Курс “Методы линейной алгебры в теории приближений”. Лекция 1. Введение. Спектр эрмитовых матриц.

6 октября 2023 г.

Линейная алгебра вездесуща и жизненно необходима в таких областях математики, как теория функций и теория приближений. В этом спецкурсе мы углубиться в линейную алгебру — её понятия, методы, основополагающие теоремы. Поскольку нас так же интересует теория приближений, мы рассмотрим и разнообразные применения линейной алгебры к этой теории.

Начнём с простейших понятий. Линейное отображение $\mathcal{A}\colon V \to V$ линейного пространства в себя называется линейным оператором (в зависимости от контекста, линейным оператором могут называться также линейные отображения между разными пространствами). В этом курсе мы в основном будем рассматривать конечномерные пространства. Базис $\{v_1, \ldots, v_n\}$ пространства V позволяет записать оператор $\mathcal{A}$ в виде квадратной матрицы $A = (A_{i,j})$:

$$\mathcal{A}v_j = \sum A_{i,j} v_i.$$

При этом, если вектор $v \in V$ отождествить с вектором его координат в базисе $\{v_j\}$: $v = \sum c_k v_k$, то действие оператора будет соответствовать умножению матрицы A на вектор–столбец $(c_1, \ldots, c_n)^t$. Как обычно, t обозначает транспонирование.

Некоторые характеристики оператора являются инвариантами, т.е. не зависят от выбора базиса. Например, след

$$\operatorname{Tr} \mathcal{A} := \sum A_{i,i}.$$

Другие примеры инвариантов: ранг (далее обсудим его подробно), определитель, характеристический многочлен $\det(\mathcal{A} - t\mathcal{E})$. Множество собственных значений = нули характеристического многочлена.

Линейные пространства можно рассматривать над произвольным полем k, однако сейчас нас будут интересовать только $\mathbb{R}$ и $\mathbb{C}$. Ограничиться только вещественным полем нельзя, это сузит наши возможности!

Ранг. Это важное понятие допускает множество эквивалентных определений:

- размерность образа: $\dim\{\mathcal{A}x \colon x \in V\}$;
- размерность пространства, порождаемого столбцами матрицы A;
- размерность пространства, порождаемого строками матрицы A;
- минимальное число слагаемых в разложении A в виде суммы одноранговых матриц (т.е. матриц вида $B_{i,j} = u_i v_j$); отметим, что это определение симметрично по отношению к строкам и столбцам и допускает обобщение на тензоры;
- минимальная размерность r, такая что найдутся вектора $x_i, y_j \in \mathrm{k}^r$, такие что $A_{i,j} = \langle x_i, y_j \rangle$.

Конечно, нас будут интересовать не только инвариантные величины. В некоторых задачах базис уже задан и нужно работать в нём! Например, важная для нас норма $\|A\|_\infty := \max_{i,j} |A_{i,j}|$ меняется при изменении базиса. Впрочем, алгебраический, инвариантный взгляд на операторы в любом случае очень полезен.

Собственные значения и вектора. Мы только кратко напомним необходимые определения. Число λ называется собственным значением оператора $\mathcal{A}$, если оператор $\mathcal{A} - \lambda\mathrm{Id}$ вырожден, т.е. существует ненулевой вектор v, такой что $\mathcal{A}v = \lambda v$. Кратностью собственного значения называется размерность пространства собственных векторов, т.е. $\dim \operatorname{Ker}(\mathcal{A} - \lambda\mathrm{Id})$.

Собственные значения являются корнями характеристического многочлена и кратности совпадают. (Проверьте!)

Симметричные и эрмитовы операторы и матрицы. Пусть на пространстве V (над полем $\mathbb{R}$ или $\mathbb{C}$) задано скалярное произведение. В случае $\mathbb{C}$ оно полуторалинейно:

$$\langle \lambda x, y\rangle = \overline{\lambda}\langle \lambda x, y\rangle, \quad \langle x, \lambda y\rangle = \lambda\langle x, y\rangle.$$

При наличии этой дополнительной структуры имеет смысл рассматривать не произвольные базисы, а ортонормированные: $\langle f_i, f_j\rangle = \delta_{i,j}$.

Если зафиксировать в V ортонормированный базис и работать с векторами-столбцами, то скалярное произведение можно записывать кратко как $x^t y$ (в вещественном случае) или $x^* y$ (в комплексном случае). Как обычно, $x^* = \overline{(x^t)}$.

Оператор A на пространстве со скалярным произведением называется симметрическим (синоним: эрмитовым, если хотим подчеркнуть, что всё над полем $\mathbb{C}$), если выполнено равенство

$$\langle \mathcal{A}x, y\rangle = \langle x, \mathcal{A}y\rangle, \quad \forall x, y \in V.$$

Синоним — самосопряжённый оператор (есть тонкое отличие для не всюду определённых операторов в бесконечномерном случае; нам это не актуально).

Это условие равносильно тому, что в ортонормированном базисе матрица оператора:

- (в вещественном случае) симметрична, $A^t = A$;
- (в комплексном случае) эрмитова, $A^* = A$.

Спектральная теорема. Самосопряжённый оператор (над $\mathbb{R}$ или $\mathbb{C}$) диагонализуется в некотором ортонормированном базисе:

$$\mathcal{A}x = \sum_{k=1}^{n} \lambda_k \langle f_i, x\rangle f_i.$$

В терминах матриц: если вещественная матрица A симметрична (соотв., комплексная матрица A эрмитова), то существует ортогональная $U^t U = \mathrm{Id}$ (соотв., унитарная $U^* U = \mathrm{Id}$) матрица, приводящая A к диагональному виду: $U^{-1} A U = U^* A U = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

Числа λ_k являются собственными значениями A и потому определены однозначно. Условимся записывать эти числа в невозрастающем порядке:

$$\lambda_1(A) \geqslant \lambda_2(A) \geqslant \ldots \geqslant \lambda_n(A).$$

Доказательство. Рассмотрим комплексный случай. Какое-то собственное значение всегда существует (многочлен имеет корень), назовём его λ. Заметим, что оно вещественно: действительно, пусть v это соответствующий собственный вектор,

$$\langle Av, v\rangle = \langle \lambda v, v\rangle = \overline{\lambda}|v|^2$$

и с другой стороны,

$$\langle Av, v\rangle = \langle v, Av\rangle = \langle v, \lambda v\rangle = \lambda|v|^2$$

Откуда $\lambda = \overline{\lambda}$.

Далее, подпространство $W = \operatorname{span}\{v\}$ инвариантно, т.е. $\mathcal{A}W \subset W$. Докажем, что $W^\perp$ инвариантно. Действительно, если $v \perp W$, то

$$\langle \mathcal{A}v, u\rangle = \langle v, \mathcal{A}u\rangle = \langle v, u'\rangle = 0,$$

т.е. $\mathcal{A}v \in W^\perp$.

Так мы можем разложить всё пространство в сумму одномерных инвариантных подпространств, что и требовалось.

Вещественный случай. Будем рассуждать в терминах матриц. У матрицы A есть собственное значение λ. Оно вещественно (см. выше). Значит, и собственный вектор вещественный. Дальше рассуждение аналогично. $\square$

Минимаксные представления для спектра и неравенства Всюду далее в этой лекции A — эрмитова матрица.

Пусть $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \ldots \geqslant \lambda_n(A)$ — набор её собственных значений. Как работать с величиной $\lambda_i(A)$, какие у неё свойства? Например, сумма этих величин линейна по A:

$$\sum_{i=1}^n \lambda_i(A+B) = \sum_{i=1}^n \lambda_i(A) + \sum_{i=1}^n \lambda_i(B).$$

Действительно, это сумма равна следу, а он линеен.

Некоторые возможности даёт следующая теорема.

Теорема 1 (Курант–Фишер)**.** *Пусть A — эрмитова матрица.*

$$\lambda_i(A) = \max_{\dim V = i} \min_{\substack{v\in V\\ |v|=1}} v^* A v, \tag{1}$$

$$\lambda_i(A) = \min_{\dim V = n-i+1} \max_{\substack{v\in V\\ |v|=1}} v^* A v.$$

Заметим, что при фиксированном v величина v^*Av является линейной функцией матрицы A. Взяв минимум по $v \in V$, мы получаем уже вогнутую функцию (минимум линейных). Дальше берётся ещё максимум по V и структура функции ещё усложнится. Однако, из такого представления можно всё-таки извлечь много информации.

Доказательство. Достаточно доказать первое равенство, поскольку второе получается применением первого к матрице $-A$.

Пусть A диагонализуется в ортонормированном базисе $f_1, \ldots, f_n$. Тогда $v^*Av = \sum_{k=1}^n \lambda_k |v_k|^2$, где v_k это координаты v в этом базисе. Ясно, что если взять $V = \operatorname{span}\{f_1, \ldots, f_i\}$, то на V имеем $v^*Av \leqslant \lambda_i |v|^2$; это доказывает "$\leqslant$" в (1).

Для обратного неравенства нужно в любом i-мерном пространстве V найти единичный вектор с $v^*Av \leqslant \lambda_i$. Как это сделать: рассмотрим пересечение $V \cap \operatorname{span}\{f_i, \ldots, f_n\}$. В силу формулы $\dim(U+V) = \dim(U) + \dim(V) - \dim(U \cap V)$ мы имеем

$$\dim(V \cap \operatorname{span}\{f_i, \ldots, f_n\}) \geqslant i + (n - i + 1) - n = 1,$$

пересечение не пусто и содержит единичный вектор w. Ясно, что $w^*Aw \leqslant \lambda_i$, что и требовалось. $\square$

Посмотрим на равенство (1). Величина v^*Av линейна по A при фиксированном v. Минимум линейных функций является вогнутой функцией по A. Потом берётся максимум по подпространствам и получается более сложная зависимость λ_i от A.

Неравенство Вейля:

$$\lambda_{i+j-1}(A + B) \leqslant \lambda_i(A) + \lambda_j(B).$$

(Оставим в качестве упражнения.)

Докажем следующее неравенство:

$$|\lambda_i(A + B) - \lambda_i(A)| \leqslant \|B\|_{2 \mapsto 2}.$$

Справа стоит операторная норма: $\|B\|_{2 \mapsto 2} := \max_{|v|=1} |Bv|$. Нетрудно видеть, что $\|B\|_{2 \mapsto 2} = \max(|\lambda_1(B)|, |\lambda_n(B)|)$. Поскольку это одна из наиболее часто используемых норм, часто пишут просто $\|B\|$ без уточнения; мы будем делать так, если это не приведёт к путанице.

Итак, оценим $\lambda_i(A+B)$ сверху. По Куранту–Фишеру, достаточно в любом i-мерном пространстве V найти единичный вектор v с оценкой на $v^*(A+B)v$. Для матрицы A существует $v \in V$, такой что $v^*Av \leqslant \lambda_i(A)$. Далее, $v^*Bv \leqslant \|B\|$. Итого, мы пользуемся линейностью:

$$v^*(A+B)v = v^*Av + v^*Bv \leqslant \lambda_i(A) + \|B\|.$$

Оценка снизу аналогична.

Мы получили, что отображение, сопоставляющее матрице её вектор собственных значений, липшицево:

$$\|\vec{\lambda}(A+B) - \vec{\lambda}(B)\|_\infty \leqslant \|B\|_{2\mapsto 2}.$$

Обобщённый след. Пусть U — некоторое подпространство. Определим след A на подпространстве U как

$$\mathrm{Tr}(A|_U) := \sum_{i=1}^{k} u_i^* A u_i,$$

где $\{u_1, \ldots, u_k\}$ это ортонормированный базис A. Проверьте, что определение корректно и не зависит от выбора базиса.

Докажем следующее равенство:

$$\lambda_1(A) + \ldots + \lambda_i(A) = \max_{\dim V = i} \mathrm{Tr}(A|_V).$$

Из него следует, в частности, что сумма первых i собственных значений — выпуклая функция матрицы.

Доказательство. Ясно, что равенство достигается на $V = \mathrm{span}\{f_1, \ldots, f_i\}$. Для оценки в обратную сторону нужно оценить след сверху. Сделаем это по индукции. В случае $i = 1$ это просто Курант–Фишер. Для перехода от $i-1$ к i рассмотрим $V' := V \cap \mathrm{span}\{f_2, \ldots, f_n\}$. Мы переходим в подпространство, где у A собственные значения начинаются с λ_2 и по предположению индукции

$$\mathrm{Tr}(A|_{V'}) = \sum_{k=1}^{i-1} u_i^* A u_i \leqslant \lambda_2(A),$$

где $u_1, \ldots, u_{i-1}$ базис V'. Дополним его до базиса V добавлением вектора u_0. Поскольку $u_0^* A u_0 \leqslant \lambda_1(A)$, получаем требуемое. (Что, если $\dim V' = i$?) $\square$