Тема III. Линейные операторы

§ 6. Сингулярное разложение

М.В.Волков

Уральский федеральный университет
Институт естественных наук и математики
кафедра алгебры и фундаментальной информатики

2020/2021 учебный год

Постановка задачи

На прошлой лекции мы обсуждали полезное разложение линейных операторов на пространстве со скалярным произведением (*полярное разложение*). На матричном языке оно отвечает определенному разложению квадратных матриц.

Сейчас рассмотрим ситуацию, когда имеются два пространства со скалярным произведением, U и V, и линейный оператор $\mathcal{A}\colon U \to V$.

Мы укажем некоторое разложение для матриц таких линейных операторов (*сингулярное разложение*). Заметим, что размеры матриц в этом случае произвольны.

Нам потребуются некоторые результаты осеннего семестра, которые мы сейчас повторим.

Теорема (Эрик Ивар Фредгольм, 1903)

Если $\mathcal{A}\colon U \to V$ – линейный оператор пространств со скалярным произведением над полем $F \in \{\mathbb{R}, \mathbb{C}\}$, то $(\operatorname{Im} \mathcal{A})^{\perp} = \operatorname{Ker} \mathcal{A}^*$.

Доказательство. Пусть $\mathbf{y} \in \operatorname{Ker} \mathcal{A}^*$, т.е. $\mathcal{A}^*\mathbf{y} = \mathbf{0}$. Чтобы доказать, что $\mathbf{y} \in (\operatorname{Im} \mathcal{A})^{\perp}$, возьмем любой вектор $\mathbf{x} \in \operatorname{Im} \mathcal{A}$ и проверим, что $\mathbf{y} \perp \mathbf{x}$. Поскольку $\mathbf{x} = \mathcal{A}\mathbf{u}$ для некоторого вектора $\mathbf{u} \in U$, имеем

$$\mathbf{x}\mathbf{y} = \mathcal{A}\mathbf{u}\mathbf{y} = \mathbf{u}\mathcal{A}^*\mathbf{y} = 0.$$

Обратно, пусть $\mathbf{y} \in (\operatorname{Im} \mathcal{A})^{\perp}$; тогда $\mathbf{x}\mathbf{y} = 0$ для любого вектора $\mathbf{x} \in \operatorname{Im} \mathcal{A}$. Поэтому для произвольного вектора $\mathbf{u} \in U$ имеем

$$0 = \mathcal{A}\mathbf{u}\mathbf{y} = \mathbf{u}\mathcal{A}^*\mathbf{y}.$$

Вектор $\mathcal{A}^*\mathbf{y}$ ортогонален произвольному вектору $\mathbf{u} \in U$, и потому он нулевой. Отсюда $\mathbf{y} \in \operatorname{Ker} \mathcal{A}^*$. $\square$

Изоморфизм подпространств $\mathrm{Im}\,\mathcal{A}^*$ и $\mathrm{Im}\,\mathcal{A}$

Теорема Фредгольма дает ортогональное разложение пространства V:

$$V = \mathrm{Ker}\,\mathcal{A}^* \oplus \mathrm{Im}\,\mathcal{A}.$$

Применяя ту же теорему к сопряженному оператору $\mathcal{A}^* : V \to U$, получим ортогональное разложение пространства U:

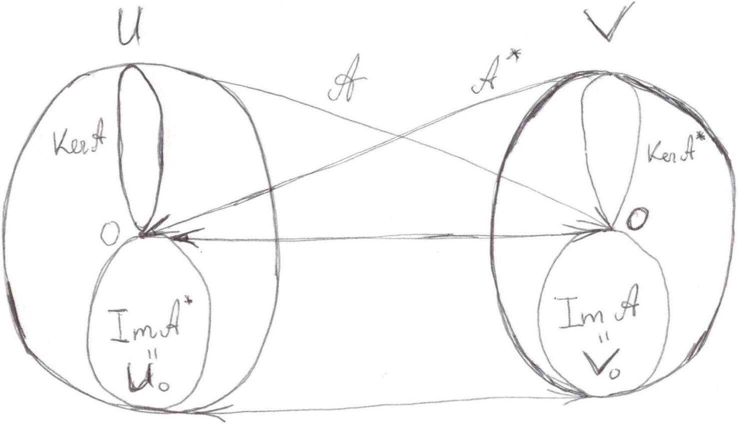
$$U = \mathrm{Ker}\,\mathcal{A} \oplus \mathrm{Im}\,\mathcal{A}^*.$$

Положим $U_0 := \mathrm{Im}\,\mathcal{A}^*$, $V_0 := \mathrm{Im}\,\mathcal{A}$ и обозначим через $\mathcal{A}_0$ *ограничение* оператора $\mathcal{A}$ на подпространство U_0. Это означает, что вектор $\mathcal{A}_0\mathbf{x}$ определен, только если $\mathbf{x} \in U_0$, и в этом случае $\mathcal{A}_0\mathbf{x} := \mathcal{A}\mathbf{x}$.

Предложение

$\mathcal{A}_0$ — изоморфизм пространства U_0 на пространство V_0.

Конфигурация из предложения



Изоморфизм подпространств $\mathrm{Im}\,\mathcal{A}^*$ и $\mathrm{Im}\,\mathcal{A}$ (2)

Доказательство. Проверим, что оператор $\mathcal{A}_0$ взаимно однозначен. Пусть $\mathcal{A}_0\mathbf{x}_1 = \mathcal{A}_0\mathbf{x}_2$ для $\mathbf{x}_1, \mathbf{x}_2 \in U_0$. Тогда $\mathcal{A}_0(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$, откуда $\mathbf{x}_1 - \mathbf{x}_2 \in \mathrm{Ker}\,\mathcal{A}$. Но $\mathrm{Ker}\,\mathcal{A} \cap U_0 = \{\mathbf{0}\}$, поэтому $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0}$, т.е. $\mathbf{x}_1 = \mathbf{x}_2$.

Проверим, что $\mathcal{A}_0$ отображает U_0 на V_0. Возьмем произвольный $\mathbf{y} \in V_0$. Поскольку $V_0 = \mathrm{Im}\,\mathcal{A}$, найдется вектор $\mathbf{x} \in U$ такой, что $\mathbf{y} = \mathcal{A}\mathbf{x}$. Пользуясь ортогональным разложением $U = \mathrm{Ker}\,\mathcal{A} \oplus U_0$, представим $\mathbf{x}$ как $\mathbf{x} = \mathbf{x}_0 + \mathbf{z}$, где $\mathbf{x}_0 \in U_0$, а $\mathbf{z} \in \mathrm{Ker}\,\mathcal{A}$. Тогда

$$\mathbf{y} = \mathcal{A}\mathbf{x} = \mathcal{A}(\mathbf{x}_0 + \mathbf{z}) = \mathcal{A}\mathbf{x}_0 + \mathcal{A}\mathbf{z} = \mathcal{A}\mathbf{x}_0 = \mathcal{A}_0\mathbf{x}_0. \qquad \square$$

В силу предложения $\dim U_0 = \dim V_0$; эта размерность равна рангу оператора $\mathcal{A}$, который мы обозначим через r. Построим «согласованные» ортонормированные базисы в U_0 и V_0.

$\nabla 1$. $\operatorname{Ker} \mathcal{A} = \operatorname{Ker} \mathcal{A}\mathcal{A}^*$.

Доказательство. Ясно, что $\operatorname{Ker} \mathcal{A} \subseteq \operatorname{Ker} \mathcal{A}\mathcal{A}^*$. Обратно, пусть $\mathbf{x} \in \operatorname{Ker} \mathcal{A}\mathcal{A}^*$, т.е. $(\mathcal{A}\mathcal{A}^*)\mathbf{x} = \mathbf{0}$. Тогда

$$0 = \mathbf{x}((\mathcal{A}\mathcal{A}^*)\mathbf{x}) = \mathbf{x}(\mathcal{A}^*(\mathcal{A}\mathbf{x})) = (\mathcal{A}\mathbf{x})(\mathcal{A}\mathbf{x}).$$

Отсюда $\mathcal{A}\mathbf{x} = \mathbf{0}$, т.е. $\mathbf{x} \in \operatorname{Ker} \mathcal{A}$. $\square$

Из $\nabla 1$ следует, что ограничение оператора $\mathcal{A}\mathcal{A}^*$ на подпространство U_0 невырождено. Тогда это ограничение – положительный оператор, так как $((\mathcal{A}\mathcal{A}^*)\mathbf{x})\mathbf{x} = (\mathcal{A}^*(\mathcal{A}\mathbf{x}))\mathbf{x} = (\mathcal{A}\mathbf{x})(\mathcal{A}\mathbf{x}) > 0$ для ненулевых векторов $\mathbf{x} \in U_0$.

Выберем в U_0 ортонормированный базис $\mathbf{u}_1, \dots, \mathbf{u}_r$ из собственных векторов оператора $\mathcal{A}\mathcal{A}^*$, принадлежащих собственным значениям $\lambda_1, \dots, \lambda_r > 0$. Образы $\mathcal{A}\mathbf{u}_1, \dots, \mathcal{A}\mathbf{u}_r$ этих векторов лежат в V_0 и попарно ортогональны. Действительно, при $i \neq j$

$$\mathcal{A}\mathbf{u}_i\mathcal{A}\mathbf{u}_j = \mathbf{u}_i(\mathcal{A}^*(\mathcal{A}\mathbf{u}_j)) = \mathbf{u}_i((\mathcal{A}\mathcal{A}^*)\mathbf{u}_j) = \mathbf{u}_i(\lambda_j\mathbf{u}_j) = \lambda_j\mathbf{u}_i\mathbf{u}_j = 0.$$

Положим $\mu_i := \sqrt{\lambda_i}$ и $\mathbf{v}_i := \mu_i^{-1}\mathcal{A}\mathbf{u}_i$ для каждого $i = 1, \dots, r$. Тогда

$$|\mathbf{v}_i|^2 = \mathbf{v}_i\mathbf{v}_i = \frac{(\mathcal{A}\mathbf{u}_i)(\mathcal{A}\mathbf{u}_i)}{\mu_i^2} = \frac{\mathbf{u}_i(\mathcal{A}^*(\mathcal{A}\mathbf{u}_i))}{\mu_i^2} = \frac{\mathbf{u}_i((\mathcal{A}\mathcal{A}^*)\mathbf{u}_i)}{\mu_i^2} = \frac{\lambda_i}{\mu_i^2} = 1.$$

Сингулярные базисы (2)

Итак, вектора $\mathbf{v}_1, \dots, \mathbf{v}_r$ образуют ортонормированный базис подпространства V_0. Дополним систему $\mathbf{v}_1, \dots, \mathbf{v}_r$ ортонормированным базисом пространства $\operatorname{Ker} \mathcal{A}^* = (\operatorname{Im} \mathcal{A})^\perp$ до ортонормированного базиса $\mathbf{v}_1, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ всего пространства V, здесь $n := \dim V$.
Аналогично, систему $\mathbf{u}_1, \dots, \mathbf{u}_r$ дополним ортонормированным базисом пространства $\operatorname{Ker} \mathcal{A} = (\operatorname{Im} \mathcal{A}*)^\perp$ до ортонормированного базиса $\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_k$ всего пространства U, здесь $k := \dim U$.
Эти ортонормированные базисы называются *сингулярными базисами* оператора $\mathcal{A}$, а числа $\mu_1, \mu_2, \dots, \mu_r$ – его *сингулярными числами*.
Обычно первые r векторов в сингулярных базисах нумеруют так, чтобы сингулярные числа шли в порядке убывания: $\mu_1 \geq \mu_2 \geq \dots \geq \mu_r > 0$.

Действие оператора $\mathcal{A}$ в его сингулярных базисах прозрачно донельзя:

$$\mathcal{A}\mathbf{u}_i = \begin{cases} \mu_i \mathbf{v}_i, & \text{если } i \leq r, \\ \mathbf{0}, & \text{если } i > r. \end{cases}$$

Поэтому матрица оператора $\mathcal{A}$ в его сингулярных базисах устроена так: в ней на местах $(1,1), (2,2), \dots, (r,r)$ стоят соответствующие сингулярные числа $\mu_1, \mu_2, \dots, \mu_r$, а на всех остальных местах – нули.

Мы доказали бóльшую часть следующего результата:

Теорема (сингулярное представление линейного оператора)

Для любого линейного оператора $\mathcal{A}\colon U \to V$ *пространств со скалярным произведением в* U *и* V *можно выбрать ортонормированные базисы, в которых его матрица имеет вид*

$$A = \begin{pmatrix} \mu_1 & 0 & \dots & 0 & 0 & \dots \\ 0 & \mu_2 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \mu_r & 0 & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{pmatrix}, \qquad (\sharp)$$

где r *– ранг* $\mathcal{A}$*, а* $\mu_1, \mu_2, \dots, \mu_r$ *– положительные действительные числа, определяемые однозначно с точностью до порядка.*

Доказательство. Осталось доказать только то, что если матрица оператора $\mathcal{A}$ в каких-то ортонормированных базисах имеет вид $(\sharp)$, то диагональные элементы этой матрицы определены однозначно с точностью до порядка.

Действительно, если B_U и B_V – какие-то ортонормированные базисы в U и V, а A – матрица оператора $\mathcal{A}$ в этих базисах, то матрица сопряженного оператора $\mathcal{A}^*$ в тех же базисах есть A^*. Если A имеет вид $(\sharp)$, то $A^* = A^T$. Поэтому матрица оператора $\mathcal{A}\mathcal{A}^*$ в базисе B_U – квадратная диагональная матрица с ненулевыми элементами диагонали, равными $\mu_1^2, \mu_2^2, \dots, \mu_r^2$. Если матрица оператора в некотором базисе диагональна, то на диагонали стоят его собственные значения. Значит, $\mu_1^2, \mu_2^2, \dots, \mu_r^2$ – собственные значения оператора $\mathcal{A}\mathcal{A}^*$, а следовательно, $\mu_1, \mu_2, \dots, \mu_r$ суть в точности сингулярные числа оператора $\mathcal{A}$ и не зависят от базисов B_U и B_V. $\square$

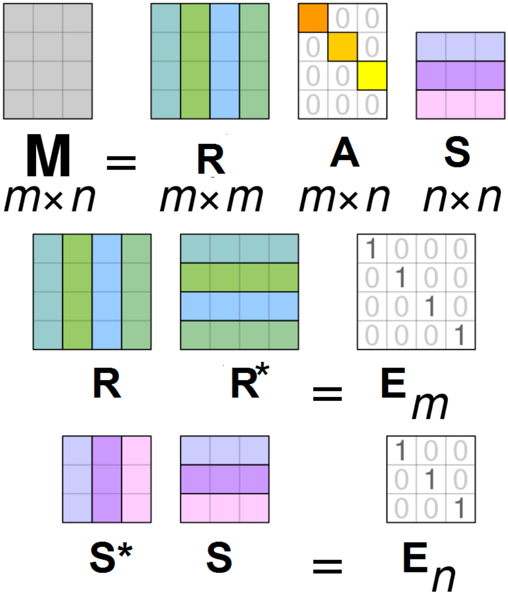
Следствие (SVD, Джеймс Джозеф Сильвестр, 1889)

Пусть F – одно из полей $\mathbb{R}$ или $\mathbb{C}$, а M – произвольная $k \times n$-матрица над F. Существует пара ортогональных (унитарных) матриц $R \in M_k(F)$ и $S \in M_n(F)$ такая, что $M = RAS$, где A – матрица вида $(\sharp)$, где r – ранг матрицы M, а $\mu_1, \mu_2, \dots, \mu_r$ – положительные действительные числа, определяемые однозначно с точностью до порядка.

Доказательство. Определим оператор $\mathcal{A}\colon F^n \to F^k$ таким правилом: $\mathcal{A}\mathbf{x} := M\mathbf{x}$. По теореме в F^n и F^k есть ортонормированные базисы, скажем, B_n и B_k, в которых матрица A оператора $\mathcal{A}$ имеет вид $(\sharp)$. С другой стороны, M есть матрица оператора $\mathcal{A}$ в стандартных ортонормированных базисах пространств F^n и F^k. Обозначим через R матрицу перехода от стандартного базиса пространства F^k к базису B_k, а через S матрицу перехода от базиса B_n к стандартному базису пространства F^n. Тогда матрицы R и S ортогональные (унитарные), и по формуле замены матрицы оператора $M = RAS$. $\square$

Замечание. Речь идет о формуле $A_Q = T_{Q\to P} A_P T_{P\to Q}$ из §III.1. Мы ее доказали для операторов на одном пространстве, но она справедлива (с тем же доказательством) для операторов из пространства U в пространство V, если под P и Q понимать *пары* (базис U, базис V).

Иллюстрация



Полярное разложение из предыдущей лекции легко получается из SVD.

Действительно, пусть M – произвольная квадратная матрица над одним из полей $\mathbb{R}$ или $\mathbb{C}$. По SVD имеем $M = RAS$, где A – матрица вида ($\sharp$), а R и S – ортогональные (унитарные) матрицы. Тогда $M = (RS)(S^{-1}AS)$, матрица RS ортогональна (унитарна), а $S^{-1}AS$ – неотрицательная симметрическая (эрмитова) матрица.
Итак, каждая квадратная действительная (комплексная) матрица представима в виде произведения ортогональной (унитарной) и симметрической (эрмитовой) матриц.
На операторном языке это означает, что любой линейный оператор $\mathcal{A}$ на евклидовом (унитарном) пространстве представим в виде $\mathcal{A} = \mathcal{B}\mathcal{U}$, где $\mathcal{B}$ – неотрицательный оператор, а $\mathcal{U}$ – ортогональный (унитарный) оператор.

Напомним, что в конце осеннего семестра мы построили для каждого линейного оператора $\mathcal{A}\colon U \to V$ пространств со скалярным произведением *псевдообратный* оператор $\mathcal{A}^+\colon V \to U$, возвращающий нормальное псевдорешение произвольной системы линейных уравнений $A\mathbf{x} = \mathbf{b}$ с матрицей A, соответствующей $\mathcal{A}$, по правой части $\mathbf{b}$ этой системы.

Псевдообратный оператор характеризуется *тождествами Пенроуза*:

$$(\mathcal{A}\mathcal{A}^+)^* = \mathcal{A}\mathcal{A}^+ \quad (1), \qquad \mathcal{A}\mathcal{A}^+\mathcal{A} = \mathcal{A} \quad (3),$$
$$(\mathcal{A}^+\mathcal{A})^* = \mathcal{A}^+\mathcal{A} \quad (2), \qquad \mathcal{A}^+\mathcal{A}\mathcal{A}^+ = \mathcal{A}^+ \quad (4).$$

Зная SVD матрицы, легко построить псевдообратную к ней.
Для простоты формулировок ограничимся евклидовым случаем.
(В унитарном случае нужно только поменять ^{T} на *.)

Предложение (псевдообратная матрица через SVD)

Пусть M — $k \times n$-матрица ранга r над полем $\mathbb{R}$, а $M = RAS$ – ее SVD, где A – $k \times n$-матрица вида ($\sharp$) с ненулевыми числами $\mu_1, \mu_2, \dots, \mu_r$ на диагонали. Тогда $M^+ = S^T A^+ R^T$, где A^+ – $n \times k$-матрица вида ($\sharp$) с ненулевыми числами $\mu_1^{-1}, \mu_2^{-1}, \dots, \mu_r^{-1}$ на диагонали.

Итак, чтобы найти псевдообратную матрицу по SVD данной матрицы, надо транспонировать SVD и обратить сингулярные числа.

$$M = R\begin{pmatrix} \mu_1 & 0 & \dots & 0 & 0 & \dots \\ 0 & \mu_2 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \mu_r & 0 & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \ddots \end{pmatrix} S \implies M^+ = S^T \begin{pmatrix} \mu_1^{-1} & 0 & \dots & 0 & 0 & \dots \\ 0 & \mu_2^{-1} & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \mu_r^{-1} & 0 & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \ddots \end{pmatrix} R^T$$

Чтобы убедиться, что формула $M^+ = S^T A^+ R^T$ действительно возвращает псевдообратную матрицу, достаточно подставить выражение $S^T A^+ R^T$ в тождества (1)–(4) и проверить, что они выполняются. Например, проверим равенство $MM^+M = M$:

$$\underbrace{RAS}_{M}\ \underbrace{S^T A^+ R^T}_{\text{кандидат в } M^+}\ \underbrace{RAS}_{M} = RA\overbrace{SS^T}^{E}A^+\overbrace{R^T R}^{E}AS = RAA^+AS = RAS = M.$$

Приложения SVD: сжатие информации

В многих практических задачах (сжатие данных, обработка сигналов, метод главных компонент, латентно-семантическое индексирование и т.д.) нужно приблизить матрицу M некоторой другой матрицей M_d с заранее заданным рангом d. При этом стремятся минимизировать $|M - M_d|^2$. (Под длиной матрицы здесь понимается длина ее *векторизации*.)

Теорема (Карл Эккарт, Гейл Янг, 1936)

Пусть M — $k \times n$-матрица ранга r над полем $\mathbb{R}$, а $M = RAS$ – ее SVD, где A – $k \times n$-матрица вида $(\sharp)$ с числами $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_r$ на диагонали, и пусть $d < r$. Тогда матрица M_d ранга d с наименьшей возможной величиной $|M - M_d|^2$ получается как $M_d := R_d A_d S_d$, где A_d – диагональная $d \times d$-матрица с числами $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_d$ на диагонали, R_d – $k \times d$-матрица, образованная первыми d столбцами матрицы R, а S_d – $d \times n$-матрица, образованная первыми d строками матрицы S.

Итак, нужно оставить d первых сингулярных чисел, а остальные занулить.