Положим $\gamma_{2,\varepsilon}(A) := \min\{\gamma_2(B)\colon \max |A_{i,j} - B_{i,j}| \leqslant \varepsilon\}$. Тогда для любой приближающей матрицы B имеем

$$\gamma_{2,\varepsilon}(A) \leqslant \gamma_2(B) \leqslant \sqrt{\operatorname{rank} B}\|B\|_\infty,$$

откуда

$$\operatorname{rank}_\varepsilon(A) \geqslant \frac{\gamma_{2,\varepsilon}(A)^2}{(\|A\|_{1\to\infty} + \varepsilon)^2}.$$

Эта оценка полезна, т.к. величину $\gamma_{2,\varepsilon}$ гораздо проще контролировать, чем $\operatorname{rank}_\varepsilon$ (выпуклая задача).

Неравенство Гротендика. Для дальнейшего нам понадобится

Теорема 1. *(Гротендик) Для любой вещественной матрицы M имеем*

$$\max_{\substack{\{x_i\},\{y_j\}\in\ell_2\\ |x_i|,|y_j|\leqslant 1}} \sum_{i,j} M_{i,j}\langle x_i, y_j\rangle \leqslant K_G \max_{x_i,y_j\in[-1,1]} \sum_{i,j} M_{i,j} x_i y_j,$$

где K_G — абсолютная постоянная.

Замечания. В левой части берём максимум по векторам в гильбертовом пространстве, но, опять-таки, можно ограничиться $\mathbb{R}^{m+n}$. В левой части можно заменить максимум по $x_i, y_j \in [-1,1]$ на максимум по $x_i, y_j \in \{-1,1\}$, поскольку максимум по выпуклому множеству (кубу) достигается в крайних точках (вершинах). Число K_G называется константой Гротендика, оно хорошо изучено, $K_G \approx 1.4$.

Доказательство. Иногда удобно реализовать гильбертово пространство как гауссово пространство. Т.е. мы считаем, что x_i, y_j это гауссовы случайные величины ξ_i, η_j, а скалярное произведение это $E\xi\eta$. (Формально: рассмотрим стандартные независимые случайные величины $Z_1, \ldots, Z_N$ и N-мерное пространство, состоящее из их линейных комбинаций, тогда $\langle \xi, \eta\rangle = \mathsf{E}\xi\eta$ задаёт на нём структуру скалярного произведения; как уже было сказано, нам достаточно брать x_i, y_j из такого пространства для $N = m+n$; хотя можно и $N = \infty$.)

Итак, нужно оценить $\sum M_{i,j}\mathsf{E}\xi_i\eta_j$, если известно что $\|\xi_i\| = (\mathsf{E}\xi_i^2)^{1/2} \leqslant 1$, $\|\eta_j\| \leqslant 1$. Пусть $K_{m,n}$ это максимум таких сумм по всем матрицам M размера $m\times n$, для которых $\sum_{x_i,y_j\in[-1,1]} M_{i,j}x_iy_j \leqslant 1$. Требуется доказать, что числа $K_{m,n}$ равномерно ограничены.

Выберем абсолютную постоянную L такую что $\mathsf{E}|\xi|^2\mathbf{1}\{|\xi| > L\} < 1/100$ для стандартной нормальной ξ. Пусть $\xi^L = \xi\mathbf{1}\{|\xi| \leqslant L\}$ это "срезка". Тогда

$$\sum M_{i,j}\mathsf{E}\xi_i\eta_j = \mathsf{E}\sum M_{i,j}\xi_i^L\eta_j^L + \sum M_{i,j}\mathsf{E}\xi_i^L(\eta_j - \eta_j^L) + \sum M_{i,j}\mathsf{E}(\xi_i - \xi_i^L)\eta_j.$$

Рассмотрим первое слагаемое. Оно поточечно не превосходит L^2, поэтому и мат. ожидание $\leqslant L^2$. Второе слагаемое не превосходит $K_{m,n}/100$, поскольку норма первого сомножителя всегда не больше единицы, а второго не больше $1/100$. Аналогично, третье слагаемое не больше $K_{m,n}/100$. Отсюда $K_{m,n} \leqslant L^2 + K_{m,n}/50$, ч.т.д. $\square$

Неравенство Гротендика позволяет найти двойственную норму

$$\gamma_2^*(A) := \max_{\gamma_2(B)\leqslant 1} \langle A, B\rangle, \quad \text{где } \langle A, B\rangle := \sum A_{i,j}B_{i,j}.$$

Условие $\gamma_2(B) \leqslant 1$ эквивалентно тому, что $B_{i,j} = \langle x_i, y_j\rangle$ для векторов в единичном евклидовом шаре, поэтому

$$\gamma_2^*(A) := \max_{|x_i|,|y_j|\leqslant 1} \sum A_{i,j}x_iy_j.$$

В силу неравенства Гротендика,

$$\|A\|_{\infty\to 1} \leqslant \max_{|x_i|,|y_j|\leqslant 1} \sum A_{i,j}x_iy_j \leqslant K_G \max_{x_i,y_j\in[-1,1]} \sum A_{i,j}x_iy_j = K_G\|A\|_{\infty\to 1}.$$

Окончательно, $\gamma_2^*(A) \asymp \|A\|_{\infty\to 1}$.

Оценка n-членных приближений. Напомним определение. Пусть D — множество в нормированном пространстве X (“словарь”), $x \in X$. Величиной наилучшего n-членного приближения x по словарю D называется

$$\sigma_n(x, D) := \inf_{g_1,\dots,g_n\in D} \inf_{c_1,\dots,c_n\in\mathbb{R}} \|x - \sum_{k=1}^{n} c_kg_k\|.$$

При оценке n-членных приближений ключевую роль играет $A_1(D)$-норма. Единичный шар в этой норме это просто $\overline{\operatorname{conv}(D \cup (-D))}$, т.е. он состоит из векторов, представимых (с любой точностью) в виде $\sum c_kg_k$, $\sum |c_k| \leqslant 1$.

Утверждение. Пусть X — линейное пространство, $Y \subset X^*$ конечное множество, $M := |Y| \geqslant 2$. Рассмотрим норму

$$\|x\|_Y := \max_{y\in Y} |\langle x, y\rangle|.$$

Пусть словарь $D \subset X$ таков что $\|g\|_Y \leqslant 1$ для всех $g \in D$. Тогда

$$\sigma_n(x, D)_Y \leqslant Cn^{-1/2} \log^{1/2}(M) \cdot \|x\|_{A_1(D)}. \tag{2}$$

Доказательство. Пусть $\|x\|_{A_1(D)} \leqslant 1$, тогда $x \approx \sum_{k=1}^{N} c_kg_k$, $g_k \in D$, $\sum |c_k| \leqslant 1$. Рассмотрим случайный элемент в X:

$$\begin{cases} \mathsf{P}(\xi = g_k) = |c_k|, \\ \mathsf{P}(\xi = 0) = 1 - \sum |c_k|. \end{cases}$$

Тогда $\mathsf{E}\xi = x$. Рассмотрим независимые копии ξ и усредним их: $\eta = \frac{1}{n}(\xi^{(1)} + \ldots + \xi^{(n)})$. Тогда для любого $y \in Y$ имеем по неравенству Хедфинга:

$$\mathsf{P}(|\langle \eta, y\rangle - \langle x, y\rangle| \geqslant \lambda) \leqslant 2\exp(-\frac{2(n\lambda)^2}{4n}) = 2\exp(-n\lambda^2/2).$$

При $n\lambda^2 > \ln M$ эта величина будет меньше $1/M$ и с положительной вероятностью $\|\eta - x\|_Y \leqslant \lambda$. Осталось заметить, что η это n-членая комбинация элементов словаря. При этом $\lambda \asymp n^{-1/2}\log^{1/2} M$. $\square$

Применим это утверждение в следующей ситуации. X это пространство матриц, Y состоит из всех функционалов $A \mapsto A_{i,j}$; словарь D состоит из одноранговых матриц с элементами из $[-1, 1]$. Что будет нормой $A_1(D)$? Проще найти двойственную норму:

$$\|A\|_{A_1(D)^*} = \sup_{\|B\|_{A_1(D)} \leqslant 1} \langle A, B\rangle = \sup_{\substack{B = \sum c_k u_k v_k^t \\ \sum |c_k| \leqslant 1, \|u_k\|_\infty, \|v_k\|_\infty \leqslant 1}} \sum c_k \langle A, u_k v_k^t\rangle.$$

Ясно, что для последняя сумма максимизируется когда $\langle A, u_k v_k^t\rangle$ максимально возможно и соотвествующий вес $c_k = 1$,

$$\|A\|_{A_1(D)^*} = \max_{\|u\|_\infty, \|v\|_\infty \leqslant 1} \langle A, uv^t\rangle = \max_{u_i, v_j \in [-1,1]} \sum A_{i,j} u_i v_j = \|A\|_{\infty\to 1}.$$

Итак, $\|A\|_{A_1(D)^*} = \|A\|_{\infty\to 1}$, поэтому $\|A\|_{A_1(D)} = \|A\|^*_{\infty\to 1} \asymp \gamma_2(A)$. Итого из (2) получаем оценку для поэлементного приближения матрицы.

Утверждение. Пусть A — вещественная $m \times n$ матрица. Тогда

$$E(A, \mathcal{R}_r)_{1\to\infty} \leqslant Cr^{-1/2}\log^{1/2}(mn) \cdot \gamma_2(A).$$