Линейная алгебра

Дима Трушин

Семинар 8

Проекции

Пусть V – некоторое векторное пространство и $U, W \subseteq V$ – некоторые подпространства. Будем говорить, что V раскладывается в прямую сумму этих подпространств, если $U \cap W = 0$ и $V = U + W$, т.е. любой вектор $v \in V$ представляется в виде $v = u + w$, где $u \in U$ и $w \in W$ (то есть $U + W = \{u + w \mid u \in U,\ w \in W\}$). Думать про это надо так, U и W – это непересекающиеся подпространства и V является наименьшим пространством их содержащим.Такое разложение всегда получается так: берем какой-нибудь базис $e_1, \ldots, e_n$ пространства V, делим его на две части $e_1, \ldots, e_k$ и $e_{k+1}, \ldots, e_n$ и полагаем $U = \langle e_1, \ldots, e_k \rangle$ и $W = \langle e_{k+1}, \ldots, e_n \rangle$. Если пространство V является прямой суммой подпространств U и W, то мы будем обозначать это дело следующим образом $V = U \oplus W$. В этом случае любой вектор v единственным образом раскладывается в виде $v = u + w$, где $u \in U$ и $w \in W$. Еще в этом случае $\dim U + \dim W = \dim V$.

Утверждение. *Пусть V – евклидово пространство и $U \subseteq V$ – произвольное подпространство. Тогда $V = U \oplus U^\perp$.*

Таким образом в евклидовом пространстве V при фиксированном подпространстве $U \subseteq V$, любой вектор $v \in V$ единственным образом раскладывается в сумму $v = \operatorname{pr}_U v + \operatorname{ort}_U v$, где $\operatorname{pr}_U v \in U$ и $\operatorname{ort}_U v \in U^\perp$.

Определение. Если V – евклидово пространство, $U \subseteq V$ – произвольное подпространство и $v \in V$, то

- Вектор $\operatorname{pr}_U v$ называется ортогональной проекцией v на U.
- Вектор $\operatorname{ort}_U v$ называется ортогональной составляющей v относительно U.

Обратите внимание, что ортогональная проекция v на U – это проекция v на U вдоль $U^\perp$, а ортогональная составляющая – проекция v на $U^\perp$ вдоль U.

Формула БАБА

Давайте я в начале разберу задачу нахождения проекции вектора на подпространство вдоль другого подпространства (здесь нам не нужно никакое скалярное произведение). Пусть V – некоторое векторное пространство и $V = U \oplus W$. Тогда на пространстве V задан оператор проекции $P\colon V \to V$ такой, что $\ker P = W$ и $P|_U = \mathrm{Id}$, то есть, если $v \in V$ раскладывается в сумму $v = u + w$, где $u \in U$ и $w \in W$, то $Pv = u$ – оператор вычисления проекции на U вдоль W.

Теперь мы хотим научиться эффективно считать P. Для этого предположим $V = \mathbb{R}^n$, $U = \langle u_1, \ldots, u_k \rangle$, $W = \{y \in \mathbb{R}^n \mid Ay = 0\}$, где $A \in \mathrm{M}_{s\,n}(\mathbb{R})$. В этом случае $P\colon \mathbb{R}^n \to \mathbb{R}^n$ задается некоторой матрицей. Наша задача – найти эту матрицу.

Предположим для простоты, что векторы $u_1, \ldots, u_k$ образуют базис U, а строки матрицы A линейно независимы. Определим матрицу $B = (u_1|\ldots|u_k) \in \mathrm{M}_{n\,k}(\mathbb{R})$. Тогда утверждаются следующие вещи:

1. Количество столбцов B совпадает с количеством строк A, то есть $k = s$.
2. Матрица AB обратима.
3. Оператор проекции задается формулой $P = B(AB)^{-1}A$. Мнемоническое правило «БАБА».

Доказательство. Матрица A задает линейное отображение $A\colon \mathbb{R}^n \to \mathbb{R}^s$ такое, что $\ker A = W$ и $\operatorname{Im} A = \mathbb{R}^s$ (так как строки матрицы A линейно независимы, то $\operatorname{rk} A = s$, но $\operatorname{rk} A = \dim \operatorname{Im} A$). Матрица B задает отображение $B\colon \mathbb{R}^k \to \mathbb{R}^n$ такое, что $\operatorname{Im} B = U$ и $\ker B = 0$ (так как столбцы B линейно независимы).

(1) Мы знаем, что

$$\begin{matrix} \dim U + \dim W = n \\ \dim \ker A + \dim \operatorname{Im} A = n \end{matrix} \quad \text{то есть} \quad \begin{matrix} k + \dim W = n \\ \dim W + s = n \end{matrix} \quad \text{откуда} \quad s = k$$

(2) Теперь рассмотрим отображение $AB\colon \mathbb{R}^k \to \mathbb{R}^k$. Заметим, что $\operatorname{Im} B \cap \ker A = U \cap W = 0$. Значит $\ker AB = 0$, то есть AB – обратимый оператор.

(3) Теперь выведем формулу для P. Пусть $v = u + w$, где $v \in \mathbb{R}^n$ – произвольный вектор, $u \in U$ и $w \in W$ – его единственное разложение по прямой сумме подпространств. Тогда $Av = Au + Aw = Au$. С другой стороны, так как $u \in U$, мы имеем $u = Bx$ для некоторого $x \in \mathbb{R}^k$. Тогда $Av = ABx$. Так как AB обратимая квадратная матрица, имеем $x = (AB)^{-1}Av$. Значит $u = Bx = B(AB)^{-1}Av$, что и требовалось. $\square$

Обратите внимание, что проектор P на U вдоль W зависит от двух подпространств, а не только от U. Если вы измените одно из них, то проектор изменится.

Формула Атата

Теперь я хочу разобрать случай проектора на подпространство вдоль его ортогонального дополнения. Такой проектор называется ортопроектором. Пусть $V = \mathbb{R}^n$ со стандартным скалярным произведением $(x, y) = x^t y$ и пусть подпространство $U \subseteq V$ задано своим базисом $U = \langle u_1, \ldots, u_k \rangle$. Составим матрицу $A = (u_1 | \ldots | u_k) \in \operatorname{M}_{n\,k}(\mathbb{R})$. Тогда $U^\perp = \{y \in \mathbb{R}^n \mid A^t y = 0\}$. Пусть теперь $v \in V$ – произвольный вектор и $v = \operatorname{pr}_U v + \operatorname{ort}_U v$. Тогда формула «БАБА» превращается в $\operatorname{pr}_U v = A(A^tA)^{-1}A^t v$. Мнемоническое правило для запоминания: в евклидовом пространстве БАБА – это Атата.

Обратите внимание, что проектор P всегда зависит от двух подпространств: то, на которое проектируем U, и то, вдоль которого проектируем W. Но в случае ортогонального проектирования $W = U^\perp$, потому ортопроектор P реально зависит только от одного подпространства.

Метод наименьших квадратов

Пусть мы хотим решить систему $Ax = b$, где $A \in \operatorname{M}_{m\,n}(\mathbb{R})$, $b \in \mathbb{R}^m$ и $x \in \mathbb{R}^n$ – столбец неизвестных. И предположим, что система не имеет решений, но от этого наше желание ее решить не становится слабее. Давайте обсудим, как удовлетворить наши желания в подобной ситуации и когда такие ситуации обычно встречаются.

Введем на пространстве $\mathbb{R}^m$ стандартное скалярное произведение $(x, y) = x^t y$. Тогда, на процесс решения системы можно смотреть так: мы подбираем $x \in \mathbb{R}^n$ так, чтоб $|Ax - b| = 0$. Если решить систему невозможно, то этот подход подсказывает, как надо поступить. Надо пытаться минимизировать расстояние между Ax и b. То есть решить задачу

$$\begin{aligned} &|Ax - b| \to \min \\ &x \in \mathbb{R}^n \end{aligned}$$

Теперь давайте поймем, как надо решать такую задачу. Пусть матрица A имеет вид $A = (A_1 | \ldots | A_n)$, где $A_i \in \mathbb{R}^m$ – ее столбцы. Тогда система $Ax = b$ означает, $x_1 A_1 + \ldots + x_n A_n = b$. То есть система разрешима тогда и только тогда, когда $b \in \langle A_1, \ldots, A_n \rangle$. Значит наша задача минимизировать расстояние между b и $\langle A_1, \ldots, A_n \rangle$. Мы можем разложить вектор b на проекцию и ортогональную составляющую относительно $\langle A_1, \ldots, A_n \rangle$. Обычная теорема пифагора говорит, что минимум расстояния достигается на $b_0 = \operatorname{pr}_{\langle A \rangle} b$. В этом случае вместо исходной системы $Ax = b$ мы должны решить систему $Ax = b_0$. И если x_0 – ее решение, то $|Ax_0 - b|$ как раз и будет минимальным.

Давайте теперь предположим, что столбцы матрицы A линейно независимы. Тогда по формуле «Атата» мы знаем, что $b_0 = A(A^tA)^{-1}A^t b$. Кроме этого должно выполняться $b_0 = Ax_0$. Так как столбцы A линейно независимы, такое x_0 должно быть единственным. Но мы видим, что в качестве x_0 подходит $x_0 = (A^tA)^{-1}A^t b$.

Объемы

Матрица Грама Прежде чем переходить к объемам, мне понадобятся некоторые технические вещи. Пусть V – евклидово пространство и $v_1, \ldots, v_k \in V$ – произвольный набор векторов (k НЕ обязательно равно размерности пространства). Тогда матрица

$$G(v_1, \dots, v_k) = \begin{pmatrix} (v_1, v_1) & \dots & (v_1, v_k) \\ \vdots & \ddots & \vdots \\ (v_k, v_1) & \dots & (v_k, v_k) \end{pmatrix} \in \mathrm{M}_k(\mathbb{R})$$

называется матрицей Грама системы векторов $v_1, \dots, v_k$.[1]

Если $e_1, \dots, e_n$ – некоторый базис пространства V, то $B = G(e_1, \dots, e_n)$ – матрица скалярного произведения заданная в базисе $e_1, \dots, e_n$. Таким образом матрица Грама – это некоторое обобщение матрицы билинейной формы.

Пример Пусть $V = \mathbb{R}^n$ со стандартным скалярным произведением $(x, y) = x^t y$. И пусть $v_1, \dots, v_k \in \mathbb{R}^n$ – произвольный набор векторов. Тогда составим матрицу $A = (v_1 | \dots | v_k) \in \mathrm{M}_{n\,k}(\mathbb{R})$. Тогда матрица Грама будет

$$G(v_1, \dots, v_k) = A^t A$$

Обратите внимание, что k может быть больше, чем n или меньше, это совершенно не важно. Матрица Грама определена для любого количества векторов.

Утверждение. *Пусть $v_1, \dots, v_k \in V$ – произвольный набор векторов в евклидовом пространстве. Тогда*

1. *Пусть $\alpha_1, \dots, \alpha_k \in \mathbb{R}$, введем обозначение $\alpha = (\alpha_1, \dots, \alpha_k)^t$. Тогда следующие условия эквивалентны:*
   (a) $\alpha_1 v_1 + \dots + \alpha_k v_k = 0$.
   (b) $G(v_1, \dots, v_k)\alpha = 0$.
   (c) $\alpha^t G(v_1, \dots, v_k)\alpha = 0$.
2. $\operatorname{rk} G(v_1, \dots, v_k) = \dim\langle v_1, \dots, v_k \rangle$.
3. $\det G(v_1, \dots, v_k) \geqslant 0$. *При этом равенство достигается тогда и только тогда, когда векторы линейно зависимы.*
4. *Если $C \in \mathrm{M}_k(\mathbb{R})$ является матрицей элементарного преобразования I или II типа, то*

$$\det G(v_1, \dots, v_k) = \det G((v_1, \dots, v_k)C)$$

Заметим, что из третьего пункта следует вот какое наблюдение. Если $v_1, \dots, v_k$ – линейно независимый набор векторов, из которого процессом ортогонализации Грама-Шмидта мы получили набор $u_1, \dots, u_k$, то $\det G(u_1, \dots, u_k) = \det G(v_1, \dots, v_k)$.

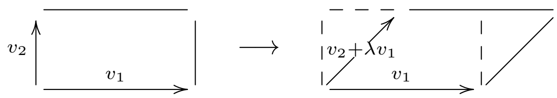
Объемы Теперь самое время прикоснуться к объемам. Надо сказать, что существует общая теория вычисления объемов в евклидовом пространстве V. Она позволяет посчитать «объем» любого подмножества в V. Данная конструкция ведет к понятию меры Лебега и далее к интегралу Лебега. Мы, конечно же, не будем развивать подобную теорию в такой общности, а всего лишь ограничимся вычислением объемов для простых и естественных с точки зрения линейной алгебры фигур – многомерных параллелепипедов.

Определение. Пусть V – евклидово пространство и $v_1, \dots, v_k \in V$ – набор векторов, тогда k-мерным параллелепипедом натянутым на $v_1, \dots, v_k$ называется следующее множество

$$\Pi(v_1, \dots, v_k) = \left\{ \sum_{i=1}^{k} x_i v_i \mid 0 \leqslant x_i \leqslant 1 \right\}$$

[1] Обратите внимание, тут важен порядок векторов. То есть формально матрица Грама зависит от набота $(v_1, \dots, v_k) \in V^k$.

VIP пример Я хочу разобрать один важный пример. Пусть $V = \mathbb{R}^2$ – плоскость со стандартным скалярным произведением $(x, y) = x^t y$ и $v_1, v_2 \in V$ – два ортогональных вектора. Если я заменю v_2 на $v_2 + \lambda v_1$, то геометрически я наклоню мой параллелепипед вдоль направления v_1 как на рисунке ниже:



На рисунке мы видим, что параллелепипед справа отличается от параллелепипеда слева перестановкой треугольника отмеченного пунктиром. А значит их площади одинаковые. Давайте посмотрим на матрицы Грама двух наборов векторов:

$$G(v_1, v_2 + \lambda v_1) = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix} G(v_1, v_2) \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \quad \text{при этом} \quad G(v_1, v_2) = \begin{pmatrix} |v_1|^2 & 0 \\ 0 & |v_2^2| \end{pmatrix}$$

То есть $\det G(v_1, v_2 + \lambda v_1) = \det G(v_1, v_2) = |v_1|^2 |v_2|^2 = S^2_{\Pi(v_1, v_2)}$. То есть определитель матрицы Грамма дает нам квадрат площади параллелограмма, натянутого на векторы v_1, v_2. Этот пример подсказывает, как надо определять объемы в общем случае.

Определение. Пусть $v_1, \ldots, v_k \in V$ – произвольный набор векторов в евклидовом пространстве. Тогда определим k-мерный объем параллелепипеда $\Pi(v_1, \ldots, v_k)$ по следующей формуле:

$$\mathrm{Vol}_k(\Pi(v_1, \ldots, v_k)) = \sqrt{\det G(v_1, \ldots, v_k)}$$

Оказывается, что про определитель линейного оператора можно думать геометрически в терминах объемов. Определитель показывает, во сколько раз линейная деформация соответствующая оператору изменяет n-мерные объемы.

Утверждение. *Пусть V – евклидово пространство размерности n, $v_1, \ldots, v_n \in V$ – некоторый набор векторов и $\varphi\colon V \to V$ – линейный оператор. Тогда*

$$\mathrm{Vol}_n(\varphi(\Pi(v_1, \ldots, v_n))) = |\det \varphi| \, \mathrm{Vol}_n(\Pi(v_1, \ldots, v_n))$$

Ориентированный объем Во-первых, я хочу сказать, что ориентированный объем в n-мерном пространстве бывает только для n-мерного объема. Во-вторых, я хочу начать с простого и понятного примера, а уж потом пару слов сказать про общий случай.

Пример Пусть $V = \mathbb{R}^n$ со стандартным скалярным произведением $(x, y) = x^t y$. Возьмем $v_1, \ldots, v_n \in \mathbb{R}^n$ произвольный набор векторов. Образуем матрицу $A = (v_1 | \ldots | v_n) \in \mathrm{M}_n(\mathbb{R})$. Тогда

- Матрица Грама по определению считается так: $G(v_1, \ldots, v_n) = A^t A$.
- Ее определитель: $\det G(v_1, \ldots, v_n) = \det A^2$.
- А значит обычный объем будет: $\mathrm{Vol}_n \Pi(v_1, \ldots, v_n) = |\det A|$.
- Если мы теперь избавимся от модуля, то получим понятие ориентированного объема: $\mathrm{Vol}_n^{or} \Pi(v_1, \ldots, v_n) = \det A$.

Заметьте, что ориентация набора (как и знак соответствующего объема) меняется при перестановке векторов в наборе на знак совершенной перестановки. Другая причина знака объема – смена направления вектора, то есть когда вектор v_i в наборе меняется на вектор $-v_i$.

Общий случай Пусть V – евклидово пространство. Для того, чтобы задать ориентированный объем, мы должны фиксировать некоторый ортонормированный базис $e_1, \ldots, e_n$, перейти к нему и тогда мы окажемся в ситуации примера выше. Таким образом ориентированный объем зависит от выбора ортонормированного базиса. Выбор базиса будет влиять только на знак объема. Если у нас есть два ортонормированных базиса $e_1, \ldots, e_n$ и $f_1, \ldots, f_n$ и пусть $(f_1, \ldots, f_n) = (e_1, \ldots, e_n)C$, то есть $C \in \mathrm{M}_n(\mathbb{R})$ – матрица перехода (она обязательно ортогональна). Тогда если $\det C > 0$ (что равносильно $\det C = 1$, то знаки объемов, которые получаются при выборе этих базисов будут одинаковыми. А если $\det C < 0$ (что равносильно $\det C = -1$), то знаки объемов, которые получаются при выборе этих базисов будут разными.

Утверждение. *Пусть V – евклидово пространство размерности n, мы зафиксировали базис $e_1, \ldots, e_n$ относительно которого будем считать ориентированный объем, $v_1, \ldots, v_n \in V$ – некоторый набор векторов и $\varphi\colon V \to V$ – линейный оператор. Тогда*

$$\operatorname{Vol}_n^{or}(\varphi(\Pi(v_1, \ldots, v_n))) = \det \varphi \operatorname{Vol}_n^{or}(\Pi(v_1, \ldots, v_n))$$

Движения и ортогональные матрицы

Так как углы и расстояния выражаются через скалярное произведение и наоборот, мы получаем следующее.

Утверждение. *Пусть теперь $\phi\colon V \to V$ – линейный оператор в евклидовом пространстве. Следующие утверждения эквивалентны:*

1. *ϕ сохраняет скалярное произведение, т.е. $(\phi(v), \phi(u)) = (v, u)$ для любых $v, u \in V$.*
2. *ϕ сохраняет длины и углы, т.е. $|\phi(v)| = |v|$ и $\alpha_{\phi(v),\phi(u)} = \alpha_{v,u}$ для всех $v, u \in V$.*
3. *ϕ сохраняет длины, т.е. $|\phi(v)| = |v|$ для всех $v \in V$.*

Линейные операторы, обладающие одним из эквивалентных свойств выше, называются *движениями*. Пусть в V выбрали ортонормированный базис. Это значит, что V можно отождествить с $\mathbb{R}^n$ и при этом скалярное произведение превращается в стандартное $(x, y) = x^t y$. Пусть отображение $\phi\colon \mathbb{R}^n \to \mathbb{R}^n$ задано матрицей $A \in \mathrm{M}_n(\mathbb{R})$. Тогда условие движения записывается так $(Ax, Ay) = (x, y)$. То есть $x^t A^t A y = x^t y$ для любых $x, y \in \mathbb{R}^n$. То есть $A^t A = E$. Теперь заметим следующее.

Утверждение. *Для матрицы $A \in \mathrm{M}_n(\mathbb{R})$ следующие условия эквивалентны:*

1. *$A^t A = E$.*
2. *$AA^t = E$.*
3. *$A^t = A^{-1}$.*

Матрица обладающая одним из этих эквивалентных условий называется *ортогональная*. Таким образом в ортонормированном базисе движение задается ортогональной матрицей.

Утверждение. *Пусть $C \in \mathrm{M}_n(\mathbb{R})$ – ортогональная матрица и пусть $\lambda \in \mathbb{C}$ – ее собственное значение. Тогда*

1. *$\bar{\lambda}$ тоже является собственным значением для C.*
2. *$|\lambda| = 1$.*

Примеры

1. Пусть $V = \mathbb{R}^2$ со стандартным скалярным произведением. Тогда любое движение это:
   (a) центральная симметрия относительно начала координат $C = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.
   (b) симметрия относительно какой-то прямой $C = D^t \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) D$, где D – матрица поворота (см. далее).
   (c) поворот на некоторый угол, $C = \left(\begin{smallmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{smallmatrix}\right)$ – матрица поворота.
2. Пусть $V = \mathbb{R}^3$ со стандартным скалярным произведением и $C \in \mathrm{M}_3(\mathbb{R})$ – ортогональная матрица. Тогда $\chi_C(t)$ – многочлен степени 3. Любой многочлен нечетной степени имеет хотя бы один вещественный корень.[2] А значит это ± 1. То есть соответствующий собственный вектор v либо неподвижен, либо отражается в $-v$ под действием C. Кроме того, ортогональное дополнение $\langle v \rangle^{\perp}$ является двумерной плоскостью, на которой C действует одним из трех способов описанных в предыдущем пункте. Короче

[2] Потому что такое многочлен устроен $\chi(t) = t^n(1 + o(1))$ при $t \to \pm\infty$. То есть на плюс бесконечности многочлен уходит в плюс бесконечность, а на минус бесконечности – в минус бесконечность. То есть по не прерывности он где-то должен был пересечь горизонтальную ось координат. А эта точка и есть корень.

говоря, если задано движение в трехмерном пространстве, то в каком-то ортонормированном базисе оно имеет один из следующих видов:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}$$

Первое из них является поворотом вокруг некоторой оси, а второе является поворотом вокруг оси и отражение вдоль оси.

Утверждение. *Пусть V евклидово пространство и $\phi\colon V \to V$ – некоторый оператор. Тогда эквивалентно*

1. *ϕ является движением (ортогональный оператор).*

2. *В некотором ортонормированном базисе матрица оператора ϕ имеет вид:*

$$A_\phi = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_r \end{pmatrix}, \quad \textit{где} \quad A_i \ \textit{либо} \ 1, \ \textit{либо} \ -1, \ \textit{либо} \ \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

Замечания

- Для ортогональной матрицы $\det C = \pm 1$ (примените $\det$ к равенству $C^tC = E$). Если $\det C = 1$, движение называется *собственным* и если $\det C = -1$, то *несобственным*.

- Если $e_1, \ldots, e_n$ и $f_1, \ldots, f_n$ – ортонормированные базисы пространства V и пусть $(f_1, \ldots, f_n) = (e_1, \ldots, e_n)C$, где C – матрица перехода. Тогда C является ортогональной матрицей. Это вторая ситуация, когда появляются ортогональные матрицы.

Самосопряженные операторы

Пусть V – евклидово пространство и пусть $\phi\colon V \to V$ – линейный оператор. Тогда *сопряженный* к нему линейный оператор ϕ^* – это такой оператор, что $(\phi(v), u) = (v, \phi^*(u))$ для всех $v, u \in V$. Оператор называется *самосопряженным*, если $\phi^* = \phi$.

Теперь разберемся, что происходит в ортонормированном базисе. В этом случае $V = \mathbb{R}^n$, $(x, y) = x^t y$, а $\phi(x) = Ax$, а $\phi^*(x) = Bx$. Тогда условие $(Ax, y) = (x, By)$ означает $x^t A^t y = x^t B y$. То есть $B = A^t$. То есть матрица для ϕ^* это A^t. Значит самосопряженный оператор в ортонормированном базисе задается симметричной матрицей.

В случае произвольного базиса скалярное произведение задается $(x, y) = x^t B y$, где B – симметричная невырожденная положительно определенная матрица. Тогда если $\phi x = Ax$ и $\phi^* x = A'x$, то условие $(Ax, y) = (x, A'y)$ расписывается так: $(Ax)^t B y = x^t B A' y$. То есть $x^t A^t B y = x^t B A' y$ для всех $x, y \in \mathbb{R}^n$. Последнее значит, что $A^t B = BA'$. Значит $A' = B^{-1} A^t B$ – это формула связывает матрицу ϕ и ϕ^* в произвольных базисах.

Утверждение. *Пусть $\phi\colon V \to V$ – самосопряженный оператор в евклидовом пространстве. Тогда*

1. *Все его собственные значения вещественны.*

2. *Собственные вектора с разными собственными значениями ортогональны друг другу.*

3. *Существует ортонормированный базис пространства V состоящий из собственных векторов ϕ.*

4. *В некотором ортонормированном базисе матрица ϕ имеет диагональный вид, с вещественными числами на диагонали.*

Переформулируем это утверждение на языке матриц.

Утверждение. *Пусть $A \in \mathrm{M}_n(\mathbb{R})$ – симметрическая матрица. Тогда*

1. *Все собственные значения A вещественные.*

2. *Все собственные вектора с разными собственными значениями ортогональны.*

3. *Существует ортогональная матрица $C \in \mathrm{M}_n(\mathbb{R})$ такая, что $C^{-1}AC$ является диагональной вещественной матрицей.*[3]

Самосопряженный оператор называется *положительным*, если все его собственные значения **неотрицательные**. Да, да, именно так. Нулевая матрица тоже считается положительным оператором. Вот такая вот дурацкая терминология.

Алгоритм разложения симметрических матриц

Дано Матрица $A \in \mathrm{M}_n(\mathbb{R})$ такая, что $A^t = A$.

Задача Найти разложение $A = C\Lambda C^t$, где $C \in \mathrm{M}_n(\mathbb{R})$ – ортогональная матрица, $\Lambda \in \mathrm{M}_n(\mathbb{R})$ – диагональная матрица.

Алгоритм

1. Найти собственные значения матрицы A.
   (a) Составить характеристический многочлен $\chi(\lambda) = \det(A - \lambda E)$.
   (b) Найти корни $\chi(\lambda)$ с учетом кратностей: $\{(\lambda_1, n_1), \ldots, (\lambda_k, n_k)\}$, где λ_i – корни, n_i – кратности.
2. Для каждого λ_i найти ортонормированный базис в пространстве собственных векторов отвечающему λ_i.
   (a) Найти ФСР системы $(A - \lambda_i E)x = 0$. Пусть это будет $v_1^i, \ldots, v_{n_i}^i$. Обратите внимание, что их количество будет в точности равно кратности n_i.
   (b) Ортогонализовать $v_1^i, \ldots, v_{n_i}^i$ методом Грама-Шмидта. Обратите внимание, после ортогонализации останется ровно n_i векторов.
   (c) Сделать каждый вектор длинны один: $v_j^i \mapsto \frac{v_j^i}{|v_j^i|}$.
3. Матрица Λ будет диагональной с числами $\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_k, \ldots, \lambda_k$ на диагонали, где каждое λ_i повторяется n_i раз. Обратите внимание, всего получится n чисел.
4. Матрица C будет составлена из столбцов $v_1^1, \ldots, v_{n_1}^1, v_1^2, \ldots, v_{n_2}^2, \ldots, v_1^k, \ldots, v_{n_k}^k$. Обратите внимание, порядок собственных векторов соответствует порядку собственных значений в матрице Λ.

Сингулярное разложение (SVD)

Это утверждение я в начале сформулирую на матричном языке.

Утверждение. *Пусть дана матрица $A \in \mathrm{M}_{m\,n}(\mathbb{R})$. Тогда*

1. *Существует $U \in \mathrm{M}_m(\mathbb{R})$ такая, что $U^t U = E$.*
2. *Существует $V \in \mathrm{M}_n(\mathbb{R})$ такая, что $V^t V = E$.*
3. *Существует последовательность вещественных чисел $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0$.*

такие, что $A = U\Sigma V^t$, где

$$\Sigma = \left.\overbrace{\begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix}}^{n}\right\} m$$

При этом последовательность чисел $\sigma_1, \sigma_2, \ldots, \sigma_r$ определена однозначно.

[3]Обратите внимание, чтот тут нет разницы между $C^{-1}AC$ и C^tAC, так как C ортогональная.

Пусть столбцы матрицы U – это вектора u_i, а столбцы матрицы V – это вектора v_i. Тогда утверждение означает, что

$$A = \sigma_1 u_i v_i^t + \ldots + \sigma_s u_s v_s^t$$

То есть мы представили матрицу A в виде «ортогональной» суммы матриц ранга один, в том смысле, что все u_i ортогональны друг другу и все v_i ортогональны друг другу.

Геометрически сингулярное разложение означает следующее.

Утверждение. *Пусть V и U – евклидовы или эрмитовы пространства и $\phi\colon V \to U$ – линейное отображение. Тогда существует ортонормированный базис $e_1, \ldots, e_n$ в V, ортонормированный базис $f_1, \ldots, f_m$ в U и последовательность вещественных чисел $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0$ такие, что матрица ϕ имеет вид*

$$\phi(e_1, \ldots, e_n) = (f_1, \ldots, f_m) \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix}$$

При этом числа $\sigma_1, \ldots, \sigma_r$ определены однозначно и называются сингулярными значениями отображения ϕ.

Компактное сингулярное разложение Пусть $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $U \in \mathrm{M}_{m\,s}(\mathbb{R})$, $V \in \mathrm{M}_{n\,s}(\mathbb{R})$ и $\Sigma \in \mathrm{M}_s(\mathbb{R})$ – диагональная матрица с числами $\sigma_1 \geqslant \ldots \geqslant \sigma_s > 0$ на диагонали. Предположим, что столбцы матриц U и V ортонормированны (то есть все между собой ортогональны и длины один). Тогда равенство вида $A = U\Sigma V^t$ называется компактным сингулярным разложением.

Если нам известно сингулярное разложение $A = U\Sigma V^t$, то компактное из него делается так: 1) составим матрицу U', состоящую из первых s столбцов матрицы U, 2) составим матрицу V', состоящую из первых s столбцов матрицы V, 3) определим матрицу Σ' как квадратную s на s матрицу с диагональю из матрицы Σ. Тогда $A = U'\Sigma'(V')^t$ будет компактным разложением.

Усеченное сингулярное разложение Пусть $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ и пусть для определенности $m \leqslant n$. И пусть $A = U\Sigma V^t$ – сингулярное разложение, где $U \in \mathrm{M}_m(\mathbb{R})$ и $V \in \mathrm{M}_n(\mathbb{R})$ – ортогональные матрицы, а $\Sigma \in \mathrm{M}_{m\,n}(\mathbb{R})$ – диагональная с числами $\sigma_1 \leqslant \ldots \leqslant \sigma_k > 0$ на диагонали ($k \leqslant m$). Тогда пусть $\Sigma' \in \mathrm{M}_m(\mathbb{R})$ будет матрица полученная из Σ отрезанием нулей вдоль длинной стороны так, чтобы она стала квадратной. И пусть $V' \in \mathrm{M}_{n\,m}(\mathbb{R})$ получена из V отрезанием последних столбцов так, чтобы осталось ровно m. Тогда видно, что $A = U\Sigma'(V')^t$. Такое разложение называется усеченным сингулярным разложением.

Замечание Философский смысл этого разложения следующий. Пусть наша матрица – это прямоугольная черно-белая картинка, где числа – интенсивности черного цвета. На вектора v_i и u_i надо смотреть как на «ортогональные» компоненты «базовых» цветовых интенсивностей. А λ_i – это мощности этих самых сигналов. Потому, если λ_i достаточно малы, то наш глаз не способен различить соответствующие сигналы. Потому, если мы выкинем их из нашей матрицы, то на глаз, матрица A не будет отличаться от полученной.

Обычно в реальной жизни выходит, что достаточно только первых штук пять слагаемых. Тогда $A' = \lambda_1 v_1 u_1^t + \ldots + \lambda_5 v_5 u_5^t$ будет на глаз не отличима от A. В чем же польза от такого? На хранение матрицы A нам потребуется mn чисел. Для хранения матрицы A' нам надо 5 чисел λ_i и еще 5 пар векторов v_i и u_i, на хранение каждого из которых надо m и n чисел соответственно. И того затраты $5 + 5m + 5n = 5(m + n + 1)$. Это дает огромный выигрыш в количестве хранимой информации и является основой для многих алгоритмов архивации с потерей данных вроде JPG.

Алгоритм нахождения компактного сингулярного разложения

Дано Матрица $A \in \mathrm{M}_{m\,n}(\mathbb{R})$.[4]

Задача Найти разложение $A = U\Sigma V^t$, где $U \in \mathrm{M}_{m\,s}(\mathbb{R})$, $V \in \mathrm{M}_{n\,s}(\mathbb{R})$ – матрицы с ортонормированными столбцами, $\Sigma \in \mathrm{M}_s(\mathbb{R})$ – диагональная матрица с элементами $\sigma_1 \geqslant \ldots \geqslant \sigma_s > 0$ на диагонали.

[4] Этот алгоритм рекомендуется применять при $m \leqslant n$, в противном случае, применить его к матрице A^t, а потом транспонировать полученное разложение.

Алгоритм

1. Составим матрицу $S = AA^t \in \mathrm{M}_m(\mathbb{R})$. Тогда $S = U\Sigma^2 U^t$.

2. Так как $S^t = S$. То с помощью алгоритма для симметрических матриц найдем ее разложение $S = CDC^t$.[5] Причем, обязательно получится, что диагональная матрица $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ состоит из неотрицательных элементов и мы можем выбрать порядок так, чтобы $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_m \geqslant 0$.

3. Пусть $C = (C_1|\ldots|C_m)$, тогда положим $U = (C_1|\ldots|C_s) \in \mathrm{M}_{m\,s}(\mathbb{R})$. А матрица $\Sigma \in \mathrm{M}_s(\mathbb{R})$ будет диагональной с числами $\sigma_i = \sqrt{\lambda_i}$ на диагонали, то есть $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_s)$.

4. Теперь надо найти V из условия $A = U\Sigma V^t$.[6] Положим $U = (u_1|\ldots|u_s)$ и $V = (v_1|\ldots|v_s)$. Тогда $A^t U\Sigma^{-t} = V$, то есть $v_i = \frac{1}{\sigma_i} A^t u_i$ при $1 \leqslant i \leqslant s$.

Алгоритм нахождения сингулярного разложения

Дано Матрица $A \in \mathrm{M}_{m\,n}(\mathbb{R})$.

Задача Найти разложение $A = U\Sigma V^t$, где $U \in \mathrm{M}_m(\mathbb{R})$ ортогональная, $V \in \mathrm{M}_n(\mathbb{R})$ ортогональная, $\Sigma \in \mathrm{M}_{m\,n}(\mathbb{R})$ содержит на диагонали элементы $\sigma_1 \geqslant \ldots \geqslant \sigma_s > 0$, а все остальные нули.

Алгоритм

1. Составим матрицу $S = AA^t \in \mathrm{M}_m(\mathbb{R})$. Тогда $S = U\Sigma\Sigma^t U^t$.

2. Так как $S^t = S$. То с помощью алгоритма для симметрических матриц найдем ее разложение $S = CDC^t$. Причем, обязательно получится, что диагональная матрица $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ состоит из неотрицательных элементов и мы можем выбрать порядок так, чтобы $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_m \geqslant 0$.

3. Тогда $U = C$, а $\Sigma\Sigma^t = D$. То есть $\sigma_i^2 = \lambda_i$. Так как $\sigma_i \geqslant 0$, то они находятся как $\sigma_i = \sqrt{\lambda_i}$.

4. Теперь надо найти V из условия $A = U\Sigma V^t$.[7] Пусть $\sigma_1 \geqslant \ldots \geqslant \sigma_s > 0$. Положим $U = (u_1|\ldots|u_m)$ и $V = (v_1|\ldots|v_n)$. Тогда $A^t U = V\Lambda^t$, то есть $v_i = \frac{1}{\sigma_i} A^t u_i$ при $1 \leqslant i \leqslant s$. Так мы находим первые s столбцов матрицы V.

5. Найдем ФСР для $\{y \in \mathbb{R}^n \mid Ay = 0\}$ и ортонормировать его (ортогонализуем Грамом-Шмидтом, а потом нормируем). Полученные векторы и будут оставшиеся $v_{s+1}, \ldots, v_n$.

Замечания

1. Надо заметить, что нельзя попытаться составить матрицу $A^t A$ и из нее найти матрицу V. Так как матрицы V и U определены не однозначно и зависят друг от друга. Если вы нашли какую-то матрицу U, то к ней подойдет не любая найденная матрица V, а только та, что является решением $A = U\Sigma V^t$.

2. Приведенным выше алгоритмом имеет смысл пользоваться, если у матрицы A количество строк меньше, чем количество столбцов. Если же столбцов меньше, чем строк, то надо найти сингулярное разложение для $A^t = U\Sigma V^t$. Тогда $A = V\Sigma^t U^t$ будет искомым сингулярным разложением для A.

[5] Здесь D будет диагональной матрицей, а C ортогональной.
[6] Обратите внимание, что Σ квадратная и обратимая матрица.
[7] Обратите внимание Σ не обязательно квадратная и тем более не обязательно обратимая.