Основы матричных вычислений. ПМИ. Семинар 2.
Нормальные матрицы. Положительная определённость. SVD-разложение.

Весна 2025. Н.Медведь (по материалам М.Рахубы, Л.Высоцкого и др.)

1 Задачи для семинара

Обсуждение 1 (Критерий эрмитовости)**.** Пусть матрица A такова, что для любого $x \in \mathbb{C}^n$ выполнено $x^*Ax \geqslant 0$ (в частности, это число вещественно). Докажите, что A эрмитова, то есть $A = A^*$.

Пояснение: получаем, что в комплексном случае в определении положительной определённости не обязательно включать эрмитовость.

Решение: разложим матрицу A в сумму $\frac{A+A^*}{2} + i\frac{A-A^*}{2i}$. Обозначим $H = \frac{A+A^*}{2}$ и $K = \frac{A-A^*}{2i}$. Из этих формул видно, что $H^* = H$ и $K^* = K$. Тогда для любого x имеем $\overline{x^*Hx} = x^T\overline{H}\overline{x} = (x^T\overline{H}\overline{x})^T = x^*Hx$, то есть число x^*Hx вещественно. Аналогично для K. Тогда $x^*Ax = (x^*Hx) + i(x^*Kx)$, откуда из положительной определённости получаем, что второе слагаемое равно нулю.

Если для всех $x \in \mathbb{C}^n$ выполнено $x^*Kx = 0$, то на самом деле $K = 0$. Действительно, $K = U\Lambda U^*$, откуда легко видим $\Lambda = 0$.

Тогда из $A = H + iK$ получаем просто $A = H$, что и требовалось доказать.

Обсуждение 2 (Некритерий симметричности)**.** Пусть матрица A с вещественными элементами такова, что для любого $x \in \mathbb{R}^n$ выполнено $x^TAx \geqslant 0$. Тогда из этого не следует, что $A = A^T$.

Контрпример: пусть $B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Тогда с одной стороны эта матрица, очевидно, не симметрична, а с другой стороны $x^TBx = (x_1 x_2)\begin{pmatrix} x_2 \\ -x_1 \end{pmatrix} = x_1x_2 - x_2x_1 = 0$ для всех x. Тогда в качестве матрицы A рассмотрим $A = I + B$. Получим положительно определённую несимметричную матрицу.

Видно, что в качестве B можно взять поворот не на $\pi/2$, а на острый угол, тогда скалярное произведение повёрнутого вектора с исходным будет положительно.

Обсуждение 3 (Явная формула для фробениусовой нормы и матричной 2-нормы)**.** Пусть дана произвольная матрица A. У неё существует SVD-разложение $A = U\Sigma V^*$ с унитарными U и V. Тогда по свойствам унитарных матриц $\|U\Sigma V^*\|_2 = \|\Sigma\|_2$ и $\|U\Sigma V^*\|_F = \|\Sigma\|_F$.

Для фробениусовой нормы остаётся заметить, что $\|\Sigma\|_F = \sqrt{\sigma_1(A)^2 + \sigma_2(A)^2 + \ldots + \sigma_r(A)^2}$.

Для 2-нормы нужно найти супремум.

$$\|\Sigma\|_2 = \sup_{x\neq 0} \frac{\|\Sigma x\|_2}{\|x\|_2} = \sup_{x\neq 0} \frac{\sqrt{\sigma_1^2|x_1|^2 + \ldots + \sigma_r^2|x_r|^2}}{\sqrt{|x_1|^2 + \ldots + |x_r|^2}} \leqslant \sup_{x\neq 0} \frac{\sqrt{\sigma_1^2|x_1|^2 + \ldots + \sigma_1^2|x_r|^2}}{\sqrt{|x_1|^2 + \ldots + |x_r|^2}} =$$

$$= \sigma_1 \sup_{x\neq 0} \frac{\sqrt{|x_1|^2 + \ldots + |x_r|^2}}{\sqrt{|x_1|^2 + \ldots + |x_r|^2}} = \sigma_1$$

При этом очевидно, что в месте где мы один раз написали неравенство, оно точное: оно достигается на $(1, 0, \ldots, 0)$.

Задача 4 (Пример неустойчивости наивного алгоритма поиска svd)**.** Рассмотрим матрицу $\begin{pmatrix} 1 & 1 \\ 0 & \sqrt{\varepsilon} \end{pmatrix}$, где ε меньше машинного ноля ε_{mach}, но $\sqrt{\varepsilon}$ уже не столь мало.

Тогда $A^*A = \begin{pmatrix} 1 & 0 \\ 1 & \sqrt{\varepsilon} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & \sqrt{\varepsilon} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1+\varepsilon \end{pmatrix}$, откуда

$$\det(A^*A - \lambda I) = \begin{vmatrix} 1-\lambda & 1 \\ 1 & 1+\varepsilon-\lambda \end{vmatrix} = \lambda^2 - (2+\varepsilon)\lambda + \varepsilon = 0$$

$$\lambda_{1,2} = \frac{2+\varepsilon \pm \sqrt{\varepsilon^2+4\varepsilon+4-4\varepsilon}}{2} = \frac{2+\varepsilon \pm \sqrt{\varepsilon^2+4}}{2} = \frac{2+\varepsilon \pm 2\sqrt{1+(\varepsilon/2)^2}}{2} = \frac{2+\varepsilon \pm (2+O(\varepsilon^2))}{2}$$

$$(\text{Ряд для корня:} \sqrt{1+x} = 1 + \frac{x}{2} + O(x^2))$$

Таким образом, видим, что сингулярные значения — корни из полученных чисел — равны $\sqrt{2+\varepsilon/2+O(\varepsilon^2)} = \sqrt{2}+O(\varepsilon)$ и $\sqrt{\varepsilon/2+O(\varepsilon^2)} = \sqrt{\varepsilon/2}+O(\varepsilon^{3/2})$. Грубо говоря, хотелось бы, чтобы с машинной точностью получались ответы $\sqrt{2}$ и $\sqrt{\varepsilon/2}$.

Но при вычислении A^*A мы можем получить много разного с точностью до машинного нуля, в том числе, например, $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Легко посчитать, что её сингулярные значения — $\sqrt{2}$ и 0. Таким образом, одно из сингулярных значений может измениться как минимум на величину порядка $\sqrt{\varepsilon/2}$ (возможно, если как-то по-другому манипулировать клетками матрицы, можно и ещё ухудшить результат?).

Обсуждение 5 ((Не)единственность svd)**.** Если останется время, кратко обсуждаю, что при построении svd при совпадении сингулярных чисел появляется свобода; для иллюстрации достаточно единичной матрицы. И наоборот — для несовпадающих свободы нет.

Обсуждение 6 (Ядерная норма)**.** Напомню, что в лекции упоминалась *ядерная норма* (nuclear norm = trace norm = Ky Fan norm):

$$\|A\|_* = \|A\|_{1,Shatten} = \sigma_1(A) + \dots + \sigma_r(A).$$

Давайте поймём, почему это норма. Неочевидно только неравенство треугольника: $\|A+B\|_* \leqslant \|A\|_* + \|B\|_*$. Давайте введём матрицы $\widetilde{A} = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$ и $\widetilde{B} = \begin{bmatrix} 0 & B \\ B^* & 0 \end{bmatrix}$.

Заметим, что первые n (от большего к меньшим) собственных значений матрицы $\widetilde{A}$ равны сингулярным числам матрицы A. Действительно, если $A = U\Sigma V^*$, то $\widetilde{A} = \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \begin{bmatrix} 0 & \Sigma \\ \Sigma^* & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix}^*$ (проверьте!), откуда остаётся найти собственные числа матрицы $\begin{bmatrix} 0 & \Sigma \\ \Sigma^* & 0 \end{bmatrix}$, что несложно.

Таким образом, остаётся доказать, что

$$\lambda_1(\widetilde{A}+\widetilde{B}) + \dots + \lambda_k(\widetilde{A}+\widetilde{B}) \leqslant \lambda_1(\widetilde{A}) + \dots + \lambda_k(\widetilde{A}) + \lambda_1(\widetilde{B}) + \dots + \lambda_k(\widetilde{B}),$$

где $\widetilde{A}$ и $\widetilde{B}$ эрмитовы (важно!). Нам достаточно случая $k = n$, но проще доказать для произвольного k.

Лемма (принцип максимума Ки–Фана): Для эрмитовой матрицы A выполнено равенство $\sum\limits_{j=1}^{k} \lambda_j(A) = \max \sum\limits_{j=1}^{k} x_j^T A x_j$, где максимум ищется среди всех возможных ортонормированных наборов $x_1, \ldots, x_k$.

Вариант этого утверждения будет доказан ближе к концу курса на лекциях, а пока можно воспринимать это как небольшой взгляд в будущее. Я не буду доказывать это утверждение строго, а поясню его смысл в вещественном случае. Пусть матрица A симметрична и положительно определена (положительная определённость нам тоже не дана, но на самом деле она в доказательстве и не нужна), тогда про неё можно думать как про матрицу квадратичной формы. Тогда её можно привести ортогональной заменой к главным осям. Тогда λ_j – это во сколько раз A как оператор растягивает j-тый базисный вектор в этих координатах, откуда скалярное произведение $x_j^T A x_j = (x_j, Ax_j)$ равно λ_j. Поэтому для x_j – ортов главных осей – выполнено равенство. Остаётся объяснить, почему это максимальное значение.

Почему из принципа максимума Ки–Фана следует наше неравенство

$$\lambda_1(\widetilde{A}+\widetilde{B}) + \cdots + \lambda_k(\widetilde{A}+\widetilde{B}) \leqslant \lambda_1(\widetilde{A}) + \cdots + \lambda_k(\widetilde{A}) + \lambda_1(\widetilde{B}) + \cdots + \lambda_k(\widetilde{B})?$$

Потому что его можно переписать как

$$\max \sum_{j=1}^{k} x_j^T (\widetilde{A}+\widetilde{B}) x_j \leqslant \max \sum_{j=1}^{k} x_j^T \widetilde{A} x_j + \max \sum_{j=1}^{k} x_j^T \widetilde{B} x_j,$$

но если на каком-то наборе векторов выражение слева достигает максимум, то по линейности выражение справа на этом же наборе векторов принимает то же значение:

$$\sum_{j=1}^{k} x_j^T (\widetilde{A}+\widetilde{B}) x_j \leqslant \sum_{j=1}^{k} x_j^T \widetilde{A} x_j + \sum_{j=1}^{k} x_j^T \widetilde{B} x_j.$$

Тогда максимумы справа могут быть только ещё больше.