Лекция 2. Малоранговые аппроксимации 1

М. В. Рахуба

21 января 2023 г.

1 Нормальные матрицы

Прошлую лекцию мы закончили на разложении Шура, которое как используется при построении алгоритмов вычислительной линейной алгебры, так и помогает для теоретического анализа. Первое, в чем нам поможет разложение Шура, — это без труда описать множество всех унитарно диагонализуемых матриц, то есть матриц, которые записываются в виде

$$A = U\Lambda U^*,$$

где U — унитарная, Λ — диагональная.

Определение 1. *Матрица A, удовлетворяющая условию*

$$AA^* = A^*A,$$

называется нормальной.

Утверждение 1. *Матрица является унитарно диагонализумой тогда и только тогда, когда она нормальная.*

Доказательство. Пусть $A = U\Lambda U^*$, тогда $AA^* = U\Lambda\Lambda^*U^*$ и $A^*A = U\Lambda^*\Lambda U^*$. Получаем, что $AA^* = A^*A$ в силу того, что любые две диагональные матрицы коммутируют.

Докажем в другую сторону. Запишем разложение Шура $A = UTU^*$. Из условия нормальности получаем, что должно выполняться:

$$TT^* = T^*T,$$

то есть верхнетреугольная T должна быть нормальной. Но так может быть только в случае, если T — диагональная. Действительно, запишем T в виде:

$$T = \begin{bmatrix} t_1 & s \\ 0 & T_1 \end{bmatrix}, \quad t_1 \in \mathbb{C}.$$

Тогда

$$TT^* = \begin{bmatrix} |t_1|^2 + \|s\|_2^2 & \star \\ \star & T_1T_1^* \end{bmatrix}, \quad T^*T = \begin{bmatrix} |t_1|^2 & \star \\ \star & s^*s + T_1^*T_1. \end{bmatrix},$$

Приравнивая $|t_1|^2 + \|s\|_2^2 = |t_1|^2$, получаем $s = 0$. Далее по индукции. $\square$

Приведем примеры нормальных матриц:

- эрмитовы: $A = A^*$;
- унитарные: $A^{-1} = A^*$;
- косоэрмитовы: $A = -A^*$.

Не стоит думать, что нормальные матрицы ограничиваются перечисленными классами. Например, нам в курсе еще встретятся циркулянтные матрицы, которые также являются нормальными, но вообще говоря могут не быть ни эрмитовыми, ни унитарными, ни косоэрмитовыми.

Из Утверждения 1 получаем, что все эти типы матриц являются унитарно диагонализуемыми. Для эрмитовых матриц мы получили результат аналогичный действительному случаю (любая симметричная действительная матрица диагонализуется с помощью некоторой ортогональной матрицы).

Следствие 1. *Обозначим за $\lambda(A)$ множество всех собственных значений матрицы A.*

- A *— эрмитова* $\iff$ A *— нормальна и* $\lambda(A) \subset \mathbb{R}$*;*
- A *— унитарная* $\iff$ A *— нормальна и* $|\lambda| = 1$*,* $\forall \lambda \in \Lambda(A)$*;*
- A *— косоэрмитова* $\iff$ A *— нормальна и* $\lambda(A) \subset i\mathbb{R}$*.*

Докажите Следствие 1 в качестве упражнения.

2 Знакоопределенные матрицы

Определение 2.

1. *Эрмитова* $A \in \mathbb{F}^{n\times n}$*,* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *называется положительно определенной, если*
$$x^* A x > 0 \quad \forall x \in \mathbb{F}^n \setminus \{0\}. \tag{1}$$
2. *Эрмитова* $A \in \mathbb{F}^{n\times n}$*,* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *называется неотрицательно определенной (положительно полуопределенной), если*
$$x^* A x \geq 0 \quad \forall x \in \mathbb{F}^n. \tag{2}$$

Если для какого-то x, выражение $x^* A x$ содержит ненулевую мнимую часть, то такая матрица не может быть знакоопределенной.

Аналогично определяется отрицательная определенность и отрицательная полуопределенность.

Замечание 1. *Можно проверить (материал семинара), что все матрицы, удовлетвроющие* (1) *или* (2) *над* $\mathbb{C}$ *являются эрмитовыми. То есть в комплексном случае требование эрмитовости является избыточным. Однако для* $\mathbb{R}$ *это уже не так. Действительно,*
$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \Longrightarrow \quad x^\top A x = 0 \quad \forall x \in \mathbb{R}^2,$$
то есть матрица является неотрицательно определенной, но не является симметричной.

Несмотря на то, что понятие знакоопределенности часто дается только для симметричных/эрмитовых матриц, все равно принято говорить “симметричная/эрмитова положительно определенная матрица”.

Матрица A^*A является матрицей Грама столбцов $A = [a_1, \dots, a_n]$.

Пример 1. *Пусть $A \in \mathbb{C}^{m\times n}$. Рассмотрим матрицу*

$$G = A^*A$$

и обсудим ее свойства.

1. *G — неотрицательно определенная. Действительно,*

$$x^*Gx = x^*(A^*A)x = (Ax)^*Ax = \|Ax\|_2^2 \geq 0 \quad \forall x \in \mathbb{C}^n.$$

2. *Из знакоопределенности и Замечания 1 сразу следует эрмитовость, однако ее можно проверить и явно:*

$$G^* = (A^*A)^* = A^*A^{*^*} = A^*A = G.$$

Утверждение 2. *Собственные значения неотрицательно (положительно) определенной матрицы являются неотрицательными (положительными).*

Доказательство. Пусть λ_i, v_i — собственная пара неотрицательно определенной матрицы A, причем $\|v_i\|_2 = 1$. Тогда

$$v_i^*Av_i = \lambda_i v_i^* v_i = \lambda_i.$$

Но из неотрицательной определенности матрицы следует, что $v_i^*Av_i \geq 0$, а значит и $\lambda_i \geq 0$. Для положительно определенной матрицы показывается аналогично. □

3 Сингулярное разложение (SVD)

Сформулируем теорему о существовании сингулярного разложения (singular value decomposition, SVD) для произвольной комплексной матрицы.

Теорема 1. *Пусть $A \in \mathbb{C}^{m\times n}$ имеет ранг r. Тогда существуют унитарные матрицы $U \in \mathbb{C}^{m\times m}$ и $V \in \mathbb{C}^{n\times n}$ и числа $\sigma_1 \geq \dots \geq \sigma_r > 0$, такие что:*

$$A = U\Sigma V^*,$$

где $\Sigma \in \mathbb{R}^{m\times n}$ является диагональной матрицей:

$$\Sigma = \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}, \quad \Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r\times r}.$$

Определение 3. *Столбцы матриц U и V будем называть соответственно левыми и правыми сингулярными векторами матрицы A. Числа (включая нулевые), стоящие на главной диагонали Σ будем называть сингулярными числами.*

Доказательство.

1. A^*A — эрмитова матрица, а значит, унитарно диагонализуема, то есть найдется унитарная $V \in \mathbb{C}^{n\times n}$:
$$A^*A = V\Lambda V^*,$$
где Λ – диагональная матрица с собственными значениями матрицы A^*A на диагонали.

   Из Утверждения 2 следует, что все собственные числа A^*A неотрицательны, а значит, мы можем записать $\Lambda = \Sigma_n^2$ с некоторой диагональной матрицей Σ_n с неотрицательными числами на диагонали. Отсюда имеем:
$$V^*(A^*A)V = \Sigma_n^2, \quad \Sigma_n = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}, \quad \sigma_1 \geq \dots \geq \sigma_n \geq 0. \qquad (3)$$

Путем перестановки собственных векторов в матрице V мы всегда можем добиться желаемой очередности собственных значений.

2. Пусть $r\colon \sigma_r > 0, \sigma_{r+1} = 0$. Введем обозначение
$$\Sigma_r = \mathrm{diag}(\sigma_1, \dots, \sigma_r).$$
Так как V — унитарна, то каждый из ее r первых столбцов будет ортогонален каждому из $n-r$ последних. Обозначим
$$V = \Big[\underbrace{V_r}_{r} \quad \underbrace{V_r^\perp}_{n-r}\Big].$$
Перепишем равенство (3) в блочном виде с учетом новых обозначений:
$$\begin{bmatrix} V_r^*A^*AV_r & \star \\ \star & V_r^{\perp *}A^*AV_r^\perp \end{bmatrix} = \begin{bmatrix} \Sigma_r^2 & O \\ O & O \end{bmatrix}$$
Приравняем соответствующие блоки.

Если окажется, что $\sigma_n \neq 0$, то положим $r = n$.

Левый верхний блок:
$$V_r^*A^*AV_r = \Sigma_r^2$$
$$\left(\Sigma_r^{-1}V_r^*A^*\right)\underbrace{\left(AV_r\Sigma_r^{-1}\right)}_{U_r} = I$$
$$AV_r\Sigma_r^{-1} = U_r$$
$$AV_r = U_r\Sigma_r$$

Правый нижний блок:
$$V_r^{\perp *}A^*AV_r^\perp = O$$
$$\underbrace{\mathrm{Tr}\left(V_r^{\perp *}A^*AV_r^\perp\right)}_{\|AV_r^\perp\|_F^2} = 0$$
$$AV_r^\perp = O$$

Отсюда получаем:
$$AV = A\begin{bmatrix} V_r & V_r^\perp \end{bmatrix} = \begin{bmatrix} AV_r & AV_r^\perp \end{bmatrix} =$$
$$= \begin{bmatrix} U_r\Sigma_r & O \end{bmatrix} = \begin{bmatrix} U_r & U_r^\perp \end{bmatrix}\begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix},$$

где $U_r^\perp$ получена путем произвольного дополнения U_r до ортогонального базиса. Таким образом, $AV = U\Sigma$, откуда $A = U\Sigma V^{-1} = U\Sigma V^*$.

3. Найдем, чему равно r. Имеем $\Sigma = U^{-1}A(V^*)^{-1} = U^*AV$, а значит, $r = \text{rank}(A)$.

□

Замечание 2. *В случае $A \in \mathbb{R}^{m\times n}$ также существует сингулярное разложение*

$$A = U\Sigma V^\top,$$

с вещественными ортогональными матрицами U, V.

Напомним, что вещественные унитарные матрицы называются ортогональными.

Перемножая $A = U\Sigma V^*$ и $A^* = V\Sigma^* U^*$, получим собственное разложения матриц A^*A и AA^*:

$$A^*A = V \begin{bmatrix} \Sigma_r^2 & O \\ O & O \end{bmatrix}_{n\times n} V^*,$$

$$AA^* = U \begin{bmatrix} \Sigma_r^2 & O \\ O & O \end{bmatrix}_{m\times m} U^*.$$

Из полученных выражений и единственности собственных значений матриц следует единственность сингулярных чисел. Более того,

$$\sigma_i = \sqrt{\lambda_i(A^*A)} = \sqrt{\lambda_i(AA^*)}, \quad i = 1, \dots, r. \tag{4}$$

Несложно убедиться, что для нормальных матриц (4) сводится к $\sigma_i = |\lambda_i(A)|$. Однако важно помнить, что эта формула перестает работать для произвольных матриц.

Для сингулярных чисел с номером больше r справедливо:

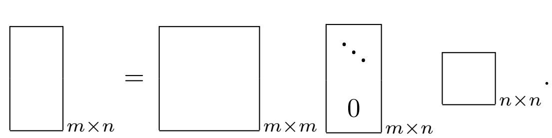
$$\sigma_i = 0, \quad i = r+1, \dots, \min(m,n).$$

Также полезно иметь в виду соотношения между левыми и правыми сингулярными векторами:

$$\begin{aligned} AV = U\Sigma &\implies \\ A^*U = V\Sigma^* &\implies \end{aligned} \boxed{\begin{aligned} Av_i &= \sigma_i u_i, \quad i \le r, \\ A^*u_i &= \sigma_i v_i, \quad i \le r. \end{aligned}}$$

4 *Виды записи SVD*

Существует несколько вариантов записи SVD в зависимости от размеров матриц. Разложение, рассмотренное ранее в Теореме 1, называется **полным SVD**. В случае $m \ge n$ имеет вид:



У разложения в такой форме есть недостаток при сильном отличии m и n. Например, если требуется разложить матрицу A размера 1000×2, то при использовании полного SVD пришлось бы формировать матрицу U размера 1000×1000, чего хотелось бы избежать. По этой причине на практике часто работают с **тонким (thin) SVD**:

$$\boxed{}_{m\times n} = \boxed{}_{m\times K} \boxed{\ddots}_{K\times K} \boxed{}_{K\times n}, \quad K = \min(n, m). \tag{5}$$

Такой вид сразу следует из полного SVD, если обратить внимание на нулевой блок Σ[n:,:n] при $m > n$. В numpy по умолчанию возвращается полное SVD. Чтобы получить тонкое SVD, необходимо “выключить” аргумент `full_matrices`:

```
np.linalg.svd(A, full_matrices=False).
```

В случае, если несколько последних сингулярных чисел являются нулевым, мы имеем возможность еще сильнее уменьшить размеры матриц, входящих в тонкое SVD. А именно, мы можем отбросим соответствующее число последних столбцов в U и V. Получим $A = U_r\Sigma_r V_r^*$ или визуально (5) с $K = r$. Такое разложение называется **компактным SVD**.

Если же строить малоранговое приближение с рангом $k < r$ на основе SVD (смотри Теорему 2), то такое приближение будем называть **усеченным SVD**.

5 Наивный алгоритм поиска SVD

Для конкретики в этой секции будем использовать $m \geq n$. Доказательство Теоремы 1 дает нам конструктивный алгоритм для построения SVD: найти V и Σ через собственное разложение AA^*, затем найти $U_r = AV_r\Sigma_r^{-1}$. Если требуется полное SVD, то U_r необходимо дополнить до ортогонального базиса. Такой алгоритм поиска SVD будем называть наивным алгоритмом, так как, как мы увидим далее, он плохо подходит для вычислений.

Если бы у нас был случай $m \leq n$, то имело бы смысл рассматривать матрицу A^*A вместо AA^*.

Приведем формальное описание наивного алгоритма поиска сингулярного разложения с подсчетом числа арифметических операций с плавающей точкой (FLOPs).

1. Вычислим $A^*A \in \mathbb{C}^{n\times n}$. Эта операция требует $\mathcal{O}(mn^2)$ FLOPs, так как вычисление каждого элемента A^*A является ска-

лярным произведением векторов длины m.

$$n\Big[\underset{m}{\qquad\quad}\Big]\left[\quad\;\right]$$

2. Диагонализация A^*A:

$$A^*A = V\begin{bmatrix}\Sigma_r^2 & 0\\ 0 & 0\end{bmatrix}V^*$$

— нашли V и Σ_r. Позже в курсе мы узнаем, что поиск собственных всех значений и собственных векторов диагонализуемой матрицы требует $\mathcal{O}(n^3)$ FLOPs.

3. $U_r = AV_r\Sigma_r^{-1}$ (нам нужны только первые r столбцов U). Требует $\mathcal{O}(mnr) = \mathcal{O}(mn^2)$, так как $r \le n$.

Итоговая сложность алгоритма: $\mathcal{O}(mn^2)$ или $\mathcal{O}(mn\,\text{min}(m,n))$ в общем случае. Однако у описанного алгоритма возникают нежелательные вычислительные особенности: можно потерять корень из машинной точности. Действительно, можно показать[1] (см. семинар), что при $|\varepsilon| < 1$ наименьшее собственное значение матрицы

$$A = \begin{bmatrix}1 & 1\\ 0 & \sqrt{\varepsilon}\end{bmatrix}$$

ведет себя как $\sigma_2 = \sqrt{\varepsilon/2} + \mathcal{O}(\varepsilon^{3/2})$, в то время как при $|\varepsilon| \le \varepsilon_{\text{m}}$ (ε_{m} — машинное эпсилон) и вычислениях с плавающей точкой мы можем получить вырожденную матрицу, то есть матрицу с сингулярным числом точно равным 0:

$$A^*A = \begin{bmatrix}1 & 1\\ 1 & 1+\varepsilon\end{bmatrix} \overset{\text{float. point}}{=} \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},$$

а значит, ошибка в наименьшем сингулярном числе оказалась равной $\approx \sqrt{\varepsilon_{\text{m}}/2}$. Например, если $\varepsilon_{\text{m}} \approx 10^{-16}$, то $\sqrt{\varepsilon_{\text{m}}/2} \approx 10^{-8}$, что в 10^8 раз больше ε_{m}. Нам бы хотелось поддерживать точность порядка ε_{m}. Позже в курсе мы узнаем подходы для борьбы с этой проблемой.

[1] J. W. Demmel. *Applied numerical linear algebra.* SIAM, 1997

6 *Теорема Эккарта-Янга-Мирского*

Для формулировки теоремы Эккарта-Янга-Мирского нам необходимо ввести понятие унитарно-инвариантной нормы.

Определение 4. *Норма $\|\cdot\|$ называется унитарно-инвариантной, если для любой матрицы $A \in \mathbb{C}^{m\times n}$ и для любых унитарных матриц $U \in \mathbb{C}^{m\times m}$, $V \in \mathbb{C}^{n\times n}$, выполняется*

$$\|UAV\| = \|A\|.$$

Очевидно, что любая унитарно-инвариантная норма является функцией сингулярных чисел матрицы, так как $\|A\| = \|U\Sigma V\| = \|\Sigma\|$. Оказывается, что p-нормы вектора сингулярных чисел являются матричными нормами и называются нормами Шаттена:

$$\|A\|_{p,Shatten} = \left(\sigma_1^p(A) + \cdots + \sigma_r^p(A)\right)^{1/p}, \quad p \geq 1.$$

На прошлой лекции мы с вами показали унитарную инвариантность двух частных случаев Шаттеновских норм: $\|\cdot\|_2$ и $\|\cdot\|_F$. Действительно, используя сингулярное разложение получим:

$$\begin{aligned}
\|A\|_2 &= \|\Sigma\|_2 = \sigma_1(A),\\
\|A\|_F &= \|\Sigma\|_F = \sqrt{\sigma_1^2(A) + \ \ldots \ + \sigma_r^2(A)}.
\end{aligned}$$

Шаттеновскую норму при $p=1$ называют ядерной нормой (nuclear norm) и обозначают:

$$\|A\|_* = \sigma_1(A) + \ \ldots \ + \sigma_r(A).$$

На практике часто минимизацию ранга заменяют именно на минимизацию ядерной нормы.

Теорема 2 (Эккарта-Янга-Мирского). *Пусть $A = U\Sigma V^*$ — сингулярное разложение матрицы A ранга r. Зафиксируем $k \leq r$ и обозначим за U_k и V_k матрицы соответственно первых k левых и правых сингулярных векторов A, а за Σ_k — ведущую $k\times k$ подматрицу матрицы Σ. Пусть также $A_k = U_k \Sigma_k V_k^*$ (усеченное SVD). Тогда*

$$\min_{B:\mathrm{rank}(B)\leq k} \|A - B\| = \|A - A_k\|$$

для любой унитарно-инвариантной нормы $\|\cdot\|$.

Теорема приводится без доказательства.

Отметим, что для введенных унитарно-инвариантных норм очевидно получаются выражения для ошибки $\|A - A_k\|$:

$$\begin{aligned}
\|A - A_k\|_2 &= \sigma_{k+1}(A),\\
\|A - A_k\|_F &= \sqrt{\sigma_{k+1}^2(A) + \cdots + \sigma_r^2(A)},\\
\|A - A_k\|_* &= \sigma_{k+1}(A) + \cdots + \sigma_r(A).
\end{aligned}$$

Список литературы

[1] J. W. Demmel. *Applied numerical linear algebra*. SIAM, 1997.

[2] Е. Е. Тыртышников. *Методы численного анализа*. Академия, 2007.

[3] Е. Е. Тыртышников. *Матричный анализ и линейная алгебра*. Физматлит, 2007.

[4] R. A. Horn and Ch. R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[5] G.H. Golub and C.F. Van Loan. *Matrix Computations, 4th Edition*. The Johns Hopkins University Press, Baltimore, 2013.