# *Лекция 6. Тензорные разложения*

*М. В. Рахуба*

*28 февраля 2024 г.*

В этом курсе мы в основном работаем с матрицами — двумерными массивами, но данные могут иметь и бо́льшую размерность. Тензором размерности $d$ будем называть многомерный массив:

$$A = \{a_{i_1 \dots i_d}\}_{i_1,\dots,i_d=1}^{n_1,\dots,n_d} \in \mathbb{R}^{n_1 \times \dots \times n_d}.$$

Обращаем внимание, что такое определение тензора нередко можно встретить в публикациях по прикладной математике или глубинному обучению. Однако классически под тензором понимается более общий объект.

При $n_1 = \ldots = n_d = n$ хранение такого массива в компьютере требует $n^d$ ячеек памяти, что при больших $d$ может оказаться невозможным из-за экспоненциального роста $n^d$ с размерностью $d$. Например, $2^{300}$ намного больше оценки числа атомов во Вселенной. Это проявление так называемого *проклятия размерности*. Поэтому в случае с тензорами высокой размерности нас будут интересовать способы их представления с помощью небольшого числа параметров или другими словами нас будет интересовать сжатие тензоров. В этом нам помогут тензорные разложения, которые по сути являются обобщением скелетного разложения матриц на тензоры размерности $d \geq 3$.

На практике не часто встречаются тензоры, которые можно сжать с помощью тензорных разложений без потери качества. Поэтому нас будет также интересовать вопрос построения приближений с контролируемой ошибкой (по аналогии как мы это делали для матриц через SVD и теорему Эккарта-Янга-Мирского).

## 1 Кронекерово произвдение и векторизация

Свойства тензоров и тензорных разложений часто удается свести к матричному случаю. Чтобы научиться такие матрицы выписывать через элементы тензорных разложений, нам понадобятся две новые операции: кронекерово произведение и векторизация. Отметим также, что эти операции оказываются полезными и в других случаях, не связанных с тензорными разложениями. Например, при работе с изображениями.

### 1.1 Кронекерово произведение

**Определение 1.** *Кронекеровым произведением матриц $A \in \mathbb{C}^{m \times n}$ и $B \in \mathbb{C}^{p \times q}$ называется:*

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix} \in \mathbb{C}^{mp \times nq}.$$

Обратим внимание, что Кронекерово произведение определено для матриц любых размеров и содержит в себе всевозможные попарные произвдения элементов матриц $A$ и $B$.

**Пример 1.**

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

**Утверждение 1.** *Пусть $A, B, C, D$ — некоторые комплекснозначные матрицы. Тогда выполняются следующие свойства:*

1. *Ассоциативность:*
   $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.
2. *Дистрибутивность:*
   $(A + B) \otimes C = A \otimes C + B \otimes C$,
   $A \otimes (B + C) = A \otimes B + A \otimes C$.
3. $(A \otimes B)^\top = A^\top \otimes B^\top$ *и* $(A \otimes B)^* = A^* \otimes B^*$.
4. $A \otimes B = (A \otimes I)(I \otimes B) = (I \otimes B)(A \otimes I)$.
5. $(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$.
6. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ *для обратимых* $A$,$B$.

Формально свойство 4 записывается как $A \otimes B = (A \otimes I_p)(I_n \otimes B) = (I_m \otimes B)(A \otimes I_q)$, где $A \in \mathbb{C}^{m\times n},\ B \in \mathbb{C}^{p\times q}$, но для удобства мы будем позволять себе опускать размеры матриц $I$, так как в большинстве случаев они однозначно определяются из контекста.

Обратите особое внимание на свойство 5, которое будет нам неоднократно встречаться в дальнейшем. Также свойства 4 и 6 элементарно из него следуют.

*Доказательство.* Первые три свойства очевидно следуют из свойств транспонирования, комплексного сопряжения и умножения матриц, и их справедливость можно показать, явно расписав $ij$-ый элемент каждой итоговой матрицы. Здесь же докажем остальные.

4. Пользуясь правилом блочного умножения матриц:

$$(A \otimes I)(I \otimes B) = \begin{bmatrix}a_{11}I & \dots & a_{1n}I\\ \vdots & \ddots & \vdots\\ a_{m1}I & \dots & a_{mn}I\end{bmatrix}\begin{bmatrix}B & & \\ & \ddots & \\ & & B\end{bmatrix} = $$
$$= \begin{bmatrix}a_{11}B & \dots & a_{1n}B\\ \vdots & \ddots & \vdots\\ a_{m1}B & \dots & a_{mn}B\end{bmatrix} = A \otimes B.$$

   Аналогично показывается $(I \otimes B)(A \otimes I) = A \otimes B$.

5. Используя свойство 4 имеем:

$$\begin{aligned}(A \otimes C)(B \otimes D) &= (A \otimes I)(I \otimes C)(B \otimes I)(I \otimes D) =\\ &= (A \otimes I)(B \otimes I)(I \otimes C)(I \otimes D) =\\ &= ((AB) \otimes I)\,(I \otimes (CD))\,,\end{aligned}$$

где последнее равенство следует из $(A \otimes I)(B \otimes I) = (AB) \otimes I$ и $(I \otimes C)(I \otimes D) = I \otimes (CD)$. Действительно,

$$(I \otimes C)(I \otimes D) = \begin{bmatrix} C & & \\ & \ddots & \\ & & C \end{bmatrix} \begin{bmatrix} D & & \\ & \ddots & \\ & & D \end{bmatrix} = \\ = \begin{bmatrix} CD & & \\ & \ddots & \\ & & CD \end{bmatrix} = I \otimes CD.$$

Свойство $(A \otimes I)(B \otimes I) = (AB) \otimes I$ проверьте самостоятельно.

6. Пусть матрицы $A \in \mathbb{C}^{m \times m}$, $B \in \mathbb{C}^{n \times n}$ обратимы, тогда:

$$(A \otimes B)(A^{-1} \otimes B^{-1}) = (AA^{-1}) \otimes (BB^{-1}) = I_m \otimes I_n = I_{mn},$$

а значит $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

$\square$

**Замечание 1.** *Кронекерово произведение не является коммутативным. Смотри пример 1.*

На этом все свойства кронекерова произведения матриц, конечно, не заканчиваются. Мы познакомились только с самыми базовыми из них. Многие другие свойства можно получить используя свойство 5 из Утверждения 1 и применяя его к некоторым матричным разложениям. Например, несложно убедиться, что сингулярные числа $A \otimes B$ являются всевозможными попарными произведениями сингулярных чисел $A$ и $B$. Отсюда сразу следует, что

$$\|A \otimes B\|_2 = \|A\|_2 \|B\|_2.$$

## *1.2 Векторизация*

**Определение 2.** *Определим* $\mathrm{vec}\,(\cdot) \colon \mathbb{C}^{m \times n} \to \mathbb{C}^{mn}$ *как векторизацию матрицы по столбцам, то есть:*

$$\mathrm{vec}\left(\begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}\right) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad \textit{где} \quad a_k \in \mathbb{C}^m.$$

Мы не зря вводим векторизацию вместе с кронекеровым произведением. Покажем, как эти две операции связаны связаны друг с другом.

**Утверждение 2.** *Пусть определено произведение $AXB$. Тогда:*

$$\text{vec}\,(AXB) = (B^\top \otimes A)\,\text{vec}\,(X).$$

*Доказательство.* Сначала заметим, что

$$\text{vec}\left(uv^\top\right) = \text{vec}\left(\begin{bmatrix} v_1 u & v_2 u & \dots & v_n u \end{bmatrix}\right) = \begin{bmatrix} v_1 u \\ v_2 u \\ \vdots \\ v_n u \end{bmatrix} = v \otimes u.$$

Пусть $X = uv^\top$, тогда из выписанного выше выражения и свойства 5 Утверждения 1 получим:

$$\text{vec}\left(A(uv^\top)B\right) = \text{vec}\left((Au)(B^\top v)^\top\right) = (B^\top v) \otimes (Au) =$$
$$= (B^\top \otimes A)(v \otimes u) = (B^\top \otimes A)\,\text{vec}\left(uv^\top\right) = (B^\top \otimes A)\,\text{vec}\,(X).$$

Произвольный же $X$ является суммой матриц ранга 1, поэтому для него данное утверждение тоже справедливо в силу линейности $\text{vec}\,(\cdot)$. $\square$

## 2 Тензорные разложения

В этом разделе мы поговорим о двух разложениях, которые обобщают скелетное разложение матриц на случай тензоров размерности больше 2. Эти разложения называются каноническим разложением и разложением Таккера. В отличие от разложения Таккера, канонические разложение по своим свойствам будет заметно отличаться от скелетного матричного разложения. Как мы увидим, у каждого из разложений будут свои достоинства и недостатки при использовании в приложениях. Также отметим, что существует и множество других тензорных разложений, которые остаются за рамками настоящего курса, но представляют значительный интерес с практической точки зрения.

### 2.1 Каноническое разложение

**Определение 3.** *Пусть $A \in \mathbb{R}^{n_1 \times \ldots \times n_d}$, $B \in \mathbb{R}^{m_1 \times \ldots \times m_D}$ — тензоры размерностей $d$ и $D$ соответственно. Тогда их тензорным произведением назовем тензор $A \circ B \in \mathbb{R}^{n_1 \times \ldots \times n_d \times m_1 \times \ldots \times m_D}$ размерности $(d + D)$ с элементами:*

$$(A \circ B)_{i_1 \ldots i_d j_1 \ldots j_D} = a_{i_1 \ldots i_d} b_{j_1 \ldots j_D}.$$

Иногда тензорное произведение $\circ$ обозначают как $\otimes$, но в нашем курсе $\otimes$ уже занято под кронекерово произведения.

Посмотрим, как мы можем использовать тензорное произведение для записи скелетного разложения матриц. Во-первых, заметим, что любая матрица ранга 1 имеет вид $A = uv^\top$ или $a_{ij} = u_i v_j$ в индексной записи. Отсюда, используя только что введенное определение, получим:

$$A = uv^\top = u \circ v.$$

По аналогии можно определить трехмерный тензор ранга 1, как ненулевой тензор с элементами $a_{ijk} = u_i v_j w_k$. С использованием тензорного произведения получим:

$$A = u \circ v \circ w.$$

Обратите внимание, что мы можем не расставлять скобки в силу ассоциативности тензорного произведения.

Обсудим теперь случай ранга $r$. Вспомним, что любая матрица ранга $r$ записывается с помощью матриц $U = [u_1, \ldots, u_r]$ и $V = [v_1, \ldots, v_r]$:

$$A = UV^\top = u_1 \circ v_1 + \cdots + u_r \circ v_r, \tag{1}$$

причем ранг матрицы есть минимально возможное число слагаемых в (1). Обобщим это разложение на случай трехмерных массивов (обобщение для массивов размерности $d > 3$ делается аналогично).

**Определение 4.** *Каноническим разложением* $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *назовем разложение вида:*

$$A = \sum_{\alpha=1}^{r} u_\alpha \circ v_\alpha \circ w_\alpha.$$

*Матрицы* $U = [u_1, \ldots, u_r]$, $V = [v_1, \ldots, v_r]$ *и* $W = [w_1, \ldots, w_r]$ *назовем каноническими факторами, а минимально возможное* $r$ *— каноническим рангом. Для краткости также будем использовать следующее обозначение:*

$$A = [\![U, V, W]\!].$$

Теперь обсудим, что мы можем сказать про существование и единственность канонического разложения, и возможно ли найти наилучшее приближение тензора тензором меньшего канонического ранга. А также сравним эти свойства со скелетным матричным разложением. Подробнее про свойства канонического разложения можно ознакомится в [1].

[1] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009

**Замечание 2** (Существование)**.** *Каноническое разложение существует для любого тензора, причем* $r \le n_1 n_2 n_3$*, так как*

$$A = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} a_{ijk}\, e_i \circ e_j \circ e_k, \tag{2}$$

*где* $e_i$ *— векторы стандартного базиса. Более тонкое наблюдение заключается в том, что канонический ранг удовлетворяет* $r \leq \min(n_1 n_2, n_1 n_3, n_2 n_3)$ *(проверьте). Таким образом, как и в случае со скелетным разложением, оно всегда существует.*

**Замечание 3** (Единственность)**.** *Скелетное разложение для матриц не является единственным. Действительно, для произвольной невырожденной* $S$*:*

$$UV^\top = (US)\left(S^{-1}V^\top\right).$$

*Оказывается, что при некоторых предположениях на тензор, каноническое разложение может быть единственным с точностью до перестановки слагаемых и шкалирования векторов. В случае с матрицами это бы означало, что* $S$ *могло бы быть только произведением матрицы перестановки и диагональной матрицы, а не произвольной невырожденной матрицей.*

**Замечание 4** (Существование наилучшего приближения)**.** *Согласно теореме Эккарта-Янга-Мирского для любой ненулевой матрицы всегда найдется наилучшее приближение (в смысле унитарно-инвариантной нормы) матрицей меньшего ранга. Для канонического разложения это уже не так. Множество тензоров ранга* $\leq r$ *не является замкнутым, а значит можно, например, построить последовательность тензоров* $\{A_n\}\colon\ \operatorname{rank} A_n = 2$*, которая сходится к тензору* $A\colon\ \operatorname{rank} A = 3$*. Это может оказаться чревато разными неприятными эффектами при попытке построить наилучшее приближение заданного ранга с помощью численных алгоритмов. Подробнее об этом будет рассказано на семинаре.*

Таким образом, каноническое разложение обладает как некоторыми свойствами, которые могут оказаться полезными на практике, например, единственность, так и неприятными свойствами, как проблемы с построением наилучшего приближения заданного ранга. Далее мы обсудим разложение Таккера, которое по свойствам больше напоминает матричный случай.

## 2.2 Разложение Таккера

Для следующей попытки обобщить скелетное разложение воспользуемся следующей его записью:

$$A = UGV^\top = \sum_{\alpha,\beta=1}^{r,r} g_{\alpha\beta}\, u_\alpha \circ v_\beta,$$

для некоторой $G \in \mathbb{R}^{r\times r}$. Для матриц нет разницы, в каком виде выбирать скелетное разложение, так как один его вид сводится к другому: $A = \widetilde{U}V^\top$, где $\widetilde{U} = UG$. Для тензоров размерности $d \geq 3$ различия уже будут.

**Определение 5.** *Разложением Таккера тензора $A \in \mathbb{R}^{n_1\times n_2\times n_3}$ называют разложение вида:*

$$A = \sum_{\alpha=1}^{r_1}\sum_{\beta=1}^{r_2}\sum_{\gamma=1}^{r_3} g_{\alpha\beta\gamma}\, u_\alpha \circ v_\beta \circ w_\gamma,$$

*где тензор*

$$G = \{g_{\alpha\beta\gamma}\}_{\alpha,\beta,\gamma=1}^{r_1,r_2,r_3} \in \mathbb{R}^{r_1\times r_2\times r_3},$$

*называется таккеровским ядром, а матрицы*

$$\begin{aligned} U &= [u_1,\ldots,u_{r_1}] \in \mathbb{R}^{n_1\times r_1},\\ V &= [v_1,\ldots,v_{r_2}] \ \in \mathbb{R}^{n_2\times r_2},\\ W &= [w_1,\ldots,w_{r_3}] \in \mathbb{R}^{n_3\times r_3}. \end{aligned}$$

*называются таккеровскими факторами. Кортеж чисел*

$$\textit{mlrank}(A) = (r_1, r_2, r_3)\,,$$

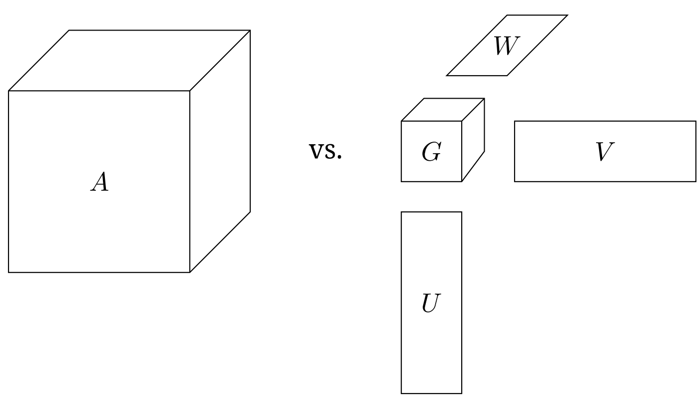
*где каждое $r_i$ является минимально возможным для $A$ значением, называется мультилинейным или таккеровским рангом. Обозначать разложение Таккера будем следующим образом:*

$$A = [\![G; U, V, W]\!]\,.$$

Дальше мы узнаем, что минимально возможные значения $r_i$ для заданного тензора $A$ достигаются одновременно, а значит данное нами определение мультилинейного ранга является корректным.

Существование разложения Таккера для любого тензора с рангами $(n_1, n_2, n_3)$ также следует из (2).

Оценим, насколько разложение Таккера экономит память для случая $n_1 = n_2 = n_3 = n$ и $r_1 = r_2 = r_3 = r$, если $r \ll n$.



Каждый из факторов Таккера займет в памяти $nr$ ячеек, ядро Таккера – еще $r^3$, и все это вместо $n^3$ для хранения изначального тензора $A$:

$$\frac{\texttt{mem}([\![G; U, V, W]\!])}{\texttt{mem}(A)} = \frac{r^3 + 3nr}{n^3} = \left(\frac{r}{n}\right)^3 + \frac{3}{r}\left(\frac{r}{n}\right)^2 \ll 1.$$

*Векторизация тензора.* Чтобы лучше понять, как устроено разложение Таккера, нам необходимо свести его к известным нам объектам, а именно к векторам и матрицам. По аналогии с матричным случаем определим векторизацию тензора как применение функции `reshape` с использованием аргумента `order='f'` (то есть по столбцам в порядке увеличения индексов слева направо). Начнем с обобщения Утверждения 2.

**Утверждение 3.** *Пусть определено* $[\![G; U, V, W]\!]$. *Тогда:*

$$\text{vec}\left([\![G; U, V, W]\!]\right) = \big(W \otimes V \otimes U\big)\,\text{vec}\,(G).$$

*Доказательство.* Утверждение доказывается аналогично Утверждению 2. Необходимо только проверить, что

$$\text{vec}\,(u \circ v \circ w) = w \otimes v \otimes u$$

и что

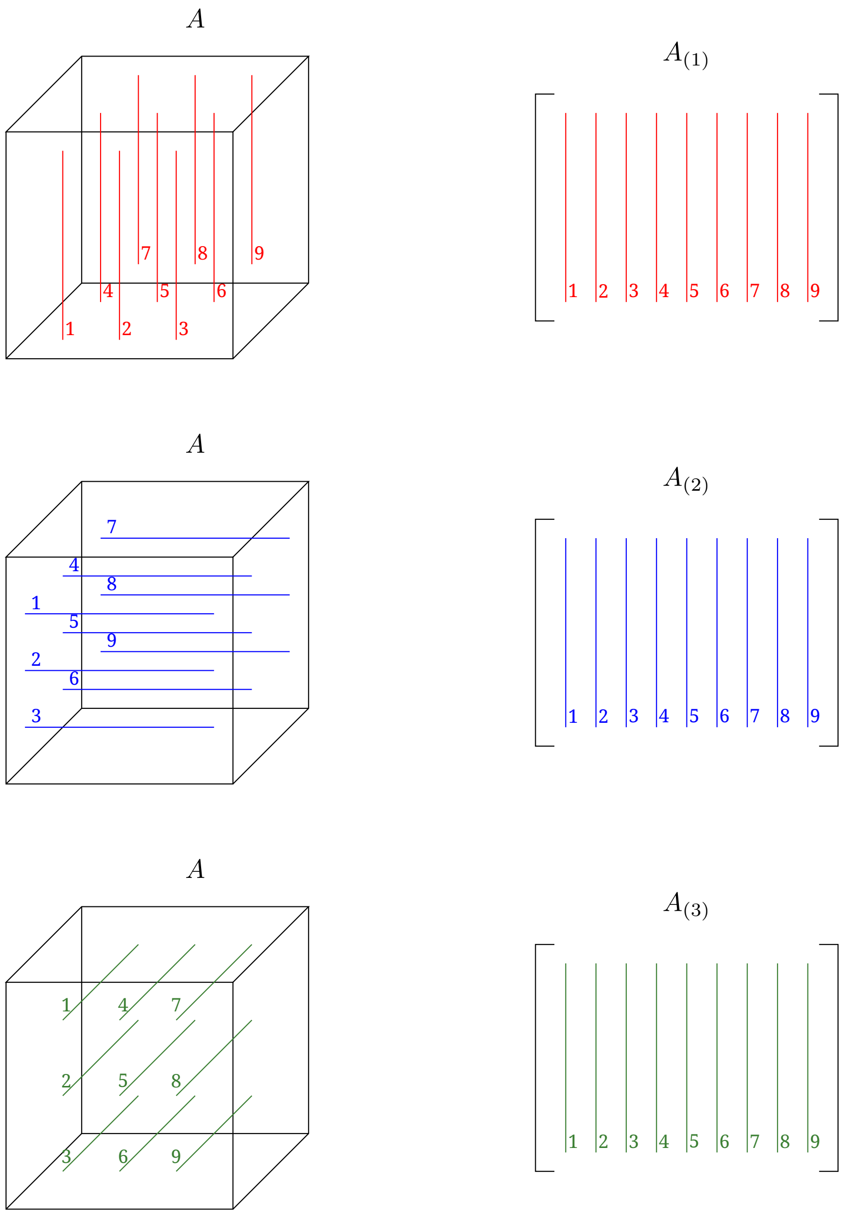
$$[\![(a \circ b \circ c); U, V, W]\!] = (Ua) \circ (Vb) \circ (Wc).$$

Далее пользуемся (2) для $G$. □

*Матрицизации тензора.* Таким образом, мы нашли, через какую матрицу соотносятся векторизации тензора $A$ и тензора $G$. Теперь попробуем понять, как связаны их матрицизации (переупорядочивание элементов в матрицу). Дадим сначала определение одного из видов превращения тензора в матрицу — так называемую развертку по $k$-й моде.

**Определение 6.** *Разверткой по $k$-моде тензора $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ назовем матрицу $A_{(k)} \in \mathbb{R}^{n_k \times (n_1 \cdot \ldots \cdot n_d)/n_k}$, столбцы которой являются $a_{i_1,\ldots,i_{k-1},:\,,i_{k+1},\ldots,i_d} \in \mathbb{R}^{n_k}$, расположенными в “фортрановском” порядке (увеличение индексов слева направо).*

Визуализируем, как выглядят развертки для трехмерного массива $A \in \mathbb{R}^{3\times3\times3}$:




Несложно также записать, как будет выглядеть развертка по $k$-й моде ($k \in \{1, 2, 3\}$) на Python:

```
order = list(range(3))
order[0], order[k-1] = order[k-1], order[0]
Ak = np.transpose(A, order)
Ak = np.reshape(Ak, (nk, (n1*n2*n3)/nk), order='f')
```

Запишем теперь формулы для матрицизаций разложения Таккера тензора $A$.

**Утверждение 4.** $A = [\![G; U, V, W]\!]$ *эквивалентно любой из:*

$$A_{(1)} = UG_{(1)}(W \otimes V)^\top,$$
$$A_{(2)} = VG_{(2)}(W \otimes U)^\top,$$
$$A_{(3)} = WG_{(3)}(V \otimes U)^\top.$$

*Доказательство.* Докажем формулу для $A_{(1)}$. Остальные доказываются аналогично. Ключевым наблюдением является тот факт, что $\text{vec}\,(A) = \text{vec}\,(A_{(1)})$. Отсюда, используя утверждения 2 и 3 получим:

$$\begin{aligned}&\text{vec}\,(A_{(1)}) = \text{vec}\,(A) = \\ &\quad = (W \otimes V \otimes U)\ \text{vec}\,(G) = (W \otimes V \otimes U)\ \text{vec}\,(G_{(1)}) = \\ &\quad = ((W \otimes V) \otimes U)\ \text{vec}\,(G_{(1)}) = \text{vec}\,(UG_{(1)}(W \otimes V)^\top),\end{aligned}$$

а значит $A_{(1)} = UG_{(1)}(W \otimes V)^\top$. □

Теперь мы готовы сформулировать теорему, которая проясняет наше определение мультилинейного ранга и объясняет, как этот ранг можно свести к хорошо знакомому нам матричному рангу.

**Теорема 1.** *Для мультилинейного ранга тензора* $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *справедливо:*

$$\textit{mlrank}(A) = \left(\text{rank}\,(A_{(1)})\,, \text{rank}\,(A_{(2)})\,, \text{rank}\,(A_{(3)})\right),$$

*причем разложение Таккера со значением* $(r_1, r_2, r_3) = \textit{mlrank}(A)$ *существует.*

*Доказательство.* Пусть дано разложение Таккера с $G \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ с некоторыми $r_1, r_2, r_3$. Заметим, что

$$A_{(1)} = \underbrace{U}_{n_1 \times r_1} \underbrace{G_{(1)}(W \otimes V)^\top}_{r_1 \times n_2 n_3},$$

то есть число $r_1$ не может быть меньше ранга $A_{(1)}$. Покажем, как уменьшить $r_1$ в точности до $\text{rank}\,(A_{(1)})$. Пусть $A_{(1)} = U_1 \Sigma_1 V_1^\top$ — компактное SVD матрицы $A_{(1)}$, то есть $U_1 \in \mathbb{R}^{n_1 \times \text{rank}(A_{(1)})}$. Тогда:

$$A_{(1)} = U_1 U_1^\top A_{(1)} = U_1 \underbrace{U_1^\top U G_{(1)}}_{\widetilde{G}_{(1)}} (W \otimes V)^\top = U_1 \widetilde{G}_{(1)} (W \otimes V)^\top,$$

что по Утверждению 4 эквивалентно

$$A = [\![\widetilde{G}; U_1, V, W]\!], \quad \widetilde{G} \in \mathbb{R}^{\text{rank}(A_{(1)}) \times r_2 \times r_3}.$$

Аналогично сводим $r_2 \to \text{rank}(A_{(2)})$ и $r_3 \to \text{rank}(A_{(3)})$. То есть мы явно построили разложение с рангами разверток. □

Таким образом, компоненты мультилинейного ранга есть размерности пространств натянутых соответственно на столбцы, строки и волокна (по 3-й размерности). Обратите внимание, что в случае $d \geq 3$ и $n_1, n_2, n_3 > 1$ эти размерности не обязаны совпадать, тогда так как при $d = 2$ нам хорошо известно, что столбцовый и строчные ранги равны друг другу.

### 2.3 HOSVD алгоритм (без доказательства)

Обсудим теперь HOSVD (higher-order SVD)[2] — один из известных алгоритмов построения малорангового приближения в виде разложения Таккера. Для этого нам понадобится ввести норму на $\mathbb{R}^{n_1 \times n_2 \times n_3}$. Простейшим выбором является обобщение понятия нормы Фробениуса с матричного случая:

[2] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000

$$\|A\|_F = \sqrt{\sum_{i,j,k} |a_{ijk}|^2} \equiv \|\text{vec}\,(A)\|_2\,.$$

Теперь сформулируем сам алгоритм.

1. С помощью SVD от $A_{(k)}$ для $k = 1, 2, 3$, найдем $U_k \in \mathbb{R}^{n_k \times r_k}$ — первые $r_k$ левых сингулярных векторов $A_{(k)}$, причем:
$$\left\|A_{(k)} - U_k U_k^\top A_{(k)}\right\|_F \leqslant \varepsilon_k \|A_{(k)}\|_F.$$

Алгоритм можно запускать как по заданным значениям $r_1, r_2, r_3$, из которых уже находим $\varepsilon_1, \varepsilon_2, \varepsilon_3$, а можно сначала зафиксировать $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$, и по ним подобрать значения $r_1, r_2, r_3$.

2. Вычислим ядро по формуле:
$$G = [\![A; U_1^\top, U_2^\top, U_3^\top]\!].$$
3. Положим $A_{\mathtt{HOSVD}} = [\![G; U_1, U_2, U_3]\!]$.

Можно показать, что в таком случае:

$$\|A - A_{\mathtt{HOSVD}}\|_F \leqslant \sqrt{\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2}\,\|A\|_F.$$

К сожалению, в отличие от двумерного случая, нам оказывается недоступным построение оптимального приближения, однако можно показать, что HOSVD алгоритм выдает квазиоптимальное приближение, а именно:

$$\|A - A_{\mathtt{HOSVD}}\|_F \leq \sqrt{3} \inf_{\text{mlrank}(B) \leq (r_1, r_2, r_3)} \|A - B\|_F.$$

Под $\text{mlrank}(B) \leq (r_1, r_2, r_3)$ подразумевается покомпонентное сравнение.

Результат также обобщается и на многомерный случай с константой квазиоптимальности $\sqrt{d}$ вместо $\sqrt{3}$.

## 3 Другие тензорные разложения

Разложения Таккера и каноническое разложения несложно записать и для размерности $d > 3$. Например, разложение Таккера тензора $A \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ будет иметь вид:

$$A = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_d=1}^{r_d} g_{\alpha_1 \ldots \alpha_d} \, u^{(1)}_{\alpha_1} \circ \cdots \circ u^{(d)}_{\alpha_d}.$$

Проблема заключается в том, что нам необходимо хранить таккеровское ядро $G \in \mathbb{R}^{r_1 \times \cdots \times r_d}$, которое имеет такую же размерность, как и исходной тензор $A$. Значит даже при $r_k \ll n_k$, оно все еще подвержено проклятью размерности и при больших $d$ оказывается непрактичным. Это мотивировало разработку и использование альтернативных тензорных разложений, для которых бы существовали надежные алгоритмы на основе сингулярного разложения, но которые при этом не были бы подвержены проклятью размерности. Например, разложение тензорного поезда[3]. Существуют разложения и более общего вида, удобно записывающиеся с помощью формализма так называемых тензорных диаграмм.

[3] Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011

Автор благодарит Анну Федорову, которая помогала с подготовкой картинок и набором части теха этой лекции.

## *Список литературы*

[1] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[2] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[3] Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[4] Е. Е. Тыртышников. *Матричный анализ и линейная алгебра*. Физматлит, 2007.

[5] G.H. Golub and C.F. Van Loan. *Matrix Computations, 4th Edition*. The Johns Hopkins University Press, Baltimore, 2013.