Семинар 6

Ортопроекции

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 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 $\operatorname{pr}_U v = A(A^t A)^{-1} A^t v$. 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 $b_0 = A(A^tA)^{-1}A^tb$. 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 $x_0 = (A^tA)^{-1}A^tb$.

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 $(x, y) = x^ty$. 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 $x^tA^tAy = x^ty$ для любых $x, y \in \mathbb{R}^n$. То есть $A^tA = E$, то есть A должна быть ортогональной матрицей. Напомню три эквивалентных определения для нее.

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

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

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. $|\lambda| = 1$.
3. *Собственные векторы для разных собственных значений ортогональны.*

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.[1] 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

[1] 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:*

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

Для ортогональной матрицы $\det C = \pm 1$ (примените $\det$ к равенству $C^t C = E$). 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 $A^t B = B A'$. 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.*[2]

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.

[2]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.

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.

Задача о низкоранговом приближении

Теперь я хочу пару слов сказать о том, в каком смысле описанные выше процедуры являются оптимальными или попросту говоря «самыми лучшими». Пусть $A \in \mathrm{M}_{m\,n}(\mathbb{R})$. Зададим на пространствах матриц скалярное произведение по формуле $(A, B) = \mathrm{tr}(A^t B)$. Длина относительно заданного скалярного произведения называется нормой фробениуса и выражается следующим образом:

$$\|A\|_F = \sqrt{\sum_{ij} a_{ij}^2}$$

Если матрица A имеет вид $A = (A_1|\ldots|A_n)$, тогда $\|A\|_F^2 = \sum_{i=1}^n |A_i|^2$, где $|A_i|$ – длина относительно стандартного скалярного произведения для столбца A_i.

Теперь наша задача – заменить матрицу A на матрицу B ранга не выше k, причем мы хотим выбрать B ближайшей в смысле нормы фробениуса. То есть мы зафиксируем матрицу A и число k и будем решать задачу

$$\begin{cases} \|A - B\|_F \to \min\limits_B \\ \operatorname{rk} B \leqslant k \end{cases}$$

Важно понимать, что множество матриц ранга не выше k не образуют линейное подпространство в пространстве матриц. А значит, тут не получится решить эту задачу просто применением ортопроекторов. Кроме того, задача может иметь не единственное решение, в некоторых ситуациях ближайших матриц может оказаться бесконечное число.

Обратите внимание, что если $k \geqslant \operatorname{rk} A$, то ответом будет сама матрица A. А если $k < \operatorname{rk} A$, то оказывается, что SVD дает нужный ответ к данной задаче. Нужно найти для матрицы A сингулярное разложение. После чего, выбрать в качестве нужной матрицы матрицу

$$B_k = \sigma_1 u_1 v_1^t + \ldots + \sigma_k u_k v_k^t$$

Алгоритм нахождения компактного сингулярного разложения

Дано Матрица $A \in \mathrm{M}_{m\,n}(\mathbb{R})$.[3]

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.

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 $S = CDC^t$.[4] Причем, обязательно получится, что диагональная матрица $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$ 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 $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_s)$.
4. Теперь надо найти V из условия $A = U\Sigma V^t$.[5] 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 $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$ 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 $A = U\Sigma V^t$.[6] 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 $v_{s+1}, \ldots, v_n$.

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

[4]Здесь D будет диагональной матрицей, а C ортогональной.

[5]Обратите внимание, что Σ квадратная и обратимая матрица.

[6]Обратите внимание Σ не обязательно квадратная и тем более не обязательно обратимая.

Замечания

1. Надо заметить, что нельзя попытаться составить матрицу A^tA 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.

Матричные нормы

Здесь нас ждет пример первого абстрактного определения. Любое такое определение устроено одинаково, оно состоит из двух частей: первая часть содержит данные, а вторая аксиомы на них.

Нормы Будем через $\mathbb{R}_+$ обозначать множество неотрицательных вещественных чисел, т.е. $\mathbb{R}_+ = \{r \in \mathbb{R} \mid r \geqslant 0\}$.

Пусть задано отображение

$$\mathrm{M}_{m\,n}(\mathbb{R}) \to \mathbb{R}_+$$

т.е. это правило, которое по матрице A выдает некоторое неотрицательное вещественное число, которое будет обозначаться $|A| \in \mathbb{R}_+$. Такое отображение называется нормой, если выполнены следующие аксиомы

1. Для любой матрицы $A \in \mathrm{M}_{m\,n}(\mathbb{R})$, $|A| = 0$ тогда и только тогда, когда $A = 0$.

2. Для любой матрицы $A \in \mathrm{M}_{m\,n}(\mathbb{R})$ и любого числа $\lambda \in \mathbb{R}$ выполнено $|\lambda A| = |\lambda| \cdot |A|$.[7]

3. Для любых двух матриц $A, B \in \mathrm{M}_{m\,n}(\mathbb{R})$ выполнено $|A + B| \leqslant |A| + |B|$.

Стоит отметить, что $\mathbb{R}^n$ можно отождествить с матрицами $\mathrm{M}_{n\,1}(\mathbb{R})$. Потому определение выше дает понятие нормы на векторах из $\mathbb{R}^n$.

Субмультипликативность Пусть на квадратных матрицах $\mathrm{M}_n(\mathbb{R})$ задана некоторая норма. Тогда она называется субмультипликативной, если выполнено следующее свойство: для любых матриц $A, B \in \mathrm{M}_n(\mathbb{R})$ выполнено $|AB| \leqslant |A| \cdot |B|$.

Простые примеры

1. 1-норма на $\mathrm{M}_n(\mathbb{R})$:
$$|A|_1 = \sum_{ij} |a_{ij}|, \quad A \in \mathrm{M}_n(\mathbb{R})$$

2. ∞-норма на $\mathrm{M}_n(\mathbb{R})$:
$$|A|_\infty = \max_{ij} |a_{ij}|, \quad A \in \mathrm{M}_n(\mathbb{R})$$

3. Норма Фробениуса или 2-норма на $\mathrm{M}_n(\mathbb{R})$:
$$|A|_F = |A|_2 = \sqrt{\sum_{ij} |a_{ij}|^2}, \quad A \in \mathrm{M}_n(\mathbb{R})$$

Большинство из норм выше не являются субмультипликативными.

[7] Здесь $|\lambda|$ означает модуль числа, а $|A|$ – норма от матрицы.

Индуцированная (согласованная) норма Пусть $|{-}|\colon \mathbb{R}^n \to \mathbb{R}_+$ – некоторая фиксированная норма. Определим индуцированную ей норму $\|{-}\|$ на $\mathrm{M}_n(\mathbb{R})$ следующим образом:

$$\|A\| = \sup_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \frac{|Ax|}{|x|},\quad A\in \mathrm{M}_n(\mathbb{R})$$

Методом пристального взгляда мы замечаем, что отображение $\|{-}\|\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}_+$ удовлетворяет первым трем аксиомам нормы, а значит действительно является матричной нормой. Более того, верно следующее.

Утверждение. *Для любой нормы* $|{-}|\colon \mathbb{R}^n \to \mathbb{R}_+$ *индуцированная ей норма* $\|{-}\|\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}_+$ *является субмультипликативной.*

Доказательство. Из определения индуцированной нормы следует, что $|Ax|/|x| \leqslant \|A\|$ для любого ненулевого $x\in\mathbb{R}^n$. Ясно, что тогда для любого $x\in\mathbb{R}^n$, верно $|Ax|\leqslant \|A\|\cdot|x|$.

Пусть теперь $A, B\in \mathrm{M}_n(\mathbb{R})$ и нам надо показать, что $\|AB\|\leqslant \|A\|\cdot\|B\|$. Рассмотрим произвольный $x\in\mathbb{R}^n$, тогда

$$|ABx| = |A(Bx)| \leqslant \|A\|\cdot|Bx| \leqslant \|A\|\cdot\|B\|\cdot|x|$$

Значит

$$\frac{|ABx|}{|x|} \leqslant \|A\|\cdot\|B\|$$

для любого ненулевого $x\in\mathbb{R}^n$. А значит, можно перейти к супремуму по таким x, и следовательно

$$\|AB\| = \sup_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \frac{|ABx|}{|x|} \leqslant \|A\|\cdot\|B\|$$

□

Примеры индуцированных норм Индуцированные нормы хороши тем, что они субмультипликативны. Однако, обычно для них не существует явных формул для вычисления. Ниже мы приведем несколько случаев, когда такие формулы все же возможны. Все примеры будут даны без доказательств.

1. Пусть на $\mathbb{R}^n$ дана 1-норма $|x| = \sum_i |x_i|$. Тогда индуцированная норма $\|{-}\|_1$ на матрицах $\mathrm{M}_n(\mathbb{R})$ будет задаваться по формуле
$$\|A\|_1 = \max_j \sum_i |a_{ij}|$$
2. Пусть теперь на $\mathbb{R}^n$ дана ∞-норма $|x| = \max_i |x_i|$. Тогда индуцированная норма $\|{-}\|_\infty$ на матрицах $\mathrm{M}_n(\mathbb{R})$ будет задаваться по формуле
$$\|A\|_\infty = \max_i \sum_j |a_{ij}|$$
3. И наконец, пусть на $\mathbb{R}^n$ дана 2-норма $|x| = \sqrt{\sum_i |x_i|^2}$. Тогда индуцированная норма $\|{-}\|_2$ на матрицах $\mathrm{M}_n(\mathbb{R})$ уже считается более хитрым способом. Пусть $A\in \mathrm{M}_n(\mathbb{R})$. Если взять матрицу A^tA, то окажется, что ее спектр состоит целиком из неотрицательных вещественных чисел. Пусть σ_1 – максимальное такое число, тогда $\|A\|_2 = \sqrt{\sigma_1}$.[8]

Обзор применения матричных норм

Для простоты изложения, я буду рассматривать лишь квадратные матрицы ниже. Хотя какие-то вопросы и можно формулировать и для прямоугольных матриц, это не сделает материал более интересным.

[8] К этому явлению надо относиться так: есть спектр – объект из мира алгебры и есть норма – объект из мира анализа. Оказывается, что между анализом и алгеброй есть мостик через спектр и индуцированную 2-норму. Это позволяет задачи про спектр изучать аналитическими методами и наоборот задачи про сходимости изучать алгебраическими.

Сходимость Основная задача нормы – дать понятие о близости матриц друг к другу. А именно, если есть норма $|-|\colon \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}_+$, то можно определить расстояние между матрицами $A, B \in \mathrm{M}_n(\mathbb{R})$ следующим образом $\rho(A, B) = |A - B|$. А как только у нас есть понятие расстояния между объектами, мы можем ввести понятие предела и сходимости, а именно: пусть задана последовательность матриц $A_n \in \mathrm{M}_n(\mathbb{R})$, тогда скажем, что она сходится к матрице $A \in \mathrm{M}_n(\mathbb{R})$ и будем писать $A_n \to A$, $n \to \infty$ (или $\lim_n A_n = A$), если $\rho(A_n, A) \to 0$ как последовательность чисел при условии $n \to \infty$.

Эквивалентность норм Тут встает законный вопрос: у нас есть много различных норм на матрицах, а потому много расстояний, а значит получается огромное количество разных сходимостей (по одной на каждый вид нормы). Оказывается, что все возможные нормы на матрицах дают расстояния приводящие к одинаковому определению предела. Ключом к пониманию этого явления является определение эквивалентности норм. Пусть $|-|$ и $|-|'$ – две разные нормы на $\mathrm{M}_n(\mathbb{R})$. Будем говорить, что они эквивалентны, если существуют две положительные константы $c_1, c_2 \in \mathbb{R}$ такие, что $c_1|A| \leqslant |A|' \leqslant c_2|A|$ для всех матриц $A \in \mathrm{M}_n(\mathbb{R})$. Если две нормы эквивалентны, то несложно углядеть, что расстояние $\rho(A_n, A)$ в смысле нормы $|-|$ стремится к нулю тогда и только тогда, когда расстояние $\rho'(A_n, A)$ в смысле нормы $|-|'$ стремится к нулю. А значит эквивалентные нормы дают одну и ту же сходимость. Второй ключевой факт – все матричные нормы между собой эквивалентны. Это не очень сложный результат и по сути связан с тем, что единичный куб в $\mathbb{R}^n$ является компактным множеством.

Анализ для матриц Как только у нас есть понятие предела для матриц, мы можем с помощью него развивать анализ аналогичный анализу для обычных чисел. Например, можно определить хорошо известные гладкие функции от матриц. Скажем, пусть $A \in \mathrm{M}_n(\mathbb{R})$, тогда можно сказать, что значит e^A, $\ln A$, $\sin A$ или $\cos A$. Конечно, $\ln A$ будет существовать не для всех матриц A, так же как и обычный логарифм существует только для положительных чисел. Знакомые тождества вроде $e^{\ln A} = A$ и $\ln e^A = A$ будут оставаться справедливыми. Свойства $e^{A+B} = e^A e^B$ будет верным, в случае если A и B коммутируют.

Одним из простейших подходов к определению таких функций – использование степенных рядов. Например, $e^x = \sum_{k\geqslant 0} x^k/k!$. Тогда можно определить $e^A = \sum_{k\geqslant 0} A^k/k!$. Доказательство свойств экспоненты тогда сводится к игре в раскрытие скобок со степенными рядами. И в этой игре нам важно, чтобы символы были перестановочны между собой, потому какие-то свойства могут нарушиться, если исходные матрицы не коммутируют.

Еще стоит отметить такой момент. Так как для любой матрицы A существует минимальный многочлен ее зануляющий f_{min}, то, оказывается, что любую гладкую функцию от A можно приблизить многочленом. А именно, если φ – некоторая гладкая функция, то для любой матрицы A найдется такой многочлен f (зависящий от A) степени меньше, чем $\deg f_{min}$, что $\varphi(A) = f(A)$. Более того, существует общая алгоритмическая процедура по нахождению такого многочлена f. Эта процедура является эффективным способом вычисления гладких функций от матриц.