Спецкурс 2020/2021: “Геометрические и комбинаторные свойства матриц и аппроксимация” Блок лекций “Сложность матриц и аппроксимация” Лекция 4: “Факторизационная γ_2-норма”

24 ноября 2020 г.

Определение γ_2

Пусть $M \in \mathbb{R}^{m\times n}$. Через $\gamma_2(M)$ обозначим точную нижнюю грань $c > 0$, таких что M представляется в виде $M = AB$, причём для любой строки a_i матрицы A и любого столбца b^j матрицы B имеем $|a_i| \cdot |b^j| \leqslant c$.

Другими словами, $\gamma_2(M) \leqslant c$ тогда и только тогда, когда найдутся вектора $x_1, \dots, x_m$ и $y_1, \dots, y_n$ в некотором конечномерном евклидовом пространстве, такие что $M_{i,j} = \langle x_i, y_j\rangle$ для всех i, j, и $|x_i| \cdot |y_j| \leqslant c$. (Можно перенормировать так, что $\max |x_i|, |y_j| \leqslant \sqrt{c}$.)

Матрица имеет малый ранг, когда её элементы представимы в виде скалярного произведения *маломерных* векторов: $\operatorname{rank} M \leqslant r \Leftrightarrow M_{i,j} = \langle x_i, y_j\rangle,\ x_i, y_j \in \mathbb{R}^r$; в случае же малой γ_2-нормы элементы представимы в виде скалярного произведения *коротких* векторов. Это говорит о тесной связи ранга и γ_2-нормы.

Точная нижняя грань в определении γ_2 достигается. Действительно, можно считать, что размерность пространства, в котором лежат x_i, y_j, не превосходит $m + n$.

Величина γ_2 задаёт норму в пространстве матриц $\mathrm{Mat}_{m\times n}(\mathbb{R})$.
Почему $\gamma_2(M+K) \leqslant \gamma_2(M) + \gamma_2(K)$?

Данное определение есть частный случай γ_2-нормы в пространстве операторов между банаховыми пространствами X и Y. А именно, $\gamma_2(u)$ для оператора $u\colon X \to Y$ есть точная нижняя грань $c > 0$, таких что u представляется в виде $u = AB$, где $B\colon X \to H$, $A\colon H \to Y$, H – некоторое гильбертово пространство, $\|A\| \cdot \|B\| \leqslant c$.

Через $\Gamma_2(X, Y)$ обозначается банахово пространство операторов $X \to Y$ с конечной γ_2-нормой.

$\gamma_2(M)$ для матрицы равна γ_2-норме M как оператора из $\Gamma_2(\ell_1^n, \ell_\infty^m)$. Действительно, матричная норма $\|B\|_{1\to 2}$ равна чему? максимальной длине столбца матрицы B, а норма $\|A\|_{2\to\infty}$ равна максимальной длине строки матрицы A (почему?).

Из теоремы Джона вытекает следующее неравенство для оператора $u\colon X \to Y$ конечного ранга:

Statement

Для оператора $u\colon X \to Y$ имеем

$$\gamma_2(u) \leqslant \sqrt{\operatorname{rank} u} \cdot \|u\|_{X\to Y}.$$

Действительно, пусть $V = \mathrm{Im}\, u \subset Y$, $\dim V = r := \operatorname{rank} u$.
По теореме Джона, для эллипсоида максимального объёма $\mathcal{E}$, вписанного в шар $B_V = B_Y \cap V$,

$$\mathcal{E} \subset B_V \subset \sqrt{r}\mathcal{E}.$$

Пространство E с шаром $\mathcal{E}$ евклидово, $E \cong \ell_2^r$. Тождественные

операторы $f\colon (V, \|\cdot\|_Y) \to E$ и $g\colon E \to (V, \|\cdot\|_Y)$ взаимно обратны и $\|f\| \leqslant \sqrt{r}$, $\|g\| \leqslant 1$. (Иногда теорему Джона так и формулируют: существует оператор $f\colon Z \to \ell_2^{\dim Z}$ с $\|f\| \cdot \|f^{-1}\| \leqslant \sqrt{\dim Z}$.)

Тогда $u = g(fu)$,

$$\|fu\| \cdot \|g\| \leqslant \|u\| \cdot \|f\| \cdot \|g\| \leqslant \sqrt{r}\|u\|.$$

Итак,
$$\gamma_2(u) \leqslant \sqrt{\operatorname{rank} u} \cdot \|u\|_{X\to Y}.$$

Применим это утверждение для обычной γ_2-нормы матрицы M, то есть нормы $\Gamma(\ell_1^n, \ell_\infty^m)$. Чему равно $\|M\|_{1\to\infty}$?

Следствие.
$$\gamma_2(M) \leqslant \sqrt{\operatorname{rank} M} \max_{i,j} |M_{i,j}|.$$

- Чему равно $\gamma_2(\mathrm{Id})$ для единичной матрицы? $\gamma_2(\mathrm{Id}) = 1$.
- Позже мы докажем, что $\gamma_2(\Delta^N) \asymp \log N$ для верхнетреугольной матрицы из нулей и единиц.
- Для случайных сигнум $N \times N$ матриц $\gamma_2 \asymp N^{1/2}$.

γ_2-норма тесно связана с аппроксимативным рангом и *margin complexity* (мерой сложности, основанной на реализации сигнум-матриц с большим отступом). Об этом – в следующих лекциях.

Представление $M_{i,j} = \langle x_i, y_j \rangle$, $x_i, y_j \in H$, наводит на мысли о неравенстве Гротендика:

Theorem (Grothendieck's inequality)

Для любой матрицы $M \in \mathbb{R}^{m \times n}$ имеет место неравенство

$$\max_{x_i, y_j \in U(H)} \sum_{i,j} M_{i,j} \langle x_i, y_j \rangle \leqslant K_G \max_{|s_i|, |t_j| \leqslant 1} \sum_{i,j} M_{i,j} s_i t_j$$

где $U(H)$ – единичный шар в гильбертовом пространстве H, а K_G — абсолютная постоянная.

Число K_G называется *константой Гротендика*, $1.5 < K_G < 1.8$. Для комплексных матриц $1.3 < K_G^{\mathbb{C}} < 1.5$.

$$\max_{x_i, y_j \in U(H)} \sum_{i,j} M_{i,j} \langle x_i, y_j \rangle \leqslant K_G \max_{|s_i|,|t_j| \leqslant 1} \sum_{i,j} M_{i,j} s_i t_j.$$

Величина в левой части неравенства есть максимум $\sum_{i,j} M_{i,j} A_{i,j} = \langle M, A \rangle$ по всем матрицам $\gamma_2(A) \leqslant 1$. Следовательно, этот максимум есть ни что иное как сопряжённая норма $\gamma_2^*(M)$.

Максимум в правой части неравенства Гротендика равен $\|M\|_{\infty \to 1}$. Следовательно,

$$\|M\|_{\infty \to 1} \leqslant \gamma_2^*(M) \leqslant K_G \|M\|_{\infty \to 1}.$$

Доказательство теоремы Гротендика

Рассмотрим величину

$$K_{m,n} = \max\{\gamma_2^*(M)\colon M \in \mathbb{R}^{m\times n},\ \|M\|_{\infty\to 1} \leqslant 1\},$$

$$\gamma_2^*(M) := \sup_{x_i, y_j \in U(H)} |\sum M_{i,j}\langle x_i, y_j\rangle|.$$

Нужно доказать, что $K_{m,n}$ равномерно ограничена.
Зафиксируем матрицу M нормы $\|M\|_{\infty\to 1} \leqslant 1$, оценим $\gamma_2^*(M)$.
Мы можем считать, что H — Гауссово Гильбертово пространство: элементы $\xi \in H$ суть гауссовы случайные величины с нулевым средним $\mathsf{E}\xi = 0$, и

$$\|\xi\|_H^2 := \mathsf{E}\xi^2.$$

Эквивалентное условие: $\langle \xi, \eta\rangle_H = \mathsf{E}\xi\eta$.

Почему так можно: рассмотрим последовательность $Z_1, \dots, Z_n, \dots$ независимых стандартных гауссовых величин, положим

$$H = \{\xi = \sum_{k=1}^{\infty} c_k Z_k\colon c \in \ell_2\}, \quad \|\xi\|_H^2 = \|c\|_{\ell_2}^2 = \mathsf{E}\xi^2.$$

Доказательство теоремы Гротендика (продолжение)

Фиксируем $\delta \in (0, 1/2)$. Пользуясь тем, что все $\xi \in H$ имеют нормальное распределение, можем найти $L = L(\delta)$, такое что

$$\mathsf{E}|\xi|^2 \mathbf{1}_{\{|\xi|>L\}} \leqslant \delta^2 \quad \text{для } \xi \in U(H).$$

Таким образом, ξ представляется в виде равномерно ограниченной *срезки* $\xi^L := \xi \mathbf{1}_{\{|\xi|\leqslant L\}}$ и величины $\xi \mathbf{1}_{\{|\xi|>L\}}$ малой нормы.

Теперь запишем

$$\sum M_{i,j}\langle \xi_i, \eta_j \rangle = \sum M_{i,j}\langle \xi_i^L + (\xi_i - \xi_i^L), \eta_j \rangle =$$

$$= \sum M_{i,j}\langle \xi_i^L, \eta_j^L \rangle + \sum M_{i,j}\langle \xi_i^L, \eta_j - \eta_j^L \rangle + \sum M_{i,j}\langle \xi_i - \xi_i^L, \eta_j \rangle.$$

Доказательство теоремы Гротендика (окончание)

Оценим первое слагаемое:

$$\sum M_{i,j}\langle \xi_i^L, \eta_j^L\rangle = \mathsf{E}\sum M_{i,j}\xi_i^L\eta_j^L.$$

В каждой точке имеем $|\xi_i^L(\omega)|, |\eta_j^L(\omega)| \leqslant L$, значит, поточечно

$$\sum M_{i,j}\xi_i^L(\omega)\eta_j^L(\omega) \leqslant \|M\|_{\infty\to 1}L^2 \leqslant L^2.$$

Поэтому

$$\mathsf{E}\sum M_{i,j}\xi_i^L\eta_j^L \leqslant L^2.$$

Остальные слагаемые легко оценить:

$$|\sum M_{i,j}\langle \xi_i^L, \eta_j - \eta_j^L\rangle| \leqslant \delta K_{m,n},$$

$$|\sum M_{i,j}\langle \xi_i - \xi_i^L, \eta_j\rangle| \leqslant \delta K_{m,n}.$$

Отсюда $\gamma_2^*(M) \leqslant L^2 + 2\delta K_{m,n}$. Так как M произвольна, то

$$K_{m,n} \leqslant L^2 + 2\delta K_{m,n}, \quad K_{m,n} \leqslant \frac{L^2}{1-2\delta}.$$

Мультипликатор Шура

Фиксируем матрицу M. Рассмотрим *мультипликатор Шура*: оператор в пространстве матриц, действующий по формуле

$$A \mapsto A \circ M,$$

где $\circ$ – произведение Шура, т.е. $(A \circ M)_{i,j} = A_{i,j} M_{i,j}$.
Спектральная норма в пространстве матриц индуцирует норму

$$\|S_M\| := \max_{\|A\|_{2\to 2} \leqslant 1} \|A \circ M\|_{2\to 2}.$$

Theorem

Норма мультипликатора Шура равна γ_2-норме матрицы:

$$\|S_M\| = \gamma_2(M).$$

Мы приведём доказательство, основанное на технике полуопределённого программирования (*Semidefinite programming*, SDP), следуя статье T.Lee, A.Shraibman, R.Spalek (2008).

Полуопределённая оптимизация – частный случай выпуклой оптимизации. Напомним принцип Лагранжа для выпуклых задач – теорему Каруша-Куна-Таккера.

Пусть дана выпуклая задача

$$\begin{cases} f_0(x) \to \min, \\ f_j(x) \leqslant 0,\ j = 1, \ldots, m, \\ x \in \mathcal{A}, \end{cases} \qquad (*)$$

где все f_j – выпуклые функции, $\mathcal{A} \subset \mathbb{R}^d$ – выпуклое множество. Предположим также, что выполнено условие Слейтера: $\exists x_0 \in \mathcal{A}\colon f_j(x_0) < 0,\ j = 1, \ldots, m.$

Theorem (KKT)

Вектор $\hat{x} \in \mathcal{A}$ является решением задачи (*) тогда и только тогда, когда существуют множители Лагранжа $\hat{\lambda} = (1, \hat{\lambda}_1, \ldots, \hat{\lambda}_m)$, $\hat{\lambda}_j \geqslant 0$, такие что $\lambda_j(\hat{x}) f_j(\hat{x}) = 0$, $j = 1, \ldots, m$, и функция Лагранжа

$$L(x, \hat{\lambda}) = \sum_{j=0}^{m} \hat{\lambda}_j f_j(x)$$

достигает минимума в $\hat{x}$.

Пара $(\hat{x}, \hat{\lambda})$ является седловой точкой:

$$L(\hat{x}, \lambda) \leqslant L(\hat{x}, \hat{\lambda}) \leqslant L(x, \hat{\lambda}), \quad x \in \mathcal{A},\ \lambda = (1, \lambda_1, \ldots, \lambda_m),\ \lambda_j \geqslant 0.$$

Последнее означает, что $\hat{\lambda}$ является решением двойственной задачи:

$$\begin{cases} g(\lambda) := \min_{x \in \mathcal{A}} L(x, \lambda) \to \max, \\ \lambda = (1, \lambda_1, \ldots, \lambda_m),\ \lambda_j \geqslant 0, \end{cases}$$

причём значения этих задач совпадают (и равны $L(\hat{x}, \hat{\lambda})$).

Напомним: $\langle A, B\rangle = \sum A_{i,j}B_{i,j} = \mathrm{tr}(AB^t)$.
Следующая простая лемма лежит в основе SDP:

Statement

Пусть $A = A^t$. Тогда

$$\inf_{X \geqslant 0} \langle A, X\rangle = \begin{cases} 0, & A \geqslant 0, \\ -\infty, & \text{иначе.} \end{cases}$$

Докажем это утверждение. Если $A \not\geqslant 0$ и $x^tAx < 0$ для некоторого вектора x, то взяв $X = cxx^t$ и устремив $c \to +\infty$, мы получим

$$\langle A, X\rangle = c\,\mathrm{tr}(Axx^t) = cx^tAx \to -\infty.$$

Обратно, если $A \geqslant 0$ и $X \geqslant 0$, то $\langle A, X\rangle \geqslant 0$. Следовательно, минимум достигается при $X = 0$.

Типичная задача SDP записывается следующим образом:

$$\begin{cases} f_0(X) := \langle C, X \rangle \to \min, \\ f_j(X) := \langle A_j, X \rangle - b_j \leqslant 0,\ j = 1, \ldots, m, \\ X \geqslant 0. \end{cases}$$

Параметры задачи: симметричные матрицы C, A_j и вектор b. Выписывая двойственную задачу, получаем выражение

$$g(\lambda) = \min_{X \geqslant 0} \langle C + \sum_{j=1}^{m} \lambda_j A_j, X \rangle - \sum_{j=1}^{m} \lambda_j b_j.$$

Применяя доказанную лемму, видим, что для того, чтобы $g \neq -\infty$, необходимо, чтобы матрица $C + \sum_{j=1}^{m} \lambda_j A_j$ была неотрицательно определённой; при этом $g(\lambda) = -\sum_{j=1}^{m} \lambda_j b_j$. Отсюда получаем следующую двойственную задачу:

$$-\sum_{j=1}^{m} b_j \lambda_j \to \max, \quad C + \sum_{j=1}^{m} \lambda_j A_j \geqslant 0, \quad \lambda_j \geqslant 0,\ j = 1, \ldots, m.$$

Перейдём к доказательству теоремы. Сведём всё к симметричным матрицам. (Наша матрица M размера $m \times n$ может даже не быть квадратной!) Применим следующий приём для симметризации:

$$\widehat{M} := \begin{pmatrix} 0_m & M \\ M^t & 0_n \end{pmatrix}.$$

Здесь 0_n и 0_m – квадратные нулевые матрицы соответствующего размера. Ясно, что $\widehat{M}$ симметрична и имеет размер $(n+m) \times (n+m)$.

Нам также пригодится матрица $\widehat{1}_{m,n}$ – симметризация матрицы, состоящей из одних единиц. Через Id обозначаем единичную матрицу (тождественный оператор).

Теперь мы можем приступить к доказательству равенства $\gamma_2(M) = \max_{\|A\| \leqslant 1} \|A \circ M\|$. По определению, $\gamma_2(M) \leqslant \eta$ тогда и только тогда, когда найдутся вектора $x_1, \ldots, x_m$ и $y_1, \ldots, y_n$ длины не более $\sqrt{\eta}$, такие что $M_{i,j} = \langle x_i, y_j \rangle$.

Рассмотрим матрицу Грама X системы $(x_1, \ldots, x_m, y_1, \ldots, y_n)$: мы видим, что $X_{i,i} \leqslant \eta$ для всех i, и что $X \circ \widehat{1}_{m,n} = \widehat{M}$.

Отсюда ясно, что γ_2-норма матрицы M равна значению следующей SDP-задачи:

$$\begin{cases} \eta \to \min, \\ X_{i,i} \leqslant \eta, \ i = 1, \ldots, m+n, \\ X \geqslant 0, \\ X \circ \widehat{1}_{m,n} = \widehat{M}. \end{cases} \qquad (\star)$$

Оптимизация происходит по паре (η, X).

Далее мы составляем функцию Лагранжа. Условиям $X_{i,j} = M_{i,j}$ соответствуют множители $q_{i,j}$ (для (i,j) из носителя $\widehat{1}_{m,n}$). Вместе они образуют матрицу $Q = (q_{i,j})$ такую что $Q^t = Q$ и $Q \circ \widehat{1}_{m,n} = Q$. Итого:

$$L = \eta + \sum_{i=1}^{m+n} \lambda_i(\langle e_i e_i^t, X\rangle - \eta) + \langle Q, \widehat{M} - X\rangle, \quad Q \circ \widehat{1}_{m,n} = Q,\ Q^t = Q.$$

(Условий $q_{i,j} \geqslant 0$ нет, т.к. это множители при ограничениях вида равенства.) Приходим к двойственной задаче:

$$\begin{cases} \langle Q, \widehat{M}\rangle \to \max, \\ \operatorname{diag}(\lambda) - Q \geqslant 0, \\ Q \circ \widehat{1}_{m,n} = Q, \\ \lambda_j \geqslant 0,\ \sum \lambda_j = 1. \end{cases} \tag{$\star\star$}$$

(Условие $\sum \lambda_j = 1$ возникает из-за выражения $\eta(1 - \sum \lambda_j)$, если множитель не равен нулю, то $\min\limits_{\eta, X \geqslant 0} = -\infty$.)

Рассмотрим задачу $(\star\star)$:

$$\langle Q, \widehat{M}\rangle \to \max, \quad \mathrm{diag}\,(\lambda) - Q \geqslant 0, \quad Q \circ \widehat{1}_{m,n} = Q,\ \lambda_j \geqslant 0, \sum \lambda_j = 1.$$

Поскольку $\mathrm{diag}\,(\lambda) \geqslant Q$, то если какой-то $\lambda_j = 0$, то j-й столбец и j-я строка матрицы Q нулевые; мы можем их "вычеркнуть" т.к. они не влияют на значение задачи; можно считать, что все $\lambda_j > 0$.

Положим $Q'_{i,j} := Q_{i,j}\lambda_i^{-1/2}\lambda_j^{-1/2}$. Тогда условие $\mathrm{diag}\,(\lambda) \geqslant Q$ равносильно условию $\mathrm{Id} \geqslant Q'$. По построению Q' имеет вид $Q' = \widehat{S}$ для некоторой матрицы S.

Упражнение. Докажите, что следующие неравенства равносильны:

- $\mathrm{Id} \geqslant \widehat{S}$;
- $\|S\| \leqslant 1$;
- $\|\widehat{S}\| \leqslant 1$.

Итак, условие $\mathrm{Id} \geqslant Q'$ равносильно условию $\|S\| \leqslant 1$.

Наконец, пусть $\alpha_j = \lambda_j^{1/2}$, тогда целевая функция задачи принимает вид

$$\langle Q, \widehat{M} \rangle = \langle Q' \circ \alpha\alpha^t, \widehat{M} \rangle = \alpha^t (\widehat{S} \circ \widehat{M})\alpha.$$

При этом условие на λ означает что α – единичный вектор с неотрицательными координатами.
Максимизируя по α, получаем

$$\max_{|\alpha|=1, \alpha_j \geqslant 0} \alpha^t (\widehat{S} \circ \widehat{M})\alpha = \max_{|\beta|=|\gamma|=1, \beta_i \geqslant 0, \gamma_j \geqslant 0} \beta^t (S \circ M)\gamma.$$

Осталось максимизировать по S: $\|S\| \leqslant 1$; при этом условие неотрицательности координат β, γ пропадёт, т.к. умножение строки или столбца S на минус единицу не меняет норму. Итак, значение задачи равно

$$\max_{\|S\| \leqslant 1, |\beta|=|\gamma|=1} \beta^t (S \circ M)\gamma = \max_{\|S\| \leqslant 1} \|S \circ M\|.$$

Мы доказали, что γ_2-норма равна значению SDP-задачи ($\star$), а норма оператора Шура равна значению двойственной задачи ($\star\star$). Поскольку условие Слейтера выполнено, эти значения совпадают и теорема доказана!

T.Lee, A.Shraibman, R.Spalek, "A Direct Product Theorem for Discrepancy", 2008.

D.J.H. Garling, *Inequalities: A Journey into Linear Analysis*. 2007.