

МАТРИЦЫ, ТЕНЗОРЫ, ВЫЧИСЛЕНИЯ

Е. Е. ТЫРТЫШНИКОВ

1 Умножение тензоров и разложение Таккера

1.1 Матрицы и элементы матриц

Под матрицей понимается прямоугольная таблица, в которой на пересечении строки с номером i и столбца с номером j расположен элемент $a(i, j)$. Таким образом, матрица определяется функцией $a(i, j)$ от переменных $i \in I$ и $j \in J$, где I и J – некоторые конечные множества. При этом на множествах I и J фиксируется порядок просмотра их элементов, определяющий нумерацию строк и столбцов матрицы. Последнее является существенным при изображении матриц в виде таблиц. Однако, для определения основных операций над матрицами, прежде всего для *умножения матриц*, порядок не очень важен.

Если заданы две функции $a(i, j)$ и $b(j, k)$, то общий символ j означает, что данная переменная изменяется на одном и том же множестве J . Элементы произведения матриц определяются следующим образом:

$$c(i, k) = \sum_{j \in J} a(i, j)b(j, k). \quad (1)$$

Коммутативность сложения позволяет не фиксировать порядок элементов множества J .

Матрица C с элементами $c(i, k)$ называется произведением матрицы A с элементами $a(i, j)$ и матрицы B с элементами $b(j, k)$. Обычная запись: $C = AB$. Мы хотим ввести в употребление и узаконить следующий оборот речи: *матрица $c(i, k)$ является произведением матриц $a(i, j)$ и $b(j, k)$* .

Выражение $a(i, j)$ обычно означает *элемент матрицы*, но мы хотим придать этому выражению другой смысл, используя его как *обозначение матрицы*. Тогда вместо формулы $C = AB$ логично писать

$$c(i, k) = a(i, j)b(j, k). \quad (2)$$

Для теории матриц это несколько необычное обозначение, но оно представляется нам очень удобным при изучении тензоров.

Безусловно, возникает конфликт обозначений: в формулах (1) и (2) одни и те же символы имеют действительно разный смысл. Чтобы получить независимость от контекста, можно договориться, например, об использовании квадратных скобок и считать, что запись $a[i, j]$ *всегда* означает элемент матрицы $a(i, j)$. Тогда матрица

$c(i, k)$ будет определяться формулами

$$c[i, k] = \sum_j a[i, j]b[j, k].$$

Формула (2) становится логически естественной, если i, j, k трактовать как обозначения множеств для соответствующих переменных. Некоторое неудобство возникает, когда нужно выделить конкретные номера из этих множеств. Чтобы указать явно, что речь идет именно о матрице, а не о ее отдельном элементе, будем писать $a([i], [j])$. Таким образом, $a[i, j]$ – это элемент матрицы $a([i], [j])$, для которой мы используем более короткое обозначение $a(i, j)$.

Запись $a([i], j)$ означает j -й столбец матрицы $a(i, j)$. Аналогично, $a(i, [j])$ есть i -я строка той же матрицы.

1.2 Скелетное разложение и ранг матрицы

Матрица $A = a(i, j)$ с элементами вида $a[i, j] = u(i)v(j)$ называется *скелетоном*. Представление матрицы в виде суммы скелетонов называется ее *скелетным разложением* или *каноническим разложением*. Сумма r скелетонов определяется системами столбцов $u_1, \dots, u_r$ и $v_1, \dots, v_r$ и имеет вид

$$A = \sum_{\alpha=1}^r u_{\alpha} v_{\alpha}^{\top} = UV,$$

где

$$U = [u_1, \dots, u_r], \quad V = [v_1, \dots, v_r]^{\top}.$$

В принятых нами обозначениях скелетное разложение можно записать в виде

$$a(i, j) = \sum_{\alpha} u([i], \alpha) v(\alpha, [j]) = u(i, \alpha)v(\alpha, j).$$

Число скелетонов в этом разложении равно числу значений индекса α . Минимально возможное число скелетонов называется *рангом* матрицы $a(i, j)$. Оно же равно размерности линейной оболочки, натянутой на столбцы матрицы, и совпадает с размерностью линейной оболочки строк и с максимальным порядком ненулевых миноров данной матрицы.

1.3 Сингулярное разложение матрицы

Пусть имеется скелетное разложение комплексной матрицы A , записанное в виде

$$A = \sum_{\alpha=1}^r \sigma_{\alpha} u_{\alpha} v_{\alpha}^*, \tag{3}$$

где

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$$

и каждая из систем столбцов $u_1, \dots, u_r$ и $v_1, \dots, v_r$ является ортонормированной при естественном скалярном произведении:

$$u_\alpha^* u_\beta = v_\alpha^* v_\beta = \begin{cases} 1, & \alpha = \beta, \\ 0, & \alpha \neq \beta. \end{cases}$$

Скелетное разложение вида (3) с такими свойствами называется *сингулярным разложением* матрицы A .

Нетрудно доказать, что $r = \text{rank} A$ в том и только том случае, когда $\sigma_r > 0$. Числа $\sigma_1, \dots, \sigma_r$ называются *сингулярными числами* матрицы A , а векторы $u_1, \dots, u_r$ и $v_1, \dots, v_r$ называются *сингулярными векторами* матрицы A . Равенство (3) равносильно соотношениям

$$Av_\alpha = \sigma_\alpha u_\alpha, \quad A^* u_\alpha = \sigma_\alpha v_\alpha, \quad 1 \leq \alpha \leq r.$$

Квадраты ненулевых сингулярных чисел с учетом кратностей совпадают с ненулевыми собственными значениями эрмитовых неотрицательно определенных матриц $A^* A$ и AA^* .

В обязательные курсы линейной алгебры обычно включаются следующие результаты – одни из самых востребованных среди всего многообразия приложений линейной алгебры.

Теорема 1.3.1 *Любая комплексная матрица имеет сингулярное разложение с числом ненулевых скелетов, равным ее рангу.*

Теорема 1.3.2 *Пусть A_k получается из сингулярного разложения матрицы A отбрасыванием скелетов с номерами $\alpha \geq k + 1$. Тогда*

$$\min_{\text{rank} B \leq k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1},$$

$$\min_{\text{rank} B \leq k} \|A - B\|_F = \|A - A_k\|_F = \sqrt{\sum_{\alpha \geq k+1} \sigma_\alpha^2}.$$

Пусть $\sigma(\alpha, \beta)$ – диагональная матрица с диагональными элементами $\sigma[\alpha, \alpha] = \sigma_\alpha$. Тогда в согласии с нашими обозначениями сингулярное разложение матрицы $a(i, j)$ можно записать в виде

$$a(i, j) = u(i, \alpha) \sigma(\alpha, \beta) v(\beta, j),$$

где

$$u([i], \alpha) = u_\alpha, \quad v(\beta, [j]) = v_\beta^*.$$

1.4 Тензоры и элементы тензоров

Тензором *размерности d* (или, чаще говорят, *порядка d*) называется d -мерная таблица (d -мерная матрица), составленная из элементов $a(i_1, \dots, i_d)$. Иногда тензор размерности d называется также d -тензором.

Считаем, что $i_1 \in I_1, \dots, i_d \in I_d$. В конечных множествах $I_1, \dots, I_d$ обычно фиксируется порядок просмотра элементов. Ниже мы увидим, что при определении операции *умножения тензоров* этот порядок оказывается не очень важным.

Как и в случае матриц, выражение $a(i_1, \dots, i_d)$ мы будем использовать для обозначения самого тензора как функции от переменных $i_1, \dots, i_d$. Таким образом, выражение $a(i_1, \dots, i_d)$ может иметь два разных смысла: обозначение тензора или отдельного его элемента. Что имеется в виду, должно быть ясно из контекста. Если же требуется определенно указать отдельное значение этой функции, т.е. элемент тензора, будем писать $a[i_1, \dots, i_d]$.

Если множества $I_1, \dots, I_d$ содержат $n_1, \dots, n_d$ номеров соответственно, то говорят, что тензор $a(i_1, \dots, i_d)$ имеет *размер* $n_1 \times \dots \times n_d$. Для этого же тензора можно использовать букву a без скобок и индексов там, где эта информация не является существенной.

Запись $a(i_1, \dots, i_d)$ как обозначение тензора совершенно естественна, если $i_1, \dots, i_d$ трактовать как множества изменения соответствующих индексов. Если нужно указать явно, что речь идет о множествах, будем писать $[i_1], \dots, [i_d]$. Таким образом, $a[i_1, \dots, i_d]$ – это отдельный элемент тензора $a([i_1], \dots, [i_d])$, при этом $a(i_1, \dots, i_d)$ есть более короткое обозначение для того же тензора.

1.5 Линейное пространство тензоров

Операция сложения тензоров одного и того же размера определяется как поэлементное сложение. Операция умножения тензора на число определяется как умножение каждого элемента тензора на это число.

Пусть $\mathbb{P}^{n_1 \times \dots \times n_d}$ обозначает множество тензоров $a(i_1, \dots, i_d)$ размера $n_1 \times \dots \times n_d$ с элементами $a[i_1, \dots, i_d]$ из поля $\mathbb{P}$.

Теорема 1.5.1 *Множество $\mathbb{P}^{n_1 \times \dots \times n_d}$ с операциями сложения и умножения на числа из $\mathbb{P}$ образует линейное пространство размерности $\dim V = n_1 \dots n_d$.*

Доказательство элементарно и предоставляется читателю.

В случае вещественного или комплексного поля вводится естественное скалярное произведение тензоров $a(i_1, \dots, i_d)$ и $b(i_1, \dots, i_d)$. По определению,

$$(a, b) = \sum_{i_1, \dots, i_d} a[i_1, \dots, i_d] \bar{b}[i_1, \dots, i_d],$$

черта означает комплексное сопряжение. Норма $\|a\|_F = \sqrt{(a, a)}$ иногда называется *нормой Фробениуса*.

1.6 Соглашение о суммировании при умножении тензоров

Пусть заданы тензоры $a_1(i_1^1, \dots, i_{d_1}^1), \dots, a_s(i_1^s, \dots, i_{d_s}^s)$ и множество всех букв, обозначающих их индексы, состоит из букв $i_1, \dots, i_k$, встречающихся только один раз, и букв $j_1, \dots, j_l$, встречающихся два или большее число раз. Тогда *произведением*

тензоров $a_1, \dots, a_s$ называется тензор $b(i_1, \dots, i_k)$, элементы которого определяются следующим образом:

$$b[i_1, \dots, i_k] = \sum_{j_1, \dots, j_l} a_1[i_1^1, \dots, i_{d_1}^1] \dots a_s[i_1^s, \dots, i_{d_s}^s]. \quad (4)$$

Если общих индексов не имеется, то в этом определении следует убрать знак суммы. Для определенности можно было бы считать, что порядок букв $i_1, \dots, i_k$ соответствует их порядку в составной последовательности аргументов $i_1^1, \dots, i_{d_s}^s$ заданных тензоров. Однако, этот порядок удобнее выбирать в зависимости от ситуации.

ПРИМЕР 1. Произведением тензоров $a(i, j, k)$ и $b(l, m)$ будет тензор

$$c(i, j, k, l, m) = a(i, j, k) b(l, m),$$

имеющий элементы

$$c[i, j, k, l, m] = a[i, j, k] b[l, m].$$

ПРИМЕР 2. Произведением тензоров $a(i, j, k)$ и $b(k, m)$ будет тензор

$$c(i, j, m) = a(i, j, k) b(k, m),$$

имеющий элементы

$$c[i, j, m] = \sum_k a[i, j, k] b[k, m].$$

ПРИМЕР 3. Произведение двумерных тензоров с общим индексом вида

$$a(i, j) = u(i, \alpha) v(j, \alpha)$$

означает умножение матрицы u на матрицу, транспонированную к матрице v .

ПРИМЕР 4. Сингулярное разложение двумерного тензора $a(i, j)$ можно записать в виде

$$a(i, j) = u(i, \alpha) \sigma(\alpha) \bar{v}(j, \alpha),$$

где $\sigma(\alpha)$ – упорядоченные по невозрастанию сингулярные числа данной матрицы, $u(i, \alpha)$ и $v(j, \alpha)$ – матрицы, столбцы которых составлены из сингулярных векторов, черта означает комплексное сопряжение. Ортонормированность сингулярных векторов выражается соотношениями

$$u(i, \alpha) \bar{u}(i, \beta) = v(j, \alpha) \bar{v}(j, \beta) = \delta(\alpha, \beta),$$

где

$$\delta[\alpha, \beta] = \begin{cases} 1, & \alpha = \beta, \\ 0, & \alpha \neq \beta. \end{cases}$$

Отметим также справедливость соотношения

$$\bar{u}(i, \alpha) a(i, j) v(j, \beta) = \delta(\alpha, \gamma) \sigma(\gamma) \delta(\gamma, \beta).$$

Предполагается, конечно, что $[\alpha] = [\beta] = [\gamma]$.

Следует подчеркнуть, что для тензоров a и b запись ab не дает ясности, как именно определяется произведение. Чтобы это сделать, необходимо указать индексы, причем имеются в виду буквы для обозначения индексов, а не отдельные их значения. Индексы (буквы), принадлежащие одному и тому же множеству, могут быть обозначены одной и той же буквой, при определении произведения по всем общим индексам производится суммирование. Если эти же индексы обозначены разными буквами, то суммирования по этим индексам нет. Обратим внимание также на то, что множество индексов (букв) тензора ab однозначно определяется индексами тензоров a и b , но их порядок правилом умножения не предписывается.

Наше соглашение о суммировании по общим индексам формально допускает также следующий способ уменьшения числа измерений (индексов) тензора. Например, если задан тензор $a(i, j, k)$, то из него можно получить тензор $c(i, k) = a(i, i, k)$, имеющий элементы

$$c[i, k] = \sum_i a[i, i, k].$$

Конечно, это возможно лишь в том случае, когда $[i] = [j]$, т.е. индексы i и j принадлежат общему множеству.

ПРИМЕР 5. Числа $a(i, i)$ и $a(i, j)\bar{a}(i, j)$ представляют след и фробениусову норму матрицы $a(i, j)$.

1.7 Равенство тензоров

Тензоры $a(i_1, \dots, i_d)$ и $b(j_1, \dots, j_d)$ называются *равными*, если совпадают индексные множества

$$I_1 = [i_1] = [j_1], \dots, I_d = [i_d] = [j_d]$$

и определенные на них элементы

$$a[i_1, \dots, i_d] = b[i_1, \dots, i_d] \quad \text{для любых} \quad i_1 \in I_1, \dots, i_d \in I_d.$$

Таким образом, для индексных множеств одного и того же тензора в разных ситуациях могут использоваться разные обозначения. Заметим, что в соотношениях между тензорами, например при умножении тензоров, *не допускается* формальная замена тензора произвольным равным ему тензором. Это связано с принятым нами способом вычисления произведения тензоров с помощью суммирования по общим индексам.

1.8 Матрицы, ассоциированные с тензором

Для трехмерного тензора $a(i, j, k)$ размера $n_1 \times n_2 \times n_3$ естественным образом определяются *матрицы развертки* a_1 , a_2 и a_3 по каждому из трех измерений:

$$a_1[i; jk] = a_2[j; ik] = a_3[k; ij] = a[i, j, k].$$

Матрицы a_1 , a_2 и a_3 имеют соответственно размеры $n_1 \times n_2 n_3$, $n_2 \times n_1 n_3$ и $n_3 \times n_1 n_2$. Мы используем точку с запятой, чтобы разделить строчные и столбцовые индексы матрицы. Кроме того, здесь впервые у нас появляются *составные индексы*, или *мультииндексы*: jk , ik и ij .

По определению, составной индекс ij принимает столько значений, сколько имеется пар значений для i и j . При изображении матриц в виде таблиц будем считать, что для составных индексов используется *лексикографический порядок*: в паре индексов (i, j) сначала изменяется j , затем i . Если i и j принимают значения 1, 2, 3 и 1, 2 соответственно, то пары (i, j) выстраиваются в последовательность

$$(1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2).$$

Аналогичным образом определяется лексикографический порядок для составных индексов ijk , $ijkl$ и так далее. Вместе с тем заметим, что в большинстве ситуаций порядок значений составного индекса не очень важен.

ПРИМЕР. Рассмотрим тензор размера $2 \times 2 \times 2$ с элементами

$$\begin{aligned} a[1, 1, 1] &= 111, & a[1, 1, 2] &= 112, & a[1, 2, 1] &= 121, & a[1, 2, 2] &= 122, \\ a[2, 1, 1] &= 211, & a[2, 1, 2] &= 212, & a[2, 2, 1] &= 221, & a[2, 2, 2] &= 222. \end{aligned}$$

Вот матрицы развертки этого тензора по измерениям:

$$a_1 = \begin{bmatrix} 111 & 112 & 121 & 122 \\ 211 & 212 & 221 & 222 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 111 & 112 & 211 & 212 \\ 121 & 122 & 221 & 222 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 111 & 121 & 211 & 221 \\ 112 & 122 & 212 & 222 \end{bmatrix}.$$

Для d -мерного тензора $a(i_1, \dots, i_d)$ мы рассматриваем d матриц развертки по измерениям. Пусть $j_k = i_1 \dots i_{k-1} i_{k+1} \dots i_d$ – составной индекс, не включающий i_k . Тогда элементы матрицы развертки a_k по измерению i_k имеют вид

$$a_k[i_k; j_k] = a[i_1, \dots, i_d].$$

Помимо матриц развертки по измерениям, с тензором $a(i_1, \dots, i_d)$ можно связать и другие матрицы. В общем случае для построения матрицы $M = a_{\sigma, k}$, ассоциированной с тензором, нужно взять перестановку $\sigma(1), \dots, \sigma(d)$ натуральных чисел $1, \dots, d$ и нумеровать строки и столбцы матрицы M с помощью составных индексов $i_{\sigma(1)} \dots i_{\sigma(k)}$ и $i_{\sigma(k+1)} \dots i_{\sigma(d)}$. Элементы матрицы M определяются по элементам исходного тензора:

$$M[i_{\sigma(1)} \dots i_{\sigma(k)}; i_{\sigma(k+1)} \dots i_{\sigma(d)}] = a[i_1, \dots, i_d].$$

В случае комплексных тензоров можно использовать норму матрицы M для определения нормы тензора: $\|a\| := \|a_{\sigma, k}\|$. Разные ассоциированные матрицы будут давать, вообще говоря, разные нормы. Однако есть одно исключение: норма Фробениуса одна и та же для любой матрицы, ассоциированной с одним и тем же тензором.

Для матрицы $a_{\sigma, k}$, ассоциированной с тензором $a(i_1, \dots, i_d)$, в случае тождественной перестановки σ будем использовать обозначение $a(i_1 \dots i_k; i_{k+1} \dots i_d)$. По определению,

$$a[i_1 \dots i_k; i_{k+1} \dots i_d] = a[i_1, \dots, i_d].$$

Матрицы $a(i_1 \dots i_k; i_{k+1} \dots i_d)$ называются *стандартными ассоциированными матрицами*. Обратим внимание на то, что для тензора a и его стандартных ассоциированных матриц используется один и тот же символ a .

Среди стандартных ассоциированных матриц выделяется матрица $a(i_1 \dots i_d)$. Она является вектором-столбцом, составленным из упорядоченных лексикографически элементов тензора $a(i_1, \dots, i_d)$. Обобщение $a(i_1, \dots, i_d) \rightarrow a(i_1 \dots i_d)$ взаимно-однозначно и называется *векторизацией* тензора. Иногда используется обозначение

$$a(i_1 \dots i_d) = \text{vec}(a(i_1, \dots, i_d)).$$

Чтобы указать на d -мерный тензор, нередко мы употребляем запись без запятых, т.е. вообще не делаем различия между $a(i_1 \dots i_d)$ и $a(i_1, \dots, i_d)$. Таким образом, d -мерный тензор может рассматриваться как вектор-столбец с d -мерной индексацией элементов с помощью составного индекса.

1.9 Разложение Таккера

Представление тензора $a(i_1, \dots, i_d)$ в виде

$$a(i_1, \dots, i_d) = g(\alpha_1, \dots, \alpha_d) u_1(i_1, \alpha_1) \dots u_d(i_d, \alpha_d) \quad (5)$$

называется *разложением Таккера*. В силу нашего соглашения о суммировании по повторяющимся индексам элементы $a[i_1, \dots, i_d]$ тензора $a(i_1, \dots, i_d)$ выражаются через элементы тензоров $g(\alpha_1, \dots, \alpha_d)$ и двумерных тензоров $u_1(i_1, \alpha_1), \dots, u_d(i_d, \alpha_d)$ следующим образом:

$$a[i_1, \dots, i_d] = \sum_{\alpha_1=1}^{r_1} \dots \sum_{\alpha_d=1}^{r_d} g[\alpha_1, \dots, \alpha_d] u_1[i_1, \alpha_1] \dots u_d[i_d, \alpha_d]. \quad (6)$$

Мощности множеств $[\alpha_1], \dots, [\alpha_d]$, т.е. числа $r_1, \dots, r_d$, называются *рангами разложения Таккера*, их минимальные значения среди всевозможных разложений Таккера для данного тензора называются *рангами Таккера* данного тензора. Разложение Таккера, в котором каждый из рангов является минимально возможным, называется *минимальным разложением Таккера*. Тензор $g(\alpha_1, \dots, \alpha_d)$ называется *ядром разложения Таккера*, а двумерные тензоры

$$u_1(i_1, \alpha_1), \dots, u_d(i_d, \alpha_d)$$

называются *матрицами разложения Таккера*.

Для любого тензора можно написать следующее тривиальное разложение Таккера:

$$a(i_1, \dots, i_d) = a(\alpha_1, \dots, \alpha_d) \delta_1(i_1, \alpha_1) \dots \delta_d(i_d, \alpha_d),$$

где

$$\delta_k[i_k, \alpha_k] = \begin{cases} 1, & i_k = \alpha_k, \\ 0, & i_k \neq \alpha_k. \end{cases}$$

Здесь ядро Таккера совпадает с исходным тензором, и данное разложение, скорее всего, не является минимальным. В минимальном разложении Таккера размеры ядра $g(\alpha_1, \dots, \alpha_d)$ могут оказаться существенно меньше размеров исходного тензора.

Теорема 1.9.1 Для любого тензора $a(i_1, \dots, i_d)$ над произвольным полем $\mathbb{P}$ существует минимальное разложение Таккера, для которого ранги r_k совпадают с рангами Таккера данного тензора и равны рангам матриц развертки a_k по измерениям данного тензора. Матрицы любого минимального разложения Таккера имеют линейно независимые столбцы.

ДОКАЗАТЕЛЬСТВО. Пусть $j_k = i_1 \dots i_{k-1} i_{k+1} \dots i_d$ – составной индекс, не включающий i_k , и пусть ранг матрицы развертки $a_k(i_k; j_k)$ равен r_k . Тогда для матрицы a_k существует скелетное разложение

$$a_k(i_k; j_k) = u_k(i_k; \alpha_k) v_k(\alpha_k; j_k), \quad (7)$$

в котором число скелетонов равно r_k и вследствие этого столбцы матрицы u_k линейно независимы. В силу линейной независимости столбцов для $u_k(i_k, \alpha_k)$ существует матрица $\hat{u}_k(i_k, \alpha'_k)$ тех же размеров, с линейно независимыми столбцами и такая, что $\hat{u}_k^\top u_k = I_{r_k}$ (единичная матрица порядка r_k). Последнее равенство можно записать в виде

$$u_k(i_k, \alpha_k) \hat{u}_k(i_k, \alpha'_k) = \delta_k(\alpha_k, \alpha'_k), \quad (8)$$

где $[\alpha_k] = [\alpha'_k]$ и δ_k определяется как символ Кронекера:

$$\delta_k[\alpha_k, \alpha'_k] = \begin{cases} 1, & \alpha_k = \alpha'_k, \\ 0, & \alpha_k \neq \alpha'_k. \end{cases}$$

Из соотношений (7) и (8) находим

$$a(i_1, \dots, i_{k-1}, i_k, i_{k+1}, \dots, i_d) \hat{u}_k(i_k, \alpha_k) = v_k(\alpha_k, j_k).$$

Следовательно,

$$a(i_1, \dots, i_{k-1}, i_k, i_{k+1}, \dots, i_d) \hat{u}_k(i_k, \alpha_k) u_k(i'_k, \alpha_k) = a(i_1, \dots, i_{k-1}, i'_k, i_{k+1}, \dots, i_d). \quad (9)$$

Определим тензор $g(\alpha_1, \dots, \alpha_d)$ следующим образом:

$$g(\alpha_1, \dots, \alpha_d) = a(i'_1, \dots, i'_d) \hat{u}_1(i'_1, \alpha_1) \dots \hat{u}_d(i'_d, \alpha_d).$$

Последовательно d раз применяя равенство (9), получаем

$$\begin{aligned} & g(\alpha_1, \dots, \alpha_d) u_1(i_1, \alpha_1) \dots u_d(i_d, \alpha_d) = \\ & (a(i'_1, i'_2, \dots, i'_d) \hat{u}_1(i'_1, \alpha_1) u_1(i_1, \alpha_1)) (\hat{u}_2(i'_2, \alpha_2) u_2(i_2, \alpha_2)) \dots (\hat{u}_d(i'_d, \alpha_d) u_d(i_d, \alpha_d)) = \\ & a(i_1, i'_2, \dots, i'_d) (\hat{u}_2(i'_2, \alpha_2) u_2(i_2, \alpha_2)) \dots (\hat{u}_d(i'_d, \alpha_d) u_d(i_d, \alpha_d)) = \dots = a(i_1, \dots, i_d). \end{aligned}$$

Существование минимального разложения Таккера тем самым доказано.

Теперь рассмотрим произвольное минимальное разложение вида (5) и допустим, что столбцы матрицы $u_k(i_k, \alpha_k)$ линейно зависимы. В этом случае каждый из ее столбцов выражается в виде линейной комбинации столбцов некоторой матрицы

$u'_k(i_k, \beta_k)$ с меньшим числом столбцов. Собирая коэффициенты этих линейных комбинаций в столбцы матрицы $z_k(\beta_k, \alpha_k)$, получаем равенство

$$u_k(i_k, \alpha_k) = u'_k(i_k, \beta_k) z_k(\beta_k, \alpha_k).$$

С помощью этого равенства разложение (5) преобразуется следующим образом:

$$a(i_1, \dots, i_d) = g'(\alpha_1, \dots, \alpha_{k-1}, \beta_k, \alpha_{k+1}, \dots, \alpha_d) \cdot u_1(i_1, \alpha_1) \dots u_{k-1}(i_{k-1}, \alpha_{k-1}) u'_k(i_k, \beta_k) u_{k+1}(i_{k+1}, \alpha_{k+1}) \dots u_d(i_d, \alpha_d),$$

где

$$g'(\alpha_1, \dots, \alpha_{k-1}, \beta_k, \alpha_{k+1}, \dots, \alpha_d) = g(\alpha_1, \dots, \alpha_{k-1}, \alpha_k \alpha_{k+1}, \dots, \alpha_d) z_k(\beta_k, \alpha_k).$$

Данное разложение Таккера имеет k -ю матрицу с меньшим числом столбцов, чем в разложении (5), и это, очевидно, противоречит минимальности разложения (5). Теорема доказана.

Теорема 1.9.2 *Для любого тензора в любом минимальном разложении Таккера система столбцов k -й матрицы является базисом линейной оболочки столбцов матрицы развертки данного тензора по k -му измерению. Для любого выбора базисов в этих линейных оболочках существует минимальное разложение Таккера с матрицами, столбцы которых задаются именно этими базисами.*

ДОКАЗАТЕЛЬСТВО. Разложение Таккера (5) автоматически порождает некоторое скелетное разложение k -й матрицы развертки:

$$a_k(i_k, j_k) = u_k(i_k, \alpha_k) v_k(\alpha_k, j_k),$$

где $j_k = i_1 \dots i_{k-1} i_{k+1} \dots i_d$ и $v_k = g u_1 \dots u_{k-1} u_{k+1} \dots u_d$. Минимальность влечет за собой линейную независимость столбцов матрицы u_k и строк матрицы v_k . В силу линейной независимости строк v_k существует матрица w_k такая, что $v_k w_k = I$. Опираясь на равенство $a_k = u_k v_k$, отсюда выводим $u_k = a_k w_k$. Следовательно, $\text{im } u_k \subseteq \text{im } a_k$, т.е. линейная оболочка столбцов матрицы u_k содержит линейную оболочку столбцов матрицы a_k . Из равенства $a_k = u_k v_k$ следует включение $\text{im } a_k \subseteq \text{im } u_k$. Значит, $\text{im } u_k = \text{im } a_k$.

Второе утверждение теоремы вытекает из доказательства теоремы 1.9.1, в котором разложение Таккера строится на основе *любых* минимальных скелетных разложений матриц a_k . Теорема доказана.

1.10 Ортогональное разложение Таккера

Теорема 1.10.1 *Для любого комплексного тензора существует минимальное разложение Таккера, в котором каждая из матриц имеет ортонормированную систему столбцов.*

ДОКАЗАТЕЛЬСТВО. Согласно теореме 1.9.2, минимальное разложение Таккера можно получить, взяв в качестве столбцов матриц разложения любые базисы в линейных оболочках столбцов матриц a_k , т.е. в пространствах $\text{im } a_k$. Как известно, в любом линейном пространстве со скалярным произведением можно выбрать ортонормированный базис. Теорема доказана.

В качестве ортонормированных базисов в пространствах $\text{im } a_k$ можно взять левые сингулярные векторы для упорядоченных по невозрастанию ненулевых сингулярных чисел матриц развертки a_k . Соответствующее минимальное разложение Таккера называется *ортгональным разложением Таккера*.

Пусть матрица развертки a_k имеет сингулярное разложение

$$a_k(i_k; j_k) = u_k(i_k; \alpha_k) \sigma_k(\alpha_k) v_k(\alpha_k; j_k)$$

с числом скелетонов $r_k = \text{rank } a_k$. Как и раньше, $j_k = i_1 \dots i_{k-1} i_{k+1} \dots i_d$. Столбцы u_k суть левые сингулярные векторы, а строки v_k – правые сингулярные векторы матрицы a_k . Как видно из доказательства теоремы 1.9.1, ядро ортогонального разложения Таккера имеет вид

$$g(\alpha_1, \dots, \alpha_d) = a(i_1, \dots, i_d) \bar{u}_1(i_1, \alpha_1) \dots \bar{u}_d(i_d, \alpha_d).$$

Теорема 1.10.2 *Строки k -й матрицы развертки ядра ортогонального разложения Таккера для произвольного комплексного тензора ортогональны и при этом длина i -й по счету строки равна i -му сингулярному числу k -й матрицы развертки исходного тензора.*

ДОКАЗАТЕЛЬСТВО. Построим матрицы u_k до унитарных матриц U_k (любая ортонормированная система столбцов является частью некоторого ортонормированного базиса), а ядро ортогонального разложения Таккера $g(\alpha_1, \dots, \alpha_d)$ вложим в тензор $G(i'_1, \dots, i'_d)$ тех же размеров, что и исходный тензор. Если g имеет размер $r_1 \times \dots \times r_d$ и $1 \leq \alpha_k \leq r_k$, то положим

$$G[i'_1, \dots, i'_d] = \begin{cases} g[i'_1, \dots, i'_d], & 1 \leq i'_1 \leq r_1, \dots, 1 \leq i'_d \leq r_d, \\ 0, & \text{иначе.} \end{cases}$$

Тогда наряду с исходным ортогональным разложением Таккера возникает *расширенное* ортогональное разложение Таккера

$$a(i_1, \dots, i_d) = G(i'_1, \dots, i'_d) U_1(i_1, i'_1) \dots U_d(i_d, i'_d), \quad (10)$$

матрицы которого являются унитарными, а все ненулевые элементы ядра G совпадают с соответствующими ненулевыми элементами ядра g . По аналогии с предыдущим построением, из (10) вытекает

$$G(i'_1, \dots, i'_d) = a(i_1, \dots, i_d) \bar{U}_1(i_1, i'_1) \dots \bar{U}_d(i_d, i'_d). \quad (11)$$

Более того, благодаря унитарности матриц $U_1, \dots, U_d$ равенства (10) и (11) эквивалентны.

Заметим, что скалярные произведения ненулевых строк матрицы развертки G_k тензора G равны скалярным произведениям соответствующих строк матрицы развертки g_k тензора g .

Пусть $w_k(\alpha_k; j_k)$ – матрица, полученная из $v_k(\alpha_k; j_k)$ умножением строк на соответствующие сингулярные числа. Тогда строки матрицы w_k будут ортогональны и при этом длина строки с номером α_k равна $\sigma_k(\alpha_k)$. Матрицы w_k дополним нулевыми строками до матриц W_k – таким образом, чтобы число строк матрицы W_k равнялось порядку унитарной матрицы U_k . Тогда равенство $a_k = u_k w_k$ влечет за собой равенство $a_k = U_k W_k$, откуда $W_k = U_k^* a_k$, или, в соответствии с соглашением о суммировании,

$$W_k(i'_k, i_1 \dots i_{k-1} i_{k+1} \dots i_d) = a(i_1, \dots, i_{k-1}, i_k, i_{k+1}, \dots, i_d) \bar{U}_k(i_k, i'_k). \quad (12)$$

Вводя составные индексы

$$j_k = i_1 \dots i_{k-1} i_{k+1} \dots i_d \quad \text{и} \quad j'_k = i'_1 \dots i'_{k-1} i'_{k+1} \dots i'_d,$$

на основе равенств (11) и (12) находим

$$G_k(i'_k; j'_k) = W_k(i'_k; j_k) Z_k(j_k; j'_k),$$

где

$$Z_k(j_k, j'_k) = \bar{U}_1(i_1, i'_1) \dots \bar{U}_{k-1}(i_{k-1}, i'_{k-1}) \bar{U}_{k+1}(i_{k+1}, i'_{k+1}) \dots \bar{U}_d(i_d, i'_d).$$

Нетрудно проверить, что матрица Z_k является унитарной, поэтому скалярные произведения строк матрицы $W_k Z_k$ совпадают со скалярными произведениями соответствующих строк матрицы W_k . Теорема доказана.

1.11 Кронекерово произведение матриц

При изучении тензоров оказывается полезным *кронекерово произведение* матриц. Для матриц A и B оно определяется как блочная матрица, составленная из блоков $a_{ij}B$, где a_{ij} – элементы матрицы A . Для результата используется обозначение $A \otimes B$.

ПРИМЕР. Если $A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$, то $A \otimes B = \begin{bmatrix} 2B & 3B \\ 4B & 5B \end{bmatrix}$.

Кронекерово произведение может применяться для записи разложения Таккера. При этом исходный тензор рассматривается как вектор-столбец, а разложение Таккера записывается в виде

$$a[i_1 \dots i_d] = \sum_{\alpha_1, \dots, \alpha_d} g[\alpha_1, \dots, \alpha_d] u_{\alpha_1}^1 \otimes \dots \otimes u_{\alpha_d}^d, \quad (13)$$

где

$$u_{\alpha_k}^k = u_k([i_k], \alpha_k).$$

Перечислим некоторые полезные для нас свойства кронекерова произведения:

1. При условии существования произведений AC и BD имеет место равенство

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD). \quad (14)$$

2. При транспонировании матриц

$$(A \otimes B)^\top = A^\top \otimes B^\top. \quad (15)$$

3. Сингулярные числа матрицы $A \otimes B$ равны $\sigma_i(A)\sigma_j(B)$, где $\sigma_i(A)$ и $\sigma_j(B)$ суть сингулярные числа матриц A и B .

4. При кронекеровом умножении матриц их спектральные нормы перемножаются:

$$\|A \otimes B\|_2 = \|A\|_2 \|B\|_2. \quad (16)$$

1.12 Аппроксимации на основе разложения Таккера

Лемма 1.12.1 Для любого комплексного тензора $a(i_1, \dots, i_d)$, заданного разложением Таккера (5) с ядром $g(\alpha_1, \dots, \alpha_d)$ и матрицами $u_1(i_1, \alpha_1), \dots, u_d(i_d, \alpha_d)$, имеет место неравенство

$$\|a\|_F \leq \|g\|_F \|u_1\|_2 \dots \|u_d\|_2.$$

В случае ортогонального разложения Таккера

$$\|a\|_F = \|g\|_F.$$

ДОКАЗАТЕЛЬСТВО. Фробениусова норма тензора совпадает с нормой Фробениуса для любой ассоциированной с ним матрицы, в частности матрицы развертки по первому измерению: $\|a\|_F = \|a_1\|_F$. Запишем a_1 в виде произведения матриц

$$a(i_1; i_2 \dots i_d) = (u(i_1; \alpha_1)(g_1(\alpha_1; \alpha_2 \dots \alpha_d)) v(\alpha_2 \dots \alpha_d; i_2 \dots i_d),$$

$$v(\alpha_2 \dots \alpha_d; i_2 \dots i_d) = u_2(i_2, \alpha_2) \dots u_d(i_d, \alpha_d),$$

и заметим связь с операцией кронекерова умножения матриц:

$$v^\top = u_2 \otimes \dots \otimes u_d.$$

Далее, $\|u_1 g_1 v\|_F \leq \|u_1\|_2 \|g_1\|_F \|v\|_2$ и, кроме того, согласно (12), спектральная норма кронекерова произведения матриц равна произведению спектральных норм сомножителей.

Если дано ортогональное разложение Таккера для тензора размера $n_1 \times \dots \times n_d$, то перейдем к порожденному им расширенному разложению (см. доказательство теоремы 1.10.2))

$$a(i_1, \dots, i_d) = G(i'_1, \dots, i'_d) U_1(i_1, i'_1) \dots U_d(i_d, i'_d),$$

где $U_1, \dots, U_d$ – унитарные матрицы соответственно порядка $n_1, \dots, n_d$. Тогда

$$\|g\|_F = \|G\|_F,$$

а матрица развертки a_1 представляется произведением

$$a_1 = U_1 G_1 V,$$

причем

$$V^\top = U_2 \otimes \dots \otimes U_d.$$

Согласно (14) и (15) находим

$$\begin{aligned} (U_2 \otimes \dots \otimes U_d)(U_2 \otimes \dots \otimes U_d)^* &= (U_2 \otimes \dots \otimes U_d)(U_2^* \otimes \dots \otimes U_d^*) = \\ &= (U_2 U_2^*) \otimes \dots \otimes (U_d U_d^*) = I_{n_2} \otimes \dots \otimes I_{n_d} = I. \end{aligned}$$

Поэтому $V^\top$, а значит и V является унитарной матрицей. Отсюда

$$G = U_1^* a_1 V^* \quad \text{и} \quad \|G\|_F \leq \|U_1\|_2 \|a_1\|_F \|V\|_2.$$

В силу унитарности $\|U_1\|_2 = \|V\|_2 = 1$. Теорема доказана.

Следствие 1.12.1 Пусть комплексные тензоры $a(i_1, \dots, i_d)$ и $\tilde{a}(i_1, \dots, i_d)$ заданы разложениями Таккера с одними и теми же матрицами $u_1, \dots, u_d$, но разными ядрами g и $\tilde{g}$. Тогда имеет место неравенство

$$\|a - \tilde{a}\|_F \leq \|g - \tilde{g}\|_F \|u_1\|_2 \dots \|u_d\|_2.$$

В случае ортогональных разложений Таккера

$$\|a - \tilde{a}\|_F = \|g - \tilde{g}\|_F.$$

Предположим, что матрица развертки a_k по k -му измерению тензора a имеет сингулярные числа

$$\sigma_k(1) \geq \sigma_k(2) \geq \dots \geq \sigma_k(r_k) > 0.$$

Будем также считать, что $\sigma_k(l) = 0$ при $l > r_k$. Пусть для этого тензора имеется ортогональное разложение Таккера с рангами $r_1, \dots, r_d$:

$$a[i_1, \dots, i_d] = \sum_{\alpha_1=1}^{r_1} \dots \sum_{\alpha_d=1}^{r_d} g[\alpha_1, \dots, \alpha_d] u_1[i_1, \alpha_1] \dots u_d[i_d, \alpha_d].$$

На основе этого разложения можно получать аппроксимации данного тензора в виде ортогональных разложений Таккера с меньшими рангами.

Фиксируем натуральные числа $1 \leq s_k \leq r_k$ и рассмотрим *редуцированное* разложение Таккера

$$\tilde{a}[i_1, \dots, i_d] = \sum_{\beta_1=1}^{s_1} \dots \sum_{\beta_d=1}^{s_d} g[\beta_1, \dots, \beta_d] u_1[i_1, \beta_1] \dots u_d[i_d, \beta_d].$$

Теорема 1.12.1

$$\|a - \tilde{a}\|_F \leq \sqrt{\sum_{\alpha_k > s_k, 1 \leq k \leq d} \sigma_k^2(\alpha_k)}.$$

ДОКАЗАТЕЛЬСТВО. Редуцированное ортогональное разложение Таккера с рангами $s_1, \dots, s_d$ можно *расширить* до разложения Таккера с рангами $r_1, \dots, r_d$ и матрицами $u_1, \dots, u_d$ в точности такими же, как в исходном разложении Таккера. Для этого ядро расширенного разложения надо определить следующим образом:

$$\tilde{g}[\alpha_1, \dots, \alpha_d] = \begin{cases} g[\alpha_1, \dots, \alpha_d], & 1 \leq \alpha_1 \leq s_1, \dots, 1 \leq \alpha_d \leq s_d, \\ 0, & \text{иначе.} \end{cases}$$

Согласно следствию 1.12.1, $\|a - \hat{a}\|_F = \|g - \tilde{g}\|_2$. Далее замечаем, что ненулевые строки матриц развертки для тензора $g - \tilde{g}$ имеют такие же ненулевые элементы, как и соответствующие строки матриц развертки для тензора g , и применяем утверждение теоремы 1.10.2 о длинах строк матриц развертки ядра ортогонального разложения Таккера. Теорема доказана.

1.13 Вычисление разложения Таккера

Изучение разложения Таккера было конструктивным и опиралось на построение скелетных разложений для матриц развертки заданного тензора по каждому из его измерений. Таким образом, мы уже имеем некоторый алгоритм вычисления минимального разложения Таккера.

Алгоритм 1.13.1 Пусть задан тензор $a(i_1, \dots, i_d)$ размера $n_1 \times \dots \times n_d$. Тогда:

1. Для всех $k = 1, \dots, d$ для k -й матрицы развертки a_k тензора a найти скелетное разложение $a_k = u_k v_k$, в котором матрица u_k имеет $r_k = \text{rank } a_k$ линейно независимых столбцов.
2. Для каждой матрицы u_k найти матрицу $\hat{u}_k$ такую, что $\hat{u}_k^\top u_k = I_{r_k}$.
3. Найти ядро Таккера по формуле

$$g[\alpha_1, \dots, \alpha_d] = \sum_{i_1=1}^{n_1} \dots \sum_{i_d=1}^{n_d} a[i_1, \dots, i_d] \hat{u}_1[i_1, \alpha_1] \dots \hat{u}_d[i_d, \alpha_d].$$

Данный алгоритм симметричен относительно измерений тензора, но именно по этой причине может требовать избыточной вычислительной работы. Более экономичным (и более естественным, как мне кажется) является алгоритм, который приводится ниже.

Алгоритм 1.13.2 Для заданного тензора $a(i_1, \dots, i_d)$ выполнить следующие действия:

1. Положить $v_0(i_1, \dots, i_d) = a(i_1, \dots, i_d)$ и найти минимальное скелетное разложение первой матрицы развертки

$$v_0(i_1, i_2, \dots, i_d) = u_1(i_1, \alpha_1) v_1(\alpha_1, i_2, \dots, i_d).$$

2. Для $k := 2, \dots, d$ найти минимальное скелетное разложение k -й матрицы развертки V_k для уже вычисленного тензора v_{k-1} :

$$V_k(i_k; \alpha_1, \dots, \alpha_{k-1}, i_{k+1}, \dots, i_d) = \\ v_{k-1}(\alpha_1, \dots, \alpha_{k-1}, i_k, \dots, i_d) = u_k(i_k, \alpha_k) v_k(\alpha_1, \dots, \alpha_{k-1}, \alpha_k, i_{k+1}, \dots, i_d).$$

3. Искомое разложение Таккера определяется матрицами $u_1(i_1, \alpha_1), \dots, u_d(i_d, \alpha_d)$ и ядром $g(\alpha_1, \dots, \alpha_d) = v_d(\alpha_1, \dots, \alpha_d)$.

Теорема 1.13.1 Алгоритм 1.13.2 вычисляет минимальное разложение Таккера.

ДОКАЗАТЕЛЬСТВО. Пусть ранги вычисленного по алгоритму 1.13.2 разложения Таккера равны $r_1, \dots, r_d$. Нам нужно установить, что $r_k = \text{rank } a_k$, где a_k есть k -я матрица развертки заданного тензора. Минимальность скелетного разложения матрицы означает, что число скелетонов равно рангу матрицы. Поэтому очевидно, что $r_1 = \text{rank } a_1$. Аналогично, $r_k = \text{rank } V_k$. Для каждого k выполняется равенство $v_{k-1} = u_k v_k$, в котором матрица (двумерный тензор) u_k имеет линейно независимые столбцы, а тензоры v_{k-1} и v_k имеют, вообще говоря, разные размеры, но одинаковое число измерений. Вследствие линейной независимости столбцов u_k ранги матриц развертки для тензоров v_{k-1} и v_k по соответствующим измерениям совпадают. То же справедливо для тензоров v_k и a . Теорема доказана.

Преимущество алгоритма 1.13.2 заключается в том, что размеры тензора, с которым проводятся вычисления, уменьшаются от шага к шагу. Если исходный тензор имеет размер $n_1 \times \dots \times n_d$ и ранги разложения Таккера образуют последовательность $r_1, \dots, r_d$, то на k -м шаге мы имеем дело с тензором размера $r_1 \times \dots \times r_{k-1} \times n_k \times \dots \times n_d$, скелетное разложение вычисляется для его k -й матрицы развертки, которая имеет размер $n_k \times (r_1 \dots r_{k-1} n_{k+1} \dots n_d)$. Заметим, что разложение Таккера для тензора $a(i_1, \dots, i_d)$ легко получается из разложения Таккера для тензора с переставленными измерениями

$$a_\sigma(i_{\sigma(1)}, \dots, i_{\sigma(d)}) = a(i_1, \dots, i_d).$$

При этом объем вычислений может существенно зависеть от выбранной перестановки измерений.

Заметим, что с точки зрения параллельных вычислений алгоритм 1.13.2 может показаться менее привлекательным, чем алгоритм 1.13.1. На конкретной вычислительной системе, вероятно, следует использовать некоторую их комбинацию.

1.14 Вычисление ортогонального разложения Таккера

Алгоритм 1.13.1 легко адаптируется для вычисления ортогонального разложения Таккера. При получении скелетных разложений $a_k = u_k v_k$ следует в качестве u_k брать матрицу левых сингулярных векторов матрицы a_k . В этом случае строки матрицы v_k будут ортогональны, а их длины будут равны сингулярным числам матрицы a_k . Таким образом, для каждой матрицы развертки a_k потребуется вычислить ее сингулярное разложение.

Размеры матриц, для которых требуется находить сингулярное разложение, можно уменьшить – иногда существенно, если уже известно разложение Таккера

$$a(i_1, \dots, i_d) = g(\alpha_1, \dots, \alpha_d) u_1(i_1, \alpha_1) \dots u_d(i_d, \alpha_d),$$

в котором размеры ядра меньше размеров исходного тензора. Проведя ортогонализацию столбцов матриц u_k (по методу Грама–Шмидта или с помощью QR -разложения), получаем разложения

$$u_k(i_k, \alpha_k) = q_k(i_k, \beta_k) z_k(\beta_k, \alpha_k),$$

где матрицы q_k имеют ортонормированные столбцы, а затем и новое разложение Таккера

$$a(i_1, \dots, i_d) = h(\beta_1, \dots, \beta_d) q_1(i_1, \beta_1) \dots q_d(i_d, \beta_d) \quad (17)$$

с ортонормированными столбцами матриц разложения и ядром

$$h(\beta_1, \dots, \beta_d) = g(\alpha_1, \dots, \alpha_d) z_1(\beta_1, \alpha_1) \dots z_d(\beta_d, \alpha_d).$$

Ортонормированность столбцов позволяет свести вычисление ортогонального разложения Таккера для тензора $a(i_1, \dots, i_d)$ к вычислению ортогонального разложения Таккера для ядра $h(\beta_1, \dots, \beta_d)$.

Лемма 1.14.1 *Если тензоры $a(i_1, \dots, i_d)$ и $h(\beta_1, \dots, \beta_d)$ связаны соотношением (17) с ортонормированными столбцами в матрицах $q_1(i_1, \beta_1), \dots, q_d(i_d, \beta_d)$, то ненулевые сингулярные числа для k -й матрицы развертки тензора $a(i_1, \dots, i_d)$ с учетом кратностей совпадают с ненулевыми сингулярными числами для k -й матрицы развертки тензора $h(\beta_1, \dots, \beta_d)$.*

ДОКАЗАТЕЛЬСТВО. Пусть тензоры $a(i_1, \dots, i_d)$ и $h(\beta_1, \dots, \beta_d)$ имеют соответственно размеры

$$n_1 \times \dots \times n_d \quad \text{и} \quad r_1 \times \dots \times r_d.$$

Перейдем от (17) к расширенному разложению

$$a(i_1, \dots, i_d) = H(i'_1, \dots, i'_d) Q_1(i_1, i'_1) \dots Q_d(i_d, i'_d),$$

в котором столбцы матриц $Q_1, \dots, Q_d$ образуют ортонормированные базисы, включающие столбцы матриц $q_1, \dots, q_d$. Таким образом, матрицы $Q_1, \dots, Q_d$ являются унитарными матрицами порядка $n_1, \dots, n_d$, соответственно. Тензор H является окаймлением тензора h :

$$H[i'_1, \dots, i'_d] = \begin{cases} h[i'_1, \dots, i'_d], & 1 \leq i'_1 \leq r_1, \dots, 1 \leq i'_d \leq r_d, \\ 0, & \text{иначе.} \end{cases}$$

Ненулевые сингулярные числа k -й матрицы развертки для H с учетом кратностей очевидно совпадают с ненулевыми сингулярными числами k -й матрицы развертки для h . Остается заметить, что k -я матрица развертки для тензора a получается из k -й матрицы развертки для H умножением слева и справа на унитарные матрицы. Лемма доказана.

Теорема 1.14.1 *Предположим, что тензор $a(i_1, \dots, i_d)$ обладает разложением Таккера вида (17) с ортонормированными столбцами в матрицах разложения и ядром, для которого получено ортогональное разложение Таккера вида*

$$h(\beta_1, \dots, \beta_d) = f(\gamma_1, \dots, \gamma_d) p_1(\beta_1, \gamma_1) \dots p_d(\beta_d, \gamma_d). \quad (18)$$

Тогда ортогональное разложение Таккера для тензора $a(i_1, \dots, i_d)$ имеет вид

$$a(i_1, \dots, i_d) = f(\gamma_1, \dots, \gamma_d) w_1(i_1, \gamma_1) \dots w_d(i_d, \gamma_d), \quad (19)$$

$$w_k(i_k, \gamma_k) = q_k(i_k, \beta_k) p_k(\beta_k, \gamma_k).$$

ДОКАЗАТЕЛЬСТВО. Строки каждой из матриц развертки для f ортогональны, так как f является ядром некоторого ортогонального разложения Таккера. Длины строк равны сингулярным числам соответствующих матриц развертки для h . Согласно лемме 1.14.1, соответствующие матрицы развертки для a и h имеют один и тот же набор ненулевых сингулярных чисел. Ортонормированность столбцов матриц w_k вытекает из ортонормированности столбцов q_k и p_k . Ортогональность строк матрицы $w_k^* a_k$ означает, что столбцы матрицы w_k являются левыми сингулярными векторами k -й матрицы развертки a_k . Теорема доказана.

1.15 Вычисление аппроксимаций

Мы уже знаем, что аппроксимации легко строятся с помощью редукции ортогонального разложения Таккера заданного тензора. Оценка погрешности дается теоремой 1.12.1. Другой метод построения аппроксимаций гарантированной точности (по-видимому, более экономичный) получается на основе алгоритма 1.13.2.

Алгоритм 1.15.1 *Пусть задан тензор $a(i_1, \dots, i_d)$ и предписаны оценки погрешности $\varepsilon_1, \dots, \varepsilon_d$ аппроксимации матриц развертки $a_1, \dots, a_d$. Тогда:*

1. *Положим $v_0(i_1, \dots, i_d) = a(i_1, \dots, i_d)$, найдем сингулярное разложение первой матрицы развертки и запишем его в виде скелетного разложения*

$$v_0(i_1, i_2, \dots, i_d) = u_1(i_1, \alpha_1) v_1(\alpha_1, i_2, \dots, i_d),$$

в котором столбцы матрицы u_1 составлены из левых сингулярных векторов, а строки матрицы v_1 из правых сингулярных векторов, умноженных на соответствующие сингулярные числа в порядке невозрастания.

Пусть $r_1 = \text{rank } v_0$. Введем индексы α'_1 и α''_1 , изменяющиеся на отрезках

$$1 \leq \alpha'_1 \leq s_1, \quad s_1 + 1 \leq \alpha''_1 \leq r_1,$$

и выберем s_1 как наименьший номер, для которого

$$\|v_1(\alpha''_1; i_2 \dots i_d)\|_F \leq \varepsilon_1.$$

2. Для $k := 2, \dots, d$ найдем сингулярное разложение k -й матрицы развертки V_k для уже вычисленного тензора $v_{k-1}(\alpha'_1, \dots, \alpha'_{k-1}, i_k, \dots, i_d)$ и по аналогии с первым шагом запишем его в виде скелетного разложения:

$$V_k(i_k; \alpha'_1, \dots, \alpha'_{k-1}, i_{k+1}, \dots, i_d) = \\ v_{k-1}(\alpha'_1, \dots, \alpha'_{k-1}, i_k, \dots, i_d) = u_k(i_k, \alpha_k) v_k(\alpha'_1, \dots, \alpha'_{k-1}, \alpha_k, i_{k+1}, \dots, i_d).$$

Пусть $r_k = \text{rank } v_{k-1}$. Введем индексы α'_k и α''_k , изменяющиеся на отрезках

$$1 \leq \alpha'_k \leq s_k, \quad s_k + 1 \leq \alpha''_k \leq r_k,$$

и выберем s_k как наименьший номер, для которого

$$\|v_k(\alpha''_k; i_2 \dots i_d)\|_F \leq \varepsilon_k.$$

3. Аппроксимация в виде разложения Таккера с рангами $s_1, \dots, s_d$ определяется матрицами $u_1(i_1, \alpha'_1), \dots, u_d(i_d, \alpha'_d)$ и ядром $g(\alpha'_1, \dots, \alpha'_d) = v_d(\alpha'_1, \dots, \alpha'_d)$.

На k -м шаге данного алгоритма строится аппроксимация k -й матрицы развертки тензора v_{k-1} матрицей ранга не выше s_k . Заметим, что эта аппроксимация порождает аналогичную аппроксимацию матрицы a_k , причем в факторизованном виде:

$$a_k = u_k(i_k, \alpha'_k) (u_1(i_1, \alpha'_1) \dots u_{k-1}(i_{k-1}, \alpha'_{k-1}) v_k(\alpha'_1, \dots, \alpha'_{k-1}, \alpha'_k, i_{k+1}, \dots, i_d)).$$

На последующих шагах факторизуется тензор v_k . По завершении алгоритма та же матрица a_k получает аппроксимацию ранга не выше s_k в еще более факторизованном виде – следует подставить построенное в итоге разложение для тензора v_k .

Теорема 1.15.1 Алгоритм 1.15.1 вычисляет разложение Таккера, которое определяет тензор

$$\tilde{a}(i_1, \dots, i_d) = g(\alpha'_1, \dots, \alpha'_d) u_1(i_1, \alpha'_1) \dots u_d(i_d, \alpha'_d),$$

приближающий тензор $a(i_1, \dots, i_d)$ с погрешностью

$$\|a - \tilde{a}\|_F \leq \sqrt{\sum_{k=1}^d \varepsilon_k^2}.$$

ДОКАЗАТЕЛЬСТВО. На первом шаге получается расщепление $a = b + c$, где

$$b = u_1(i_1, \alpha'_1) v_1(\alpha'_1, i_2 \dots i_d), \quad c = u_1(i_1, \alpha''_1) v_1(\alpha''_1, i_2 \dots i_d),$$

$$(b, c) = 0 \quad \text{и} \quad \|c\|_F \leq \varepsilon_1.$$

Нетрудно проверить, что любой тензор вида

$$\tilde{b} = u_1(i_1, \alpha'_1) \tilde{v}_1(\alpha'_1, i_2 \dots i_d)$$

будет ортогональным тензору c . Поэтому из неравенства

$$\|b - \tilde{b}\|_F \leq \varepsilon$$

должно вытекать

$$\|a - \tilde{b}\|_F \leq \sqrt{\|a - b\|_F^2 + \|b - \tilde{b}\|_F^2} \leq \sqrt{\varepsilon_1^2 + \varepsilon^2}.$$

Теорема доказана.

Теорема 1.15.2 Пусть $E(s_1, \dots, s_d)$ обозначает погрешность наилучшего приближения

$$E(s_1, \dots, s_d) = \min_b \|a - b\|_F$$

на множестве тензоров, имеющих ранги Таккера не выше $s_1, \dots, s_d$. При выполнении алгоритма 1.15.1 с фиксированными значениями s_k вычисляется тензор $\tilde{a}(i_1, \dots, i_d)$, приближающий $a(i_1, \dots, i_d)$ с погрешностью

$$\|a - \tilde{a}\|_F \leq \sqrt{d} E(s_1, \dots, s_d). \quad (20)$$

ДОКАЗАТЕЛЬСТВО. Прежде всего, минимум существует – в силу того, что его вычисление сводится к минимизации непрерывной функции на компактном множестве. Далее заметим (оставляем читателю строгое объяснение, почему это именно так), что при аппроксимации k -й матрицы развертки матрицей ранга не выше s_k фробениусова норма погрешности наилучшего приближения не может быть больше $E(s_1, \dots, s_d)$. Поэтому в качестве ε_k можно взять $\varepsilon_k = E(s_1, \dots, s_d)$. Теорема доказана.

Теорема 1.15.3 (Делатауэр) Оценка (20) справедлива также для тензора $\tilde{a}$, полученного редукицией ортогонального разложения Таккера к разложению с рангами не выше $s_1, \dots, s_d$.

ДОКАЗАТЕЛЬСТВО. Согласно теореме 1.12.1, погрешность редуцированного разложения Таккера оценивается корнем квадратным из суммы квадратов погрешностей наилучшего приближения матриц развертки матрицами рангов $s_1, \dots, s_d$. Как уже отмечалось, эти погрешности не превосходят $E(s_1, \dots, s_d)$. Теорема доказана.

2 Каноническое трилинейное разложение

2.1 Определение и связь с разложением Таккера

Прямым обобщением скелетного разложения матриц на общий случай d -тензора является разложение вида

$$a(i_1, \dots, i_d) = u_1(i_1, \alpha) \dots u_d(i_d, \alpha).$$

В силу нашего соглашения о суммировании по повторяющимся индексам, каждый элемент тензора a записывается в виде суммы

$$a[i_1, \dots, i_d] = \sum_{\alpha=1}^R u_1[i_1, \alpha] \dots u_d[i_d, \alpha].$$

Тензор, обладающий каноническим разложением с числом членов $R = 1$, называется *скелетоном*. Таким образом, каноническое разложение тензора – это его запись в виде суммы скелетов. Число скелетов называется *рангом канонического разложения*. Минимально возможное число скелетов R называется *каноническим тензорным рангом* или просто *тензорным рангом* данного тензора, соответствующее каноническое разложение тензора называется *минимальным*. Для тензора a над полем $\mathbb{P}$ его тензорный ранг над $\mathbb{P}$ будем обозначать $\text{trank}_{\mathbb{P}}(a)$.

В случае трех измерений каноническое разложение часто называют *трилинейным разложением* тензора. Несмотря на то, что каноническое трилинейное разложение естественно воспринимается как обобщение скелетного разложения матриц, свойства соответствующих минимальных разложений отличаются коренным образом. В случае двух измерений, т.е. для матриц, ранг находится за конечное число арифметических операций, а для трех измерений мы уже не знаем какого-либо алгоритма, который мог бы с гарантией вычислить ранг произвольно выбранного вещественного или комплексного тензора. Для тензора над конечным полем задача очевидно решается перебором, но является комбинаторно сложной. Таким образом, случаи $d = 2$ и $d = 3$ *качественно разные*. Дальнейшее увеличение числа измерений уже не приводит к новому качеству.

Каноническое разложение можно рассматривать как специальный случай разложения Таккера

$$a(i_1, \dots, i_d) = g(\alpha_1, \dots, \alpha_d) u_1(i_1, \alpha_1) \dots u_d(i_d, \alpha_d)$$

с ядром $g(\alpha_1, \dots, \alpha_d)$ размера $R \times \dots \times R$, имеющим нулевые элементы всюду, кроме диагонали

$$\alpha_1 = \dots = \alpha_d.$$

Связь канонического разложения с разложением Таккера особенно наглядна в случае матриц: для матрицы A каноническое разложение совпадает с ее скелетным разложением

$$A = UV^{\top} = u_1 v_1^{\top} + \dots + u_R v_R^{\top},$$

а разложение Таккера имеет вид

$$A = UGV^{\top},$$

где G – матрица порядка R .

2.2 Существование и оценка сверху

Заметим, что любой тензор с одним единственным отличным от нуля элементом является скелетоном: пусть этот элемент находится в позиции $i_1 = i_1^0, \dots, i_d = i_d^0$ и

равен 1, тогда

$$a[i_1, \dots, i_d] = \delta(i_1, i_1^0) \dots \delta(i_d, i_d^0),$$

где δ – символ Кронекера ($\delta = 1$ при совпадении аргументов, иначе $\delta = 0$).

Отсюда сразу же вытекает, что любой тензор представим в виде суммы скелетонных, причем легко выписывается каноническое разложение, в котором число скелетонных равно числу ненулевых элементов тензора. Как правило, такое разложение не будет минимальным.

Теорема 2.2.1 *Для числа скелетонных R в минимальном каноническом разложении d -тензора размера $n_1 \times \dots \times n_d$ над произвольным полем справедлива оценка*

$$R \leq \min(N/n_1, \dots, N/n_d), \quad \text{где } N = n_1 \dots n_d.$$

ДОКАЗАТЕЛЬСТВО. Случай $d = 2$ очевиден, а при $d \geq 3$ применяем индукцию. Достаточно заметить, что d -тензор $a(i_1, \dots, i_{d-2}, i_{d-1}, i_d)$ размера

$$n_1 \times \dots \times n_{d-2} \times n_{d-1} \times n_d$$

можно рассматривать также как $(d-1)$ -тензор $a(i_1, \dots, i_{d-2}, i_{d-1}i_d)$ размера

$$n_1 \times \dots \times n_{d-2} \times (n_{d-1}n_d).$$

Запишем каноническое разложение

$$a(i_1, \dots, i_{d-2}, i_{d-1}i_d) = u_1(i_1, \alpha) \dots u_{d-2}(i_{d-2}, \alpha) u_{d-1}(i_{d-1}i_d, \alpha),$$

в котором число членов не больше $n_1 \dots n_{d-2}$. Далее, при каждом α имеем каноническое разложение двумерного тензора $u_{d-1}([i_{d-1}, i_d], \alpha)$ с числом слагаемых не больше чем $\min(n_{d-1}, n_d)$. Значит, исходный тензор записывается в виде суммы не более чем

$$n_1 \dots n_{d-2} \min(n_{d-1}, n_d) = \min(N/n_{d-1}, N/n_d)$$

скелетонных. Используя перестановку измерений, можно полагать, что

$$n_d = \max(n_1, \dots, n_d).$$

Теорема доказана.

Введем обозначение

$$\text{mrang}(n_1, \dots, n_d) = \text{mrang}_{\mathbb{P}}(n_1, \dots, n_d)$$

для максимально возможного значения тензорного ранга для всевозможных тензоров размера $n_1 \times \dots \times n_d$ над полем $\mathbb{P}$. Ниже мы увидим, в частности, что для полей $\mathbb{P} = \mathbb{C}$ и $\mathbb{P} = \mathbb{R}$ имеет место равенство

$$\text{mrang}(2, 2, 2) = 3.$$

Данный результат согласуется с общей нетривиальной оценкой максимального ранга кубических тензоров [10]

$$\text{mrang}_{\mathbb{C}}(n, n, n) \leq \frac{n(n+1)}{2}.$$

Таким образом, оценка теоремы 2.2.1 не является точной, по крайней мере для комплексных и вещественных тензоров размера $2 \times 2 \times 2$ и для комплексных тензоров размера $n \times n \times n$.

Заметим также, что совсем не очевидно, должен ли максимальный ранг на всем классе тензоров одного и того же размера быть одним и тем же для комплексного и вещественного поля. По крайней мере, скоро мы приведем пример конкретного вещественного тензора, для которого вещественный ранг *больше* комплексного ранга.

Задача. Пусть a – произвольный 3-тензор над полем $\mathbb{P}$, и пусть его ранги Таккера R_1, R_2, R_3 упорядочены таким образом, что $R_1 \leq R_2 \leq R_3$. Тогда

$$R_3 \leq \text{trank}_{\mathbb{P}}(a) \leq R_1 R_2.$$

2.3 Инвариантность относительно умножений на матрицы

Теорема 2.3.1 Пусть при умножении тензора $a(i_1, \dots, i_d)$ на матрицы $u_1(i_1, i'_1), \dots, u_d(i_d, i'_d)$ получается тензор

$$b(i'_1, \dots, i'_d) = a(i_1, \dots, i_d) u_1(i_1, i'_1) \dots u_d(i_d, i'_d).$$

Тогда для любого канонического разложения тензора $a(i_1, \dots, i_d)$ можно построить каноническое разложение тензора $b(i_1, \dots, i_d)$, в котором число скелетов не больше, чем в каноническом разложении тензора $a(i_1, \dots, i_d)$.

Следствие 2.3.1 Канонический тензорный ранг не меняется при умножении тензора на произвольные невырожденные матрицы.

Доказательства элементарны и предоставляются читателю.

В случае трех измерений тензор $a(i, j, k)$ размера $m \times n \times q$ можно задавать последовательностью его сечений

$$A_1 = a([i, j], 1), \dots, A_q = a([i, j], q).$$

Каждое сечение представляет собой матрицу размера $m \times n$, составленную из элементов одной строки матрицы развертки данного тензора по третьему измерению. В силу инвариантности при вычислении тензорного ранга можно выполнять следующие преобразования сечений:

1. Фиксировать номер строки и в каждом сечении к данной строке прибавить одну и ту же линейную комбинацию остальных строк.
2. Фиксировать номер столбца и в каждом сечении к данному столбцу прибавить одну и ту же линейную комбинацию остальных столбцов.

3. Выбрать сечение и прибавить к нему линейную комбинацию остальных сечений.

Из следствия 2.3.1 ясно, что при каждом преобразовании из перечисленных выше тензорный ранг сохраняется. Понятно также, что тензорный ранг не меняется при перестановке фиксированных строк или столбцов одновременно во всех сечениях или при произвольной перестановке сечений.

2.4 Интерпретации трилинейного разложения

Рассмотрим трилинейное разложение

$$a[i, j, k] = \sum_{\alpha=1}^R u[i, \alpha] v[j, \alpha] w[k, \alpha] \quad (21)$$

и следующие две возможности его интерпретации.

ИНТЕРПРЕТАЦИЯ 1. Равенство (21) означает, что сечение $A_k = a([i, j], k)$ есть линейная комбинация $m \times n$ -матриц B_α с элементами

$$(B_\alpha)_{ij} = u[i, \alpha] v[j, \alpha].$$

Таким образом, справедливо следующее утверждение.

Лемма 2.4.1 *Для тензора $a(i, j, k)$ размера $m \times n \times q$ существует трилинейное разложение ранга R тогда и только тогда, когда существуют матрицы $B_1, \dots, B_R$ размера $m \times n$, ранга 1 и такие, что каждое сечение A_k принадлежит их линейной оболочке:*

$$A_1, \dots, A_q \in \langle B_1, \dots, B_R \rangle.$$

При вычислении тензорного ранга нужно найти минимально возможное число матриц ранга 1, в линейной оболочке которых содержатся все сечения.

ИНТЕРПРЕТАЦИЯ 2. То же самое равенство (21) означает, что сечение A_k есть произведение трех матриц, а именно:

$$A_k = U D_k V, \quad 1 \leq k \leq q, \quad (22)$$

где $U = [u(i, \alpha)]$ и $V = [v(j, \alpha)]^\top$ – матрицы размеров $m \times R$ и $R \times n$ соответственно, одинаковые для всех k , а D_k – диагональная матрица с элементами

$$(D_k)_{\alpha\alpha} = w[k, \alpha], \quad 1 \leq \alpha \leq R.$$

Лемма 2.4.2 *Для тензора $a(i, j, k)$ размера $m \times n \times q$ существует трилинейное разложение ранга R тогда и только тогда, когда его сечения A_k имеют вид (22) для каких-то матриц U и V размеров $m \times R$ и $R \times n$ соответственно и диагональных матриц D_k размера $R \times R$.*

Недостающие детали доказательств очевидны и предоставляются читателю.

Задача 1. Докажите, что для тензоров над произвольным полем $\text{mrang}(m, n, q) = mn$, если $q \geq mn$.

Задача 2. Пусть тензор размера $n \times n \times 3$ определяется тремя $n \times n$ -матрицами его сечений A_1, A_2, A_3 . Докажите, что если хотя бы две из этих матриц являются диагональными, то канонический ранг данного тензора не выше $2n$. (В работе [10] утверждается, что этот результат сохраняет силу для комплексных тензоров размера $n \times n \times 4$, а в случае q сечений ранг не выше $\lceil q/2 \rceil n$.)

2.5 Ранги тензоров размера $2 \times 2 \times 2$

Тензор размера $2 \times 2 \times 2$ определяется двумя сечениями, т.е. двумя 2×2 -матрицами A_1 и A_2 .

Согласно лемме 2.4.2, для тензора $a(i, j, k) \in \mathbb{P}^{2 \times 2 \times 2}$ трилинейное разложение ранга 2 существует в том и только том случае, когда существуют матрицы $U, V \in \mathbb{P}^{2 \times 2}$ и диагональные матрицы $D_1, D_2 \in \mathbb{P}^{2 \times 2}$ такие, что

$$A_1 = U D_1 V, \quad A_2 = U D_2 V.$$

Предположим, что матрица A_2 невырожденная. Тогда

$$A_1 A_2^{-1} = U D U^{-1}, \quad D = D_1 D_2^{-1},$$

т.е. матрица $A_1 A_2^{-1}$ подобна диагональной матрице. Отсюда вытекает утверждение: *если матрица $A_1 A_2^{-1}$ недиагонализуема над полем $\mathbb{P}$, то тензорный ранг данного тензора над $\mathbb{P}$ не меньше 3.*

Для комплексного и вещественного поля в таких случаях канонический тензорный ранг в точности равен 3.

Теорема 2.5.1 $\text{mrang}_{\mathbb{C}}(2, 2, 2) = \text{mrang}_{\mathbb{R}}(2, 2, 2) = 3$.

ДОКАЗАТЕЛЬСТВО. Пусть тензор определяется сечениями A_1 и A_2 . Если обе матрицы A_1 и A_2 вырожденные, то в силу леммы 2.4.1 тензорный ранг не больше 2. Трилинейное разложение ранга 2 легко выписывается явно. В самом деле, пусть

$$A_1 = \begin{bmatrix} u(1, 1) \\ u(2, 1) \end{bmatrix} \begin{bmatrix} v(1, 1) & v(2, 1) \end{bmatrix}, \quad A_2 = \begin{bmatrix} u(1, 2) \\ u(2, 2) \end{bmatrix} \begin{bmatrix} v(1, 2) & v(2, 2) \end{bmatrix}.$$

Тогда при выборе

$$w(1, 1) = 1, \quad w(1, 2) = 0, \quad w(2, 1) = 0, \quad w(2, 2) = 1$$

получаем

$$a(i, j, k) = u(i, 1) v(j, 1) w(k, 1) + u(i, 2) v(j, 2) w(k, 2).$$

Теперь перейдем к случаю, когда хотя бы одна из матриц A_1, A_2 невырожденная. Для определенности пусть это будет A_2 . Умножая оба сечения на A_2^{-1} (например,

справа), получаем тензор, в котором второе сечение есть единичная матрица. Чтобы не вводить новых обозначений, будем теперь считать, что $A_2 = I$.

Если матрица $A_1 = \begin{bmatrix} a_{11} & a_{21} \\ a_{21} & a_{22} \end{bmatrix}$ диагонализуема, то тензорный ранг не выше 2. Предположим, что это не так. Тогда изменением ровно одного элемента матрицу A_1 можно превратить в диагонализуемую. Это очевидно, если $a_{12} = 0$ или $a_{21} = 0$.

Пусть оба внедиагональных элемента ненулевые. Характеристический многочлен матрицы A_1 имеет вид

$$f(\lambda) = \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}).$$

Ясно, что если $f(\lambda)$ имеет пару различных корней из $\mathbb{P}$, то матрица A_1 диагонализуема. Для этого достаточно, чтобы дискриминант Δ для $f(\lambda)$ был отличен от нуля. Дискриминант имеет вид

$$\Delta = (a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21}) = (a_{11} - a_{22})^2 + 4a_{12}a_{21}.$$

Чтобы получить $\Delta \neq 0$, достаточно изменить, скажем, a_{11} . Например, можно взять $a_{11} := a_{22}$. Таким образом, при $\mathbb{P} = \mathbb{C}$ тензорный ранг не выше 3. В вещественном случае нам нужны различные *вещественные* корни. Для этого достаточно положительности дискриминанта Δ , а этого легко добиться выбором достаточно большого значения для a_{11} . Теорема доказана.

ПРИМЕР 1. Рассмотрим тензор a с сечениями

$$A_1 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad A_2 = I.$$

Матрица A_1 имеет пару различных комплексных собственных значений $\pm i$ и поэтому диагоналуема над $\mathbb{C}$. Значит, тензорный ранг для a над $\mathbb{C}$ не выше 2. Согласно лемме 2.4.1, ранг не может равняться 0 или 1. Следовательно,

$$\text{trank}_{\mathbb{C}}(a) = 2.$$

В то же время, $a \in \mathbb{R}^{2 \times 2 \times 2}$, а матрица A_1 недиагонализуема над $\mathbb{R}$, так как не имеет вещественных собственных значений. Поэтому

$$\text{trank}_{\mathbb{R}}(a) = 3.$$

ПРИМЕР 2. Рассмотрим тензор a с сечениями

$$A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad A_2 = I.$$

Матрица A_1 является жордановой клеткой и поэтому недиагонализуема над $\mathbb{C}$, а значит и над $\mathbb{R}$. Таким образом,

$$\text{trank}_{\mathbb{C}}(a) = \text{trank}_{\mathbb{R}}(a) = 3.$$

2.6 Типичные и главные ранги

Число R называется *типичным рангом* на множестве вещественных тензоров $\mathbb{R}^{m \times n \times q}$, если

$$\text{trank}(a) = R$$

на множестве ненулевой меры Лебега в $\mathbb{R}^{m \times n \times q}$. Аналогичное определение можно принять и для комплексных тензоров.

Если ранг равен R почти всюду, то R называется *главным рангом* и обозначается $\text{grank}(m, n, q)$. Оказывается, для комплексных тензоров любые фиксированные размеры m, n, q однозначно определяют некоторое значение главного ранга, т.е. для комплексных тензоров типичный ранг только один и совпадает с главным рангом. Это утверждение не является тривиальным и опирается на некоторые факты из алгебраической геометрии. Общую теорему мы сформулируем и докажем позже. А прямо сейчас мы можем легко получить ее частный случай.

Теорема 2.6.1 *При $n \geq 2$ имеет место равенство $\text{grank}_{\mathbb{C}}(n, n, 2) = 2$.*

ДОКАЗАТЕЛЬСТВО. Ясно, что почти всюду сечение A_2 будет невырожденной матрицей и при этом матрица $A_1 A_2^{-1}$ будет иметь пару различных ненулевых собственных значений и поэтому будет диагонализуемой над $\mathbb{C}$. Теорема доказана.

Теорема 2.6.2 *Вещественные тензоры размера $2 \times 2 \times 2$ имеют в точности два значения типичных рангов, равные 2 и 3.*

ДОКАЗАТЕЛЬСТВО. Матрица A_2 будет невырожденной почти всюду на $\mathbb{R}^{m \times n \times q}$. Остается изучить случаи диагонализуемости и недиагонализуемости над $\mathbb{R}$ матрицы $A_1 A_2^{-1}$. А они, как нетрудно видеть, определяются положительностью и отрицательностью дискриминанта характеристического многочлена этой матрицы. Множество тензоров, для которых дискриминант равен нулю, имеет нулевую меру. Теорема доказана.

2.7 Скелетоны и полиномиальные отображения

Обозначим через X_r множество тензоров из $\mathbb{C}^{m \times n \times q}$, допускающих канонические разложения ранга r . Другими словами, X_r состоит из тех и только тех тензоров, которые представимы суммой скелетонов с числом слагаемых не более r .

Используя векторизацию тензоров, мы отождествляем пространство комплексных тензоров $\mathbb{C}^{m \times n \times q}$ с векторным пространством $\mathbb{C}^{mnq}$. Тогда отдельный скелетон является вектором специального вида

$$u_\alpha \otimes v_\alpha \otimes w_\alpha, \quad \text{где } u_\alpha \in \mathbb{C}^m, \quad v_\alpha \in \mathbb{C}^n, \quad w_\alpha \in \mathbb{C}^q,$$

а их сумма определяет вектор (векторизованный тензор)

$$a \in X_r \quad \Leftrightarrow \quad a = \sum_{\alpha=1}^r u_\alpha \otimes v_\alpha \otimes w_\alpha.$$

Каждая координата вектора a , очевидно, является полиномом от координат векторов $u_\alpha, v_\alpha, w_\alpha$. Эти полиномы определяют *полиномиальное отображение*

$$\phi_r : \mathbb{C}^{(m+n+q)r} \rightarrow \mathbb{C}^{mnq},$$

$$a = \phi_r(u_1, v_1, w_1, \dots, u_r, v_r, w_r) := \sum_{\alpha=1}^r u_\alpha \otimes v_\alpha \otimes w_\alpha.$$

Множество значений (образ) отображения ϕ_r отождествляется с множеством тензоров канонического тензорного ранга не выше r , т.е.

$$X_r = \phi_r(\mathbb{C}^{(m+n+q)r}). \quad (23)$$

Для того чтобы прояснить строение множеств X_r , естественно использовать некоторые знания о строении образов общих полиномиальных отображений. Эти знания относятся к основам *алгебраической геометрии*. В целях замкнутости изложения мы приведем определения и факты, нужные нам для изучения тензоров.

2.8 Алгебраические многообразия и идеалы

Множество $V \subseteq \mathbb{C}^N$ называется *алгебраическим многообразием*, если существуют полиномы

$$f_1(X_1, \dots, X_N), \dots, f_s(X_1, \dots, X_N) \in \mathbb{C}[X_1, \dots, X_N],$$

для которых множество V является множеством всех общих для них нулей:

$$V = \{(\xi_1, \dots, \xi_N) \in \mathbb{C}^N : f_l(\xi_1, \dots, \xi_N) = 0, 1 \leq l \leq s\}.$$

Удобно считать что *любая* система полиномов определяет некоторое алгебраическое многообразие. В частности, это может быть пустое множество, определяемое системой полиномов без общих нулей, или же все пространство $\mathbb{C}^N$, определяемое нулевым полиномом.

Пусть $\mathcal{M}$ – произвольное подмножество из $\mathbb{C}^N$. Обозначим через $I = I(\mathcal{M})$ множество всех полиномов из кольца $K = \mathbb{C}[X_1, \dots, X_N]$, обращающихся в нуль на всем множестве $\mathcal{M}$. Очевидно проверяется, что $I \subseteq K$ обладает следующими свойствами:

- если $f, g \in I$, то $f + g \in I$;
- если $f \in I$ и $g \in K$, то $fg \in I$.

Любое множество I элементов кольца полиномов K , удовлетворяющее этим двум требованиям, называется *идеалом* в K . Легко видеть, что $I(\mathcal{M})$ является идеалом для *любого* множества $\mathcal{M}$. Если $\mathcal{M}$ – пустое множество, то $I(\mathcal{M}) = K$.

Пусть J – произвольный идеал кольца K . Обозначим через $V(J)$ множество всех общих нулей полиномов из идеала J . Легко проверяется, что

$$\mathcal{M} \subseteq V(I(\mathcal{M})), \quad J \subseteq I(V(J)).$$

При этом каждое из включений может быть строгим. Другими словами, не каждое множество $\mathcal{M}$ является алгебраическим многообразием и не каждый идеал J имеет вид $I(V)$ для какого-либо алгебраического многообразия V .

Отметим две теоремы Гильберта, составляющие фундамент теории идеалов в кольце полиномов.

Теорема 2.8.1 *Любой идеал $J \subseteq K$ является конечно порожденным, т.е. для некоторой конечной системы полиномов $f_1, \dots, f_s \in J$ имеет место равенство*

$$J = \{f_1g_1 + \dots + f_sg_s : g_1, \dots, g_s \in K\}.$$

Теорема 2.8.2 *Пусть имеются идеал $J \subseteq K$ и полином f , обращающийся в нуль на множестве $V(J)$. Тогда $f^k \in J$ для некоторого натурального числа k .*

Идеал J называется *радикальным*, если он содержит все полиномы, которые при возведении в некоторую степень попадают в J .

Следствие 2.8.1 *Любой идеал J вида $J = I(\mathcal{M})$ является радикальным.*

Доказательства этих теорем можно найти в разделах учебников по алгебре, посвященных коммутативной алгебре и теории полиномов, а также в подготовительных разделах книг по алгебраической геометрии (см., например, [11]). Теорему 2.8.1 обычно называют *теоремой Гильберта о базисе*. Теорема 2.8.2 называется *теоремой Гильберта о нулях* и есть мода употребления для нее немецкого наименования Nullstellensatz.

Согласно следствию 2.8.1, имеет место *взаимно однозначное соответствие между алгебраическими многообразиями $\mathcal{V}$ и радикальными идеалами $\mathcal{I}$:*

$$\mathcal{V} = V(I(\mathcal{V})), \quad \mathcal{I} = I(V(\mathcal{I})).$$

Самому большому идеалу $I = K$ соответствует пустое множество. Следовательно, для любого идеала $I \subsetneq K$ соответствующее ему алгебраическое многообразие не является пустым множеством, т.е. существует точка $\xi \in \mathbb{C}^N$, в которой обращается в нуль каждый полином из I . В определенном смысле это утверждение можно считать обобщением основной теоремы алгебры.

2.9 Неприводимость

В качестве упражнения можно предложить проверку следующего полезного утверждения: *пересечение и объединение алгебраических многообразий V_1 и V_2 остаются алгебраическими многообразиями, при этом*

$$I(V_1 \cap V_2) = I(V_1) + I(V_2), \quad I(V_1 \cup V_2) = I(V_1) \cap I(V_2).$$

При изучении алгебраических многообразий (особенно в связи с важным понятием *размерности*) особая роль отводится свойству неприводимости. Алгебраическое

многообразие называется *неприводимым*, если его нельзя представить в виде объединения двух нетривиальных подмножеств, являющихся алгебраическими многообразиями.

Заметим, что *пространство $\mathbb{C}^N$ является неприводимым алгебраическим многообразием*. От противного, пусть $\mathbb{C}^N = V_1 \cup V_2$, где V_1 и V_2 – алгебраические многообразия, отличные от $\mathbb{C}^N$. Тогда существуют ненулевые полиномы

$$f \in V_1, \quad g \in V_2,$$

произведение которых fg обращается в нуль во всех точках $\mathbb{C}^N$. Такой полином должен быть нулевым и поэтому не может быть произведением ненулевых полиномов.

Теорема 2.9.1 *Любое алгебраическое многообразие является конечным объединением попарно различных неприводимых алгебраических многообразий, которые определяются однозначно по данному многообразию.*

Это утверждение того же типа, что и знаменитая *основная теорема арифметики* (об однозначном разложении целых чисел на простые множители). Оно существенно опирается на теорему Гильберта о базисе (теорема 2.8.1), и в силу этого доказательство получается не очень сложное, его можно найти в учебниках по коммутативной алгебре и теории полиномов [11].

Понятие неприводимости алгебраического многообразия связано с понятием простого идеала. Идеал $J \subseteq K$ называется *простым*, если из соотношений $fg \in J$ и $g \notin J$ вытекает $f \in J$. Справедливо следующее утверждение: *если V – алгебраическое многообразие, то идеал $I = I(V)$ является простым тогда и только тогда, когда V неприводимо*. Заметим также, что любой простой радикал является радикальным и поэтому имеет вид $I(V)$ для некоторого алгебраического многообразия V .

Для нас существенным является следующее утверждение, интуитивно вроде бы понятное, но совсем не тривиальное и не очень легкое для доказательства (хорошее изложение имеется, например, в [12]).

Теорема 2.9.2 *Пусть A и B – алгебраические многообразия, связанные соотношением $A \subsetneq B$. Если B неприводимо, то разность $B \setminus A$ всюду плотна в B (в обычной топологии).*

Заметим, что теорема 2.9.2 неверна без предположения о неприводимости B : для контрпримера можно взять $B \subset \mathbb{C}^2$, заданное уравнением $xy = 0$, и A , заданное уравнением $x = 0$.

В случае неприводимого алгебраического многообразия V принято говорить, что некоторое свойство выполняется на V *почти всюду*, если все точки, в которых оно не выполняется, принадлежат некоторому алгебраическому многообразию $W \subsetneq V$.

2.10 Размерность

Рассмотрим алгебраическое многообразие V , заданное одним полиномиальным уравнением

$$f(X_1, \dots, X_N) = 0.$$

Касательная плоскость для полинома f в точке

$$\xi = (\xi_1, \dots, \xi_N) \in \mathbb{C}^N$$

определяется вполне естественным образом – как множество точек

$$x = (x_1, \dots, x_N) \in \mathbb{C}^N,$$

удовлетворяющих линейному уравнению

$$\sum_{i=1}^N \frac{\partial f}{\partial X_i} (x_i - \xi_i) = 0.$$

Заметим, что частные производные для полиномов можно определить формальными алгебраическими правилами, не прибегая к теории дифференциального исчисления.

Если все частные производные равны нулю, то в этой точке касательная плоскость для f , очевидно, совпадает с $\mathbb{C}^N$. Скоро выяснится, что касательная плоскость для многообразия V в той же точке может оказаться лишь частью касательной плоскости для полинома f .

Если V – произвольное алгебраическое многообразие, то его *касательная плоскость* $T_\xi(V)$ в точке ξ определяется как пересечение касательных плоскостей в точке ξ для *всех* полиномов $f \in I(V)$. Подчеркнем, что требуется рассмотреть *все полиномы* из идеала $I(V)$ и, вообще говоря, нельзя ограничиться произвольной частной системой полиномов, определяющей V . Например, пусть $V \subset \mathbb{C}^2$ представляет собой прямую $x = 0$. Очевидно, V можно определить как множество нулей полинома $f(x, y) = x^2$. Однако, $T_\xi(V)$ не совпадает с касательной плоскостью для $f(x, y)$ ни в одной точке! Таким образом, касательная плоскость для полинома может не совпадать с касательной плоскостью для определяемого этим полиномом алгебраического многообразия. Тем не менее, мы не утверждаем, что касательную плоскость нельзя определить по некоторой конечной системе полиномов – нужно лишь понимать, что такая система не может быть произвольной системой с множеством общих нулей V .

Задача. Докажите, что в случае неприводимого полинома f касательная плоскость для f в любой точке совпадает с касательной плоскостью для алгебраического многообразия $f = 0$.

Размерность *неприводимого* алгебраического многообразия V определяется как *минимальная* размерность его касательных плоскостей:

$$\dim V = \min_{\xi \in V} \dim T_\xi(V).$$

Пусть V неприводимо и определяется конечной системой полиномиальных уравнений

$$\begin{aligned} f_1(X_1, \dots, X_N) &= 0, \\ &\dots \\ f_r(X_1, \dots, X_N) &= 0. \end{aligned}$$

Предположим, что данная система определяющих уравнений для V такова, что ранг матрицы Якоби

$$J_\xi = \begin{bmatrix} \frac{\partial f_1}{\partial X_1}(\xi) & \cdots & \frac{\partial f_1}{\partial X_N}(\xi) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_r}{\partial X_1}(\xi) & \cdots & \frac{\partial f_r}{\partial X_N}(\xi) \end{bmatrix}$$

в точке ξ равен числу ее строк r . Тогда можно утверждать [11], что

$$\dim V = \dim T_\xi(V) = N - r.$$

Следовательно, размерность V – это размерность касательной плоскости $T_\xi(V)$ в любой точке ξ , в которой матрица Якоби J_ξ имеет максимально возможный ранг (при условии что этот ранг равен числу строк матрицы Якоби).

Пусть $I(V) = \{f_\alpha\}$ и J_ξ – множество строк, составленных из частных производных f_α по переменным $X_1, \dots, X_N$. Можно считать, что J_ξ есть матрица с числом столбцов N и с бесконечным числом строк, которые соответствуют полиномам $f_\alpha \in I(V)$. Рассмотрим алгебраическое многообразие W точек $\xi \in V$, в которых все миноры порядка r для J_ξ обращаются в нуль. Согласно теореме 2.9.2, если $W \subsetneq V$, то множество $V \setminus W$ всюду плотно в V . Таким образом, для почти всех точек ξ неприводимого алгебраического многообразия V имеет место равенство $\dim V = \dim T_\xi(V)$.

Если V – произвольное алгебраическое многообразие, то оно, согласно теореме 2.9.1, является конечным объединением

$$V = \bigcup_{\alpha} V_{\alpha}$$

однозначно определенных попарно различных неприводимых алгебраических многообразий V_{α} . По определению, размерность V есть максимум из размерностей его неприводимых компонент V_{α} . Размерность пустого множества полагается равной -1 .

Под *размерностью алгебраического многообразия V в точке $\xi \in V$* понимается максимальная размерность содержащих эту точку неприводимых компонент:

$$\dim_{\xi} V = \max_{\alpha: \xi \in V_{\alpha}} \dim V_{\alpha}.$$

Точка ξ называется *особой*, если $\dim T_{\xi}(V) \neq \dim_{\xi} V$. В действительности для любой особой точки выполняется строгое неравенство $\dim T_{\xi}(V) > \dim_{\xi}(V)$. Известно также, что все точки пересечения любой пары различных неприводимых компонент являются особыми и почти все точки любой неприводимой компоненты не являются особыми [11].

ПРИМЕР. Пусть $V \subset \mathbb{C}^3$ определяется уравнениями $xz = 0$, $yz = 0$. Тогда V есть объединение двух неприводимых множеств: плоскости V_1 , определяемой уравнением $z = 0$, и прямой V_2 , определяемой уравнениями $x = y = 0$. Если $\xi = (0, 0, 0)$, то $\dim T_{\xi}(V) = 3$; в остальных случаях $\dim T_{\xi}(V) = 2$ при $\xi \in V_1$ и $\dim T_{\xi}(V) = 1$ при $\xi \in V_2$. Таким образом, равенство $\dim T_{\xi}(V) = \dim_{\xi}(V)$ нарушается только при $\xi = (0, 0, 0)$, в данном случае это и есть единственная особая точка, лежащая на пересечении неприводимых компонент V_1 и V_2 .

Для неприводимого алгебраического многообразия V есть и другие (эквивалентные) подходы к определению размерности, очень интересные с алгебраической точки

зрения. Во-первых, размерность совпадает с максимальным числом алгебраически независимых элементов в поле рациональных функций, определенных на V . Во-вторых, она равна максимально возможному числу r в цепочках непустых неприводимых алгебраических многообразий вида

$$V_0 \subsetneq V_1 \subsetneq V_2 \subsetneq \dots \subsetneq V_r = V.$$

Отсюда очевидно вытекает следующее свойство размерности, которое нам понадобится при доказательстве существования главного ранга для комплексных тензоров фиксированного размера.

Теорема 2.10.1 Пусть множества A и B являются алгебраическими многообразиями и при этом B неприводимо и $A \subsetneq B$. Тогда $\dim A < \dim B$.

Утверждения, на которых эта теорема базируется, не являются очевидными или простыми (см., например, [11, 12]). Отметим еще одну полезную теорему о размерности, напоминающую по смыслу теорему о размерности суммы подпространств из линейной алгебры (доказательство есть, например, в [12]).

Теорема 2.10.2 Пусть алгебраические многообразия $V_1, V_2 \in \mathbb{C}^N$ имеют непустое пересечение. Тогда

$$\dim V_1 + \dim V_2 - \dim(V_1 \cap V_2) \leq N.$$

2.11 Образы полиномиальных отображений

Наименьшее алгебраическое многообразие, содержащее заданное множество, называется его *алгебраическим замыканием*.

Рассмотрим произвольное полиномиальное отображение $g : \mathbb{C}^M \rightarrow \mathbb{C}^N$.

Лемма 2.11.1 Пусть Z – неприводимое алгебраическое многообразие в $\mathbb{C}^M$. Алгебраическое замыкание образа $g(Z)$ является неприводимым алгебраическим многообразием.

ДОКАЗАТЕЛЬСТВО. Пусть $g(Z)$ содержится в алгебраическом многообразии V и не содержится ни в каком меньшем алгебраическом многообразии. От противного, пусть $V = V_1 \cup V_2$, где V_1 и V_2 – алгебраические многообразия, отличные от V . Тогда Z есть объединение их полных прообразов:

$$Z = g^{-1}(V_1) \cup g^{-1}(V_2).$$

Легко видеть, что полные прообразы алгебраических многообразий при полиномиальном отображении будут алгебраическими многообразиями. В нашем случае они не являются пустыми множествами и не совпадают с Z , что противоречит неприводимости Z . Лемма доказана.

Следствие 2.11.1 Алгебраическое замыкание образа $g(\mathbb{C}^M)$ является неприводимым алгебраическим многообразием.

Теорема 2.11.1 Алгебраическое замыкание множества $A = g(\mathbb{C}^M)$ совпадает с его замыканием $\bar{A}$ в обычной топологии. При этом существует алгебраическое многообразие $B \subsetneq \bar{X}$ такое, что

$$\bar{A} \setminus B \subseteq A \subseteq \bar{A}.$$

Следствие 2.11.2 Множество A всюду плотно в $\bar{A}$.

Теорема 2.11.1 для наших целей весьма существенна. Ее доказательство требует изрядных усилий, доступное изложение можно найти в [11]. Для доказательства следствия достаточно принять во внимание лемму 2.11.1 и теорему 2.9.2.

Более детальное строение множества A дает следующий результат: *множество A является конечным объединением разностей алгебраических многообразий* [11]. Множества такого типа называются *конструктивными*.

При изучении тензоров нас интересуют образы $A = g(\mathbb{C}^M)$ при полиномиальных отображениях g специального вида. Пусть отображение g задается полиномами

$$\begin{aligned} X_1 &= g_1(Y_1, \dots, Y_M), \\ &\vdots \\ X_N &= g_N(Y_1, \dots, Y_M). \end{aligned}$$

Тогда для вычисления размерности неприводимого алгебраического многообразия $\bar{A}$ следует поинтересоваться рангами матрицы Якоби

$$J_\zeta(g) = \begin{bmatrix} \frac{\partial g_1}{\partial Y_1}(\zeta) & \dots & \frac{\partial g_1}{\partial Y_N}(\zeta) \\ \vdots & & \vdots \\ \frac{\partial g_N}{\partial Y_1}(\zeta) & \dots & \frac{\partial g_N}{\partial Y_N}(\zeta) \end{bmatrix}.$$

Теорема 2.11.2 Размерность замыкания $\bar{A}$ полиномиального образа $A = g(\mathbb{C}^M)$ равна максимальному рангу матриц Якоби $J_\zeta(g)$ в точках $\zeta \in \mathbb{C}^M$.

ДОКАЗАТЕЛЬСТВО. Пусть максимальный ранг матрицы Якоби равен r . Рассмотрим произвольную точку $\zeta \in \mathbb{C}^M$, в которой матрица $J_\zeta(g)$ имеет максимальный ранг и при этом точка $\xi = g(\zeta)$ не является особой точкой неприводимого алгебраического многообразия $\bar{A}$. В окрестности точки ξ многообразие $\bar{A}$ является частью A и имеет размерность, равную r . В силу того, что такими являются почти все точки из $\mathbb{C}^M$, получаем $r = \dim \bar{A}$. Теорема доказана.

Задача 1. Докажите, что множество комплексных $m \times n$ матриц, ранг которых не превышает $k \leq \min(m, n)$, является неприводимым алгебраическим многообразием размерности $k(m + n - k)$.

Задача 2. Пусть L – произвольное линейное подпространство в пространстве комплексных матриц размера $m \times n$, и пусть $1 \leq k \leq \min(m, n)$. Докажите, что если $\dim L \geq d \equiv (m - k)(n - k) + 1$, то L содержит ненулевую матрицу ранга не выше k .

Такой матрицы может не найтись в L , если $\dim L < d$. Например, пусть $m = n = 3$ и $k = 1$, тогда $d = 5$. Приведите пример четырех матриц порядка 3, линейная оболочка которых не содержит ни одной матрицы ранга 1.

2.12 Многообразие скелетонов

Пространство тензоров $\mathbb{C}^{m \times n \times q}$ отождествляется с векторным пространством $\mathbb{C}^{mnq}$. В этом пространстве скелетоны представляются векторами вида

$$u \otimes v \otimes w, \quad \text{где } u \in \mathbb{C}^m, \quad v \in \mathbb{C}^n, \quad w \in \mathbb{C}^q.$$

Множество скелетонов называется также *многообразием Сегре*.

Легко видеть, что множество скелетонов является образом полиномиального отображения

$$g(u, v, w) = u \otimes v \otimes w. \quad (24)$$

Координаты вектора $g(u, v, w) \in \mathbb{C}^{mnq}$ естественно нумеруются с помощью составных индексов ijk , перебираемых в лексикографическом порядке, и, очевидно, определяются полиномами

$$g_{ijk}(u, v, w) = u_i v_j w_k.$$

Пусть $J = J(u, v, w)$ обозначает матрицу Якоби отображения g в точке u, v, w . Она имеет mnq строк и $m + n + q$ столбцов. Строки соответствуют функциям g_{ijk} и нумеруются составными индексами ijk . Столбцы естественно разбиваются на три группы, соответствующие координатам u_i, v_j, w_k векторов u, v, w .

Таким образом, J состоит из трех прямоугольных блоков $J = [J^1, J^2, J^3]$. Обозначим через e_i^1, e_j^2, e_k^3 столбцы единичной матрицы соответственно порядка m, n, q . Нетрудно вычислить, что

$$J_{i'jk,i}^1 = \frac{\partial g_{ijk}}{\partial u_i} = (e_i^1)_{i'} v_j w_k,$$

$$J_{ij'k,j}^2 = \frac{\partial g_{ijk}}{\partial v_j} = u_i (e_j^2)_{j'} w_k,$$

$$J_{ijk',k}^3 = \frac{\partial g_{ijk}}{\partial w_k} = u_i v_j (e_k^3)_{k'}.$$

Следовательно, столбцы блоков J^1, J^2, J^3 имеют соответственно вид

$$e_i^1 \otimes v \otimes w, \quad u \otimes e_j^2 \otimes w, \quad u \otimes v \otimes e_k^3.$$

Полученный результат запишем в виде следующей леммы.

Лемма 2.12.1 *Матрица Якоби J отображения (24) имеет блочно-столбцовый вид*

$$J = [J^1, J^2, J^3],$$

$$J^1 = [e_i^1 \otimes v \otimes w], \quad J^2 = [u \otimes e_j^2 \otimes w], \quad J^3 = [u \otimes v \otimes e_k^3],$$

$$1 \leq i \leq m, \quad 1 \leq j \leq n, \quad 1 \leq k \leq q.$$

Следствие 2.12.1 *Для любых ненулевых векторов u, v, w имеет место равенство*

$$\text{rank } J(u, v, w) = m + n + q - 2.$$

ДОКАЗАТЕЛЬСТВО. Достаточно заметить, что вектор $u \otimes v \otimes w$ представляется линейной комбинацией столбцов матрицы J^1 и одновременно линейными комбинациями столбцов J^2 и столбцов J^2 . При выборе $u = e_1^1$, $v = e_1^2$, $w = e_1^3$ в матрице J получается ровно $m + n + q - 2$ линейно независимых столбцов. В случае произвольных ненулевых векторов u, v, w существуют невырожденные матрицы Q_1, Q_2, Q_3 такие, что

$$Q_1 u = e_1^1, \quad Q_2 v = e_1^2, \quad Q_3 w = e_1^3,$$

и нетрудно проверить, что

$$J(e_1^1, e_1^2, e_1^3) = (Q_1 \otimes Q_2 \otimes Q_3) J(u, v, w).$$

Ранг матрицы сохраняется при умножении на невырожденную матрицу. Лемма доказана.

Теорема 2.12.1 *Множество комплексных скелетов размера $m \times n \times q$ является неприводимым алгебраическим многообразием размерности $m + n + q - 2$.*

ДОКАЗАТЕЛЬСТВО. Если канонический ранг тензора $a(i, j, k)$ не выше 1, то его матрицы развертки по каждому из трех измерений имеют ранг не выше 1 (в точности 1, если тензор ненулевой). Поэтому в этих матрицах все миноры второго порядка равны нулю. Миноры второго порядка, очевидно, являются полиномами от элементов тензора a . Они и образуют систему полиномов, для которых скелетон a является общим нулем. Легко проверяется и то, что каждый общий нуль этих полиномов является скелетоном. Значит, множество скелетов является алгебраическим многообразием. Неприводимость вытекает из леммы 2.11.1. Для вычисления размерности опираемся на теорему 2.11.2. Согласно следствию 2.12.1, максимальный ранг матрицы Якоби отображения, образом которого является множество скелетов, в точности равен $m + n + q - 2$. Теорема доказана.

2.13 Незамкнутость множества сумм двух скелетов

Согласно ранее введенным обозначениям, множество тензоров ранга не выше 2 есть

$$X_2 = \phi_2(\mathbb{C}^{2(m+n+q)}).$$

Докажем, что X_2 не является алгебраическим многообразием.

Для этого достаточно убедиться в том, что множество X_2 незамкнуто в обычной топологии, и заметить, что любое алгебраическое многообразие является замкнутым множеством.

Пусть $a_1, b_1 \in \mathbb{C}^m$, $a_2, b_2 \in \mathbb{C}^n$, $a_3, b_3 \in \mathbb{C}^q$. Рассмотрим скелетон

$$S(\varepsilon) = (a_1 + \varepsilon b_1) \otimes (a_2 + \varepsilon b_2) \otimes (a_3 + \varepsilon b_3).$$

Раскрывая скобки, находим

$$S(\varepsilon) = S(0) + \varepsilon T + O(\varepsilon^2),$$

$$T = b_1 \otimes a_2 \otimes a_3 + a_1 \otimes b_2 \otimes a_3 + a_1 \otimes a_2 \otimes b_3.$$

Отсюда

$$T = \frac{1}{\varepsilon} S(\varepsilon) - \frac{1}{\varepsilon} S(0) + O(\varepsilon).$$

Мы видим, что в любой сколь угодно малой окрестности тензора T присутствует тензор ранга не выше 2.

Покажем, что в случае $m = n = q = 2$ векторы a_i, b_i можно выбрать таким образом, чтобы тензорный ранг T был равен 3. Возьмем

$$a_1 = a_2 = a_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad b_1 = b_2 = b_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Тогда тензор T имеет сечения

$$T_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Находим

$$T_1^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{и} \quad T_2 T_1^{-1} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Матрица $T_2 T_1^{-1}$ является жордановой клеткой и, следовательно, недиагонализуема. Учитывая теорему 2.5.1, приходим к выводу о том, что ранг тензора T равен 3 одновременно над $\mathbb{C}$ и над $\mathbb{R}$.

Задача. Докажите незамкнутость X_2 при произвольных значениях $m, n, q \geq 2$.

2.14 Главные ранги комплексных тензоров

Нам интересуют комплексные тензоры размера $m \times n \times q$. Среди них множество тензоров ранга не выше r получило ранее обозначение X_r . Согласно теореме 2.11.1, алгебраическое замыкание множества X_r совпадает с замыканием $\overline{X}_r$ в обычной топологии.

Лемма 2.14.1 Если $\overline{X}_{r-1} \subsetneq \mathbb{C}^{mnq}$, то $\overline{X}_{r-1} \subsetneq \overline{X}_r$.

ДОКАЗАТЕЛЬСТВО. Очевидно, что $X_{r-1} + X_1 \subseteq X_r$. Согласно теореме 2.12.1,

$$\overline{X}_1 = X_1.$$

Следовательно, $\overline{X}_{r-1} + X_1 \subseteq \overline{X}_r$. Если предположить, что $\overline{X}_{r-1} = \overline{X}_r$, то

$$\overline{X}_{r-1} + X_1 \subseteq \overline{X}_{r-1}$$

и, как следствие, для любого натурального числа s получаем

$$\overline{X}_{r-1} + s X_1 \subseteq \overline{X}_{r-1}.$$

В силу теоремы 2.2.1, пространство $\mathbb{C}^{mnq}$ есть сумма конечного числа множеств X_1 . Пусть $s X_1 = \mathbb{C}^{mnq}$. Тогда

$$\mathbb{C}^{mnq} \subseteq \overline{X}_{r-1},$$

что противоречит условию леммы.

Теорема 2.14.1 *Существует натуральное число r такое, что*

$$\overline{X}_1 \subsetneq \dots \subsetneq \overline{X}_{r-1} \subsetneq \overline{X}_r = \mathbb{C}^{mnq}.$$

ДОКАЗАТЕЛЬСТВО. Согласно теореме 2.10.1, если $\overline{X}_{r-1} \subsetneq \overline{X}_r$, то $\dim \overline{X}_{r-1} < \dim \overline{X}_r$. Поэтому рано или поздно размерность будет равна mnq , т.е. при некотором r получается равенство $\overline{X}_r = \mathbb{C}^{mnq}$. Теорема доказана.

Таким образом, почти все тензоры из $\mathbb{C}^{mnq}$ имеют одно и то же значение канонического ранга. Тем самым доказано существование главного ранга (по английски *generic rank*) для комплексных тензоров фиксированного размера, $r = \text{grank}(m, n, q)$.

Канонический тензорный ранг любого тензора из множества $X_k \setminus X_{k-1}$ в точности равен k . На основе приведенных выше сведений о полиномиальных отображениях нетрудно придти к выводу о том, что *тензоры фиксированного ранга образуют конструктивное множество в пространстве $\mathbb{C}^{mnq}$.*

Любой тензор $a \in \overline{X}_k \setminus \overline{X}_{k-1}$ является пределом последовательности тензоров ранга k и не может быть пределом последовательности тензоров меньшего ранга. Говорят, что k является *граничным рангом* тензора a . Обозначение: $k = \text{brank}(a)$ (от английского *border rank*). Ясно, что $\text{brank}(a) \leq \text{grank}(m, n, q)$ для любого тензора $a \in \mathbb{C}^{mnq}$, при этом главный ранг является максимально возможным значением граничных рангов среди всех тензоров размера $m \times n \times q$.

Конечно же, выполняется неравенство $\text{grank}(m, n, q) \leq \text{mrank}(m, n, q)$, причем обычно оно является строгим. Число $k = \text{mrank}(m, n, q)$ можно охарактеризовать как минимальное натуральное число, для которого $X_{k-1} = X_k$.

2.15 Вычисление главного ранга

По заданным значениям m, n, q довольно легко можно получить число, которое *должно быть* значением главного ранга $\text{grank}(m, n, q)$. После этого, зная результат, мы понимаем, что именно следует доказывать.

Все операции сводятся к вычислению обычного ранга некоторой матрицы. Речь идет о матрице Якоби полиномиального отображения

$$\phi_r : \mathbb{C}^{(m+n+q)r} \rightarrow \mathbb{C}^{mnq},$$

которое определяется формулой

$$\phi_r(u_1, v_1, w_1, \dots, u_r, v_r, w_r) = \sum_{\alpha=1}^r u_\alpha \otimes v_\alpha \otimes w_\alpha.$$

Эта матрица содержит mnq строк и $(m+n+q)r$ столбцов. Мы уже подробно описали строение матрицы Якоби $J = J_1$ для отображения

$$g(u, v, w) = \phi_1(u, v, w).$$

Теперь заметим, что

$$\phi_r(u_1, v_1, w_1, \dots, u_r, v_r, w_r) = \sum_{\alpha=1}^r g(u_\alpha, v_\alpha, w_\alpha).$$

Поэтому матрица Якоби J_r для ϕ_r состоит из последовательности блоков, представляющих собой матрицы Якоби $J(u_\alpha, v_\alpha, w_\alpha)$ для отображения g :

$$J_r = [J(u_1, v_1, w_1), \dots, J(u_r, v_r, w_r)]. \quad (25)$$

Блочнo-столбцовая структура матриц $J(u_\alpha, v_\alpha, w_\alpha)$ описана в лемме 2.12.1.

Следующий результат есть некоторая модификация утверждения, известного в литературе как *лемма Террачини*.

Теорема 2.15.1 *Главный ранг $\text{grank}(m, n, q)$ равен минимальному натуральному числу $r = r(m, n, q)$, для которого найдутся векторы*

$$u_\alpha \in \mathbb{C}^m, \quad v_\alpha \in \mathbb{C}^n, \quad w_\alpha \in \mathbb{C}^q, \quad 1 \leq \alpha \leq r,$$

на которых ранг матрицы Якоби J_r вида (25) равен mnq .

ДОКАЗАТЕЛЬСТВО. Для любого r ранг матрицы J_r принимает свое максимальное значение почти всюду на $\mathbb{C}^{(m+n+q)r}$ и, согласно теореме 2.11.2, определяет размерность неприводимого алгебраического многообразия $\overline{X}_r$, где $X_r = \phi_r(\mathbb{C}^{(m+n+q)r})$. Если $r < \text{grank}(m, n, q)$, то $\dim \overline{X}_r < mnq$, и лишь при $r = \text{grank}(m, n, q)$ получаем

$$\dim \overline{X}_r = \dim \mathbb{C}^{mnq} = mnq.$$

Теорема доказана.

Следствие 2.15.1 *Для главного ранга справедлива оценка*

$$\text{grank}(m, n, q) \geq \left\lceil \frac{mnq}{m+n+q-2} \right\rceil.$$

ДОКАЗАТЕЛЬСТВО. Согласно следствию 2.12.1, ранг каждой матрицы $J(u_\alpha, v_\alpha, w_\alpha)$ не превышает $m+n+q-2$. Отсюда

$$\text{rank} J_r \leq (m+n+q-2)r,$$

и при $r = \text{grank}(m, n, q)$ получаем $mnq \leq (m+n+q-2)r$.

Процедура вычисления главного ранга основывается на том, что почти всюду ранг J_r принимает свое максимальное значение. Поэтому при каждом r максимальное значение ранга матрицы Якоби для ϕ_r можно найти по случайно выбираемым векторам $u_\alpha, v_\alpha, w_\alpha$. Учитывая оценку следствия 2.15.1, мы можем начинать вычисление со значения $r = \left\lceil \frac{mnq}{m+n+q-2} \right\rceil$. Проверяем, будет ли максимальный ранг равен mnq . Если нет, полагаем $r := r + 1$ и повторяем вычисление максимального ранга. Полученное значение главного ранга будет с большой вероятностью правильным. Однако, строгое доказательство того, что данное значение является главным рангом, может потребовать немалых усилий.

ПРИМЕР. Пусть $m = n = q = 2$. Тогда вычисление главного ранга следует начать со значения $r = \left\lceil \frac{8}{4} \right\rceil = 2$. На векторах

$$u_1 = v_1 = w_1 := e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad u_2 = v_2 = w_2 := e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

матрица Якоби состоит из столбцов

$$\begin{array}{cccccc} e_1 \otimes e_1 \otimes e_1, & e_2 \otimes e_1 \otimes e_1, & e_1 \otimes e_1 \otimes e_1, & e_1 \otimes e_2 \otimes e_1, & e_1 \otimes e_1 \otimes e_1, & e_1 \otimes e_1 \otimes e_2, \\ e_1 \otimes e_2 \otimes e_2, & e_2 \otimes e_2 \otimes e_2, & e_2 \otimes e_1 \otimes e_2, & e_2 \otimes e_2 \otimes e_2, & e_2 \otimes e_2 \otimes e_1, & e_2 \otimes e_2 \otimes e_2. \end{array}$$

Исключив совпадающие столбцы, находим ровно 8 линейно независимых столбцов, представляющих собой столбцы единичной матрицы порядка 8. Поэтому

$$\text{grank}(2, 2, 2) = 2.$$

Этот же результат ранее был получен из других соображений (см. теорему 2.6.1).

Задача. Докажите, что $\text{grank}(3, 3, 3) = 5$.

Случай $3 \times 3 \times 3$ оказывается интересным исключением из правила. Известен следующий общий результат Ликтига для кубических тензоров [18]:

$$\text{grank}(n, n, n) = \left\lceil \frac{n^3}{3n - 2} \right\rceil \quad \text{при всех } n \neq 3. \quad (26)$$

Рассуждения Ликтига используют некоторые понятия и построения Штрассена [22], связанные по существу с изучением рангов матрицы Якоби J_r . Разбор этих работ вряд ли покажется легким чтением, причем интересно заметить, что поиск максимальной линейно независимой системы столбцов в J_r осуществляется в них с помощью техники, мало похожей на привычные нам методы вычисления ранга матрицы.

Список литературы

- [1] С. А. Горейнов, Н. Л. Замарашкин, Е. Е. Тыртышников, Псевдоскелетные аппроксимации матриц, ДАН России, 343 (2), 1995, 151-152.
- [2] С. А. Горейнов, Е. Е. Тыртышников, Квазиоптимальность скелетного приближения матрицы в чебышевской норме, ДАН России, том 438, N 5, 593-594.
- [3] Н. Л. Замарашкин, И. В. Оселедец, Е. Е. Тыртышников, Тензорная структура обратных к ленточной теплицевой матрице, ДАН России, том 428, N 2 (2009), 161-162.
- [4] И. В. Оселедец, Е. Е. Тыртышников, Рекурсивное разложение многомерных тензоров, ДАН России, том 427, N 1 (2009), 14-16.
- [5] И. В. Оселедец, О новом тензорном разложении, ДАН России, том 427, N 2 (2009), 168-169.
- [6] И. В. Оселедец, О приближении матриц логарифмическим числом параметров, ДАН России, том 428, N 1 (2009), 23-24.
- [7] Е. Е. Тыртышников, Тензорные аппроксимации матриц, порожденных асимптотически гладкими функциями, Матем. сб., том 194, N 6, 147-160 (2003).

- [8] Е. Е. Тыртышников, Матричный анализ и линейная алгебра, Физматлит, 2007.
- [9] Е. Е. Тыртышников, Методы численного анализа, Издательский центр “Академия”, 2007.
- [10] M. D. Atkinson and S. Lloyd, Bounds on the rank of 3-tensors, *Linear Algebra Appl.*, 31 (1980), pp. 19–31.
- [11] D. Cox, J. Little, D. O’Shea, *Idelas, varieties, algorithms*, 3rd edition, Springer, 2007.
- [12] K. Kendig, *Elementary algebraic geometry*, Springer-Verlag, New York, 1977.
- [13] W. Hackbusch, B.N. Khoromskij, E.E. Tyrtysnikov, Hierarchical Kronecker tensor-product approximations, *J. Numer. Math.* **13** (2005), 119–156.
- [14] W. Hackbusch, B. N. Khoromskij, E. Tyrtysnikov, Approximate iterations for structured matrices, *Numer. Math.*, vol.109, no. 3, pp. 365–383 (2008).
- [15] R. A. Harshman, Foundations of the PARAFAC procedure: model and conditions for an ‘explanatory’ multi- mode factor analysis. *UCLA Working Papers Phonet.*, 16: 1–84 (1970).
- [16] T. Jiang, N. D. Sidiropoulos, Kruskal’s Permutation Lemma and the Identification of CANDECOMP/PARAFAC and Bilinear Models with Constant Modulus Constraints, *IEEE Trans. on Signal Processing*, vol. 52, no. 9 (2004), pp. 2625–2636.
- [17] J. B. Kruskal, Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics, *Linear Algebra Appl.*, 18: 95–138 (1977).
- [18] T. Lickteig, Typical tensorial rank, *Linear Algebra Appl.*, 69 (1985), pp. 95–120.
- [19] N. D. Sidiropoulos, R. Bro, On the uniqueness of multilinear decomposition of N-way arrays, *J. Chemometrics*, 14: 229–239 (2000).
- [20] A. Stegeman, On uniqueness conditions for Candecomp/Parafac and Indscal with full column rank in one mode, *Linear Algebra and its Applications*, 431 (2009) 211–227.
- [21] A. Stegeman, N. D. Sidiropoulos, On Kruskal’s uniqueness condition for the Candecomp/Parafac decomposition, *Linear Algebra and its Applications* 420 (2007) 540–552.
- [22] V. Strassen, Rank and optimal computation of generic tensors, *Linear Algebra Appl.*, 52/53 (1983), pp. 645–685.
- [23] I. V. Oseledets, Approximation of $2^d \times 2^d$ matrices using tensor decomposition, *SIAM J. Matrix Anal. Appl.*, 31 (2010), pp. 2130–2145.
- [24] I. Oseledets, E. Tyrtysnikov, Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM J. Sci. Comput.*, vol 31, no. 5 (2009), pp. 3744–3759.

- [25] I. Oseledets, E. Tyrtyshnikov, TT-cross approximation for multidimensional arrays, *Linear Algebra Appl.*, 432 (2010), pp. 70–88.
- [26] I. V. Oseledets, E. E. Tyrtyshnikov, Algebraic wavelet transform via quantics tensor train decomposition, *SIAM J. Sci. Comp.*, vol. 31, no. 3 (2011), pp. 1315-1328.
- [27] E. Tyrtyshnikov, Incomplete cross approximation in the mosaic-skeleton method, *Computing* 64, no. 4 (2000), 367–380.
- [28] E. Tyrtyshnikov, Tensor ranks for the inversion of tensor-product binomials, *J. Comput. Appl. Math.*, vol. 234, no. 11 (2010) 3170-3174.