Спецкурс 2020/2021: “Геометрические и комбинаторные свойства матриц и аппроксимация” Блок лекций “Сложность матриц и аппроксимация” Лекция 7: “Ранг тензоров”

28 ноября 2020 г.

Тензоры

Тензор порядка d размера $n_1 \times n_2 \times \dots \times n_d$ — это массив чисел

$$T = (T_{i_1,\dots,i_d}), \quad i_1 \in [n_1], \dots, i_d \in [n_d].$$

Напомним, $[n] := \{1, 2, \dots, n\}$.

Числа $n_1, \dots, n_d$ называются *размерностями* тензора. Обозначаем множество таких тензоров как $\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$.
Очевидно, тензор размерности $(n_1, \dots, n_d)$ состоит из $N = n_1 n_2 \dots n_d$ элементов.

Можно считать, что тензор — это функция нескольких (дискретных) аргументов.

- скаляр $x \in \mathbb{R}$ — тензор порядка 0;
- вектор $x \in \mathbb{R}^n$ — тензор порядка 1;
- матрица $x \in \mathbb{R}^{n_1 \times n_2}$ — тензор порядка 2;
- тензор $x \in \mathbb{R}^{n_1 \times \dots \times n_d}$ порядка d.

Алгебраический взгляд

Матрицу можно отождествить с линейным оператором, это даёт много полезных и важных понятий (собственные и сингулярные числа, операторные нормы, инвариантность следа и определителя и т.п.). Каков алгебраический взгляд на тензоры?

В более общем смысле тензор — это элемент тензорного произведения пространств. Тензорное произведение линейных пространств U и V состоит из (формальных) линейных комбинаций выражений вида $u \otimes v$, $u \in U$, $v \in V$, в котором некоторые комбинации отождествлены — так, чтобы выполнялись условия билинейности

$$(\lambda_1 u_1 + \lambda_2 u_2) \otimes v = \lambda_1 (u_1 \otimes v) + \lambda_2 (u_2 \otimes v),$$

и аналогично для второго аргумента.
Тензоры вида $u \otimes v$ называются *разложимыми*, они порождают всё тензорное произведение $U \otimes V$ (но составляют лишь малую его часть).
Можно показать, что если $\{u_1, \ldots, u_m\}$ — базис U, а $\{v_1, \ldots, v_m\}$ — базис V, то $\{u_i \otimes v_j\}$ будет базисом $U \otimes V$. Значит, $\dim(U \otimes V) = \dim(U) \dim(V)$. Аналогично для нескольких пр-в.

Тензорное произведение и функции многих переменных

Выберем в пространстве U базис $\{u_i : i \in I\}$. Тогда U можно отождествить с функциями $f : I \to \mathbb{R}$. Аналогично для V с базисом $\{v_j : j \in J\}$: $g : J \to \mathbb{R}$.

В тензорном произведении $U \otimes V$ базис — $\{u_i \otimes v_j\}$, поэтому $U \otimes V$ можно отождествить с пространством функций $h : I \times J \to \mathbb{R}$. Разложимые тензоры соответствуют функциям вида $h(i, j) = f(i)g(j)$.

То есть, тензорное произведение для пространств функций соответствует переходу к декартовому произведению областей определения.

- многочлены: $\mathbb{R}[x] \otimes \mathbb{R}[y] \cong \mathbb{R}[x, y]$;
- разложимые функции от d переменных: $f_1(x_1)f_2(x_2)\dots f_d(x_d)$;
- многомерная тригонометрическая система $\exp(i(k_1x_1 + \dots k_dx_d))$ состоит из разложимых функций;
- для функциональных пространств то же самое, например, $C(X) \otimes C(Y) = C(X \times Y)$ для хороших (компактных) X, Y и при подходящем определении нормы тензорного произведения.

Алгебраический взгляд

В линейной алгебре тензоры (в нашем понимании) возникают как координаты элементов тензорных произведений. Пусть $U_1, \dots, U_d$ — линейные пространства с базисами $\{u_1^1, \dots, u_{n_1}^1\}$, ..., $\{u_1^d, \dots, u_{n_d}^d\}$, соответственно. Тогда вектора

$$u_{i_1}^1 \otimes u_{i_2}^2 \otimes \cdots \otimes u_{n_d}^d$$

образуют базис пространства $U_1 \otimes U_2 \otimes \cdots \otimes U_d$.
Значит, $\mathcal{T} \in U_1 \otimes \cdots \otimes U_d$ раскладывается по базису:

$$\mathcal{T} = \sum_{i_1 \in [n_1], \dots i_d \in [n_d]} T_{i_1, \dots, i_d} u_{i_1}^1 \otimes \cdots \otimes u_{n_d}^d.$$

При замене базисов в U_j координаты тензора преобразуются по известным законам; при этом некоторые свойства инвариантны относительно замены базиса (см. далее).
Пространство $\mathbb{R}^n$ обладает естественным базисом; мы можем отождествить тензорное произведение (в алгебраическом смысле) с пространством тензоров (как массивов чисел):
$\mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_d} \cong \mathbb{R}^{n_1 \times \ldots \times n_d}$.

Пример: оператор

Тензорное произведение $V^* \otimes V$ можно отождествить с пространством операторов на V:

$$V^* \otimes V \cong L(V, V), \quad (\xi \otimes v)\colon u \mapsto \xi(u)v.$$

Выберем в V базис $\{e_j\}$, в V^* возникает сопряжённый базис $\{\xi_i\}$, т.е. такой что $\xi_i(e_j) = \delta_{i,j}$. Тензор раскладыватеся по базису:

$$V^* \otimes V \ni \mathcal{T} = \sum_{i,j} T_{i,j} \xi_i \otimes e_j.$$

Вычислим $\mathcal{T} e_k$:

$$\mathcal{T} e_k = (\sum T_{i,j} \xi_i \otimes e_j) e_k = \sum_{i,j} T_{i,j} \xi_i(e_k) e_j = \sum_j T_{k,j} e_j.$$

Таким образом, $(T_{i,j})$ это транспонированная матрица оператора $\mathcal{T}$.

Пример: билинейное отображение

Пространство $L(U \times V, W)$ билинейных отображений $\mathcal{T} : U \times V \to W$ отождествляется с тензорным произведением $U^* \otimes V^* \otimes W$. Выбор базисов $\{u_i\}$, $\{v_j\}$, $\{w_k\}$ даёт разложение

$$\mathcal{T} = \sum_{i,j,k} T_{i,j,k} \xi_i \otimes \eta_j \otimes w_k.$$

При этом

$$\mathcal{T}(u_p, v_q) = \sum_{i,j,k} T_{i,j,k} \xi_i(u_p) \eta_j(v_q) w_k = \sum_k T_{p,q,k} w_k.$$

К сожалению, для тензоров порядка $d \geqslant 3$ нет настолько же удобного соответствия, как соответствие между матрицами и операторами.

Ранг матрицы

Напомним эквивалентные определения ранга матрицы:

- размерность образа $\dim \mathrm{Im}\mathcal{A}$ оператора с матрицей A;
- размерность пространства строк/столбцов;
- максимальный размер невырожденного минора: $\max\{|I| = |J| : \det A[I, J] \neq 0\}$;
- минимальная размерность r, в которой найдутся вектора $x_i \in \mathbb{R}^r$, $y_j \in \mathbb{R}^r$, такие что $A_{i,j} = \langle x_i, y_j \rangle$;
- минимальное число одноранговых матриц (т.е. вида $R_{i,j} = a_i b_j$) в представлении $A = R^{(1)} + R^{(2)} + \ldots + R^{(r)}$.

Обобщим понятие ранга на основании последнего определения.

Ранг тензора

Назовём тензор $T \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ *разложимым*, если

$$T_{i_1,\ldots,i_d} = v^1_{i_1} v^2_{i_2} \ldots v^d_{i_d}, \quad i_1 \in [n_1], \ldots, i_d \in [n_d],$$

для некоторых векторов $v^1, \ldots, v^d$. Например, для тензора порядка 3:

$$T_{i,j,k} = a_i b_j c_k, \quad a \in \mathbb{R}^{n_1}, b \in \mathbb{R}^{n_2}, c \in \mathbb{R}^{n_3}.$$

Рангом тензора назовём минимальное количество разложимых тензоров в представлении $T = R^{(1)} + R^{(2)} + \ldots + R^{(r)}$. Таким образом, "разложимый" это синоним слову "одноранговый" (кроме нулевого).

Пример: пусть $n_1 = n_2 = n_3$. Чему равен ранг тензора $T_{i,j,k} = \mathbf{1}\{i = j = k\}$? Ответ: $\operatorname{rank} T = n$.

Ранг в общем виде

Сформулируем понятие ранга в общем, инвариантном виде. Пусть

$$\mathcal{T} \in U_1 \otimes U_2 \otimes \cdots \otimes U_d.$$

Напомним, что разложимыми тензорами называются тензоры вида $u_1 \otimes u_2 \otimes \cdots \otimes u_d$, $u_j \in U_j$. Они порождают всё тензорное произведение (но составляют малую его часть).

Ранг тензора $\mathcal{T}$ в алгебраическом смысле равен рангу его координат $T \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ в фиксированном базисе. При этом понятия разложимости и ранга не зависят от выбора базиса!

Свойства ранга

Известно, что множество матриц ранга не выше r замкнуто. Действительно, если $\operatorname{rank} A > r$, то некоторый $(r+1)\times(r+1)$ минор невырожден, это же верно и в окрестности матрицы.

В пространстве тензоров естественным образом возникает топология. Можно рассматривать различные нормы, например, норму Фробениуса

$$\|T\|_F^2 := \sum_{i_1,\dots,i_d} T_{i_1,\dots,i_d}^2.$$

Множество разложимых тензоров замкнуто: если, скажем, $T^{(n)} = a^{(n)} \otimes b^{(n)} \otimes c^{(n)}$ и $T^{(n)} \to T$, то, нормируя $\|b^{(n)}\|_F = \|c^{(n)}\|_F = 1$, мы получим, что и нормы $\|a^{(n)}\|$ ограничены. Поэтому можно выбрать сходящуюся подпоследовательность

$$a^{(n_k)} \to a, \quad b^{(n_k)} \to b, \quad c^{(n_k)} \to c,$$

откуда $T^{(n)} \to a \otimes b \otimes c$.

В общем случае замкнутости уже нет.

Утверждение

В пространстве $\mathbb{R}^{n_1 \times n_2 \times n_3}$ при $n_j \geqslant 3$ существует последовательность тензоров ранга 2, предел которой равен тензору ранга 3.

В (1) слева стоят тензоры ранга 2, а справа — ранга 3 (при правильном выборе x_i, y_i):

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}((x_1 + \varepsilon y_1) \otimes (x_2 + \varepsilon y_2) \otimes (x_3 + \varepsilon y_3) - x_1 \otimes x_2 \otimes x_3) = x_1 \otimes x_2 \otimes y_3 + x_1 \otimes y_2 \otimes x_3 + y_1 \otimes x_2 \otimes x_3. \quad (1)$$

Проблема состоит в том, что одноранговые слагаемые имеют большую норму, но непостижимым образом "сокращаются" друг с другом и норма суммы оказывается малой. Если рассматривать *регулярные* представления, в которых нормы слагаемых ограничены:

$$T = \sum_{s=1}^{r} a_1^{(s)} \otimes a_2^{(s)} \otimes \cdots \otimes a_d^{(s)}, \quad \max_s \|a_1^{(s)}\|_F \cdot \|a_2^{(s)}\|_F \cdots \|a_d^{(s)}\|_F < C,$$

то подобного эффекта не будет.

Граничный ранг (border rank)

Скажем, что тензор T имеет *граничный ранг* не выше r , если T является пределом некоторой последовательности тензоров ранга не выше r. Обозначение: $\underline{\mathrm{rank}}(T)$.

Например, для $T = x \otimes y + x \otimes z + y \otimes z$ имеем $\mathrm{rank}(T) = 3$, $\underline{\mathrm{rank}}(T) = 2$.

Упражнение. Придумайте тензор T с $\mathrm{rank}\, T - \underline{\mathrm{rank}} T \geqslant 2$.

Известно (2017?), что отношение $\mathrm{rank}(T)/\underline{\mathrm{rank}\, T}$ может быть сколь угодно большим.

Тензор матричного умножения

Рассмотрим операцию умножения матриц подходящего размера. Мы "вытянем" матрицы в длинные вектора, т.е. будем считать индексы $A_{i,j}$ одним индексом (i,j). Тензор M умножения матриц имеет вид

$$C = AB \quad \Leftrightarrow \quad C_{(i,j)} = \sum_{(k,l),(p,q)} M_{(i,j),(k,l),(p,q)} A_{(k,l)} B_{(p,q)}.$$

Его размерности это $(n_1 n_3, n_1 n_2, n_2 n_3)$.
Как мы знаем, тензор умножения равняется

$$M_{(i,j),(k,l),(p,q)} = \mathbf{1}\{k = i, l = p, q = j\},$$

тогда получается привычная формула $C_{i,j} = \sum A_{i,p} B_{p,j}$.

Тензор умножения матриц

Предположим, ранг этого тензора равен r и мы имеем разложение

$$M = \sum_{s=1}^{r} \alpha^{(s)} \otimes \beta^{(s)} \otimes \gamma^{(s)}.$$

Тогда

$$C_{(i,j)} = \sum_{(k,l),(p,q)} \sum_{s=1}^{r} \alpha^{(s)}_{(i,j)} \beta^{(s)}_{(k,l)} \gamma^{(s)}_{(p,q)} A_{(k,l)} B_{(p,q)} =$$

$$= \sum_{s=1}^{r} \alpha^{(s)}_{(i,j)} \Big(\sum_{(k,l)} \beta^{(s)}_{(k,l)} A_{(k,l)}\Big)\Big(\sum_{(p,q)} \gamma^{(s)}_{(p,q)} B_{(p,q)}\Big).$$

Заметим, что множители с β и γ не зависят от (i,j). Следовательно, достаточно вычислить эти числа, сделать r умножений и найти все $C_{(i,j)}$ как линейные комбинации.

Тензор умножения матриц (2×2)

Рассмотрим теперь матрицы 2×2. Оказывается, ранг тензора умножения равен 7! Это позволяет умножить две 2×2 матрицы, сделав только 7 умножений (обычный алгоритм требует 8 умножений). Но главное в том, что можно считать элементы матриц 2×2 не числами, а матрицами (формула для умножения справедлива и для блоков)! Поэтому умножение матриц $(2n) \times (2n)$ сводится к 7 умножениям матриц $n \times n$! Общая сложность алгоритма не n^3 (как у тривиального), а $n^{\log_2 7} = n^{2.807\ldots}$. Этот алгоритм придумал Volker Strassen (1969).

Показатель степени улучшался за счёт рассмотрения тензоров умножения матриц большего размера. Есть также оценки, основанные на rank, а не на обычном ранге!

Максимальный ранг

Каков максимально возможный ранг тензора из $\mathbb{R}^{n_1 \times \ldots \times n_d}$?

Утверждение

$$\frac{n_1 n_2 \cdots n_d}{n_1 + \ldots + n_d} \leqslant \max \operatorname{rank} T \leqslant n_1 n_2 \cdots n_d / \max\{n_j\}.$$

Докажем оценку сверху. Пусть размерность n_1 максимальна. Мы можем представить T в виде суммы

$$T_{i_1,\ldots,i_d} = \sum_{i'_2,\ldots,i'_d} T_{i_1,i'_2,\ldots,i'_d} \delta_{i_2,i'_2} \cdots \delta_{i_d,i'_d}.$$

Каждое слагаемое представляет собой разложимый тензор.

Оценка снизу получается из соображений размерности. Тензор ранга r задаётся r наборами векторов $u_1^{(s)} \in \mathbb{R}^{n_1}, \ldots, u_d^{(s)} \in \mathbb{R}^{n_d}$. Следовательно, такой тензор описывается $r(n_1 + \ldots + n_d)$ параметрами. Чтобы покрыть всё $n_1 n_2 \ldots n_d$-мерное пространство $\mathbb{R}^{n_1 \times \cdots \times n_d}$, должно быть выполнено неравенство

$$r(n_1 + \ldots + n_d) \geqslant n_1 \cdots n_d.$$

Максимальный ранг

В случае тензоров $\mathbb{R}^{n\times n\times\cdots\times n}$ порядка d максимальный ранг заключён между n^{d-1}/d и n^{d-1}.

Проблема. Придумать конструктивный пример тензора из $\mathbb{R}^{n\times n\times n}$ ранга $\gg n^{1+\varepsilon}$, $\varepsilon > 0$.
Strassen доказал, что сложность вычисления трилинейной формы, задаваемой тензором T, не меньше $C\operatorname{rank} T$. Поэтому суперлинейные оценки ранга дают интересные следствия в теории сложности вычислений. (Ср. с теоремой Valiant-а).
Известны явные конструкции тензоров лишь ранга $\geqslant 3n - o(n)$.

Тензорный ранг, как мы видим, намного сложнее матричного. Доказано, что задача вычисления тензорного ранга при $d \geqslant 3$ надо полем $\mathbb{Q}$ является NP-трудной.

Multiparty communication

Рассмотрим задачу коммуникации с $k \geqslant 2$ участниками. Каждому выдаётся элемент $x_i \in X_i$, $i = 1, \ldots, k$. Требуется вычислить $f(x_1, x_2, \ldots, x_k)$.

Функцию $f\colon X_1 \times \cdots \times X_k \to \{0,1\}$ можно отождествить с тензором из $\{0,1\}^{n_1 \times n_2 \times \cdots \times n_d}$, $N_j = |X_j|$.

Рассмотрим прямое обобщение случая $k = 2$: участникам известны только свои входные данные. Будем считать, что они обмениваются информацией, записывая сообщения "на доске", так, чтобы все видели (broadcast).

Зафиксируем протокол. Как и раньше, введём понятие *истории* сообщений. Множество входов, приводящих к данной истории:

$$R_h = \{(x_1, \ldots, x_k)\colon P(x_1) = h_1,\ P(h_1, x_2) = h_2,\ P(h_1, h_2, x_3) = h_3, \ldots\}$$

представляет собой обобщённый "прямоугольник": $R_h = I_1 \times I_2 \times \cdots \times I_d$.

Отсюда, как и раньше, получаем оценку сложности

$$D(f) \geqslant \log_2 \operatorname{rank} f.$$

Multiparty communication

Перейдём к модели коммуникации “число на лбу”. Теперь каждый участник видит все x_j, кроме своего. Здесь у участников больше информации и появляются новые эффекты.

Пример. Функция равенства: $\mathrm{EQ}(x_1, \dots, x_k) = \mathbf{1}\{x_1 = x_2 = \dots = x_k\}$, для одинаковых множеств $X_1 = \dots = X_k$ мощности N.

В случае коммуникации двух участников мы доказали оценку сложности $D(\mathrm{EQ}) \geqslant \log N$. Что будет в случае $k \geqslant 3$ участников? Сложность падает до $D(\mathrm{EQ}) = 2$. Следовательно, $\log \mathrm{rank}$ оценки уже нет.

Однако, по-прежнему справедлива оценка через дискрепанс:

$$D(f) \gg \log(1/\mathrm{disc}\,(f)).$$

Вместо обобщённых прямоугольников, правда, приходится рассматривать более хитрые множества. Для $k = 3$ это множества вида

$$(I_1 \times I_2 \times X_3) \cap (I_1' \times X_2 \times I_3') \cap (X_1 \times I_2'' \times I_3'').$$

W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*.

D. Knuth, *The Art of Computer Programming*, vol.2, §4.6.