15

Tensors and Hypermatrices

Lek-Heng Lim
University of Chicago

Most chapters in this handbook are concerned with various aspects and implications of linearity; Chapter 14 and this chapter are unusual in that they are about multilinearity. Just as linear operators and their coordinate representations, i.e., matrices, are the main objects of interest in other chapters, tensors and their coordinate representations, i.e., hypermatrices, are the main objects of interest in this chapter. The parallel is summarized in the following schematic:

$$\begin{array}{ccccc} \textit{linearity} & \rightarrow & \textit{linear operators, bilinear forms, dyads} & \rightarrow & \textit{matrices} \\ \textit{multilinearity} & \rightarrow & \textit{tensors} & \rightarrow & \textit{hypermatrices} \end{array}$$

Chapter 14, or indeed the monographs on multilinear algebra such as [Gre78, Mar23, Nor84, Yok92], are about properties of a whole space of tensors. This chapter is about properties of a *single* tensor and its coordinate representation, a hypermatrix.

The first two sections introduce (1) a hypermatrix, (2) a tensor as an element of a tensor product of vector spaces, its coordinate representation as a hypermatrix, and a tensor as a multilinear functional. The next sections discuss the various generalizations of well-known linear algebraic and matrix theoretic notions, such as rank, norm, and determinant, to tensors and hypermatrices. The realization that these notions may be defined for order-d hypermatrices where $d > 2$ and that there are reasonably complete theories which parallel and generalize those for usual 2-dimensional matrices is a recent one. However, some of these hypermatrix notions have roots that go back as early as those for matrices. For example, the determinant of a $2 \times 2 \times 2$ hypermatrix can be found in Cayley's 1845 article [Cay45]; in fact, he studied 2-dimensional matrices and d-dimensional ones on an equal footing. The final section describes material that is omitted from this chapter for reasons of space.

In modern mathematics, there is a decided preference for coordinate-free, basis-independent ways of defining objects but we will argue here that this need not be the best strategy. The view of tensors as hypermatrices, while strictly speaking incorrect, is nonetheless a very useful device. First, it gives us a concrete way to think about tensors, one that allows a parallel to the usual matrix theory. Second, a hypermatrix is what we often get in practice: As soon as measurements are performed in some units, bases are chosen implicitly, and the values of the measurements are then recorded in the form of a hypermatrix. (There are of course good reasons not to just stick to the hypermatrix view entirely.)

We have strived to keep this chapter as elementary as possible, to show the reader that studying hypermatrices is in many instances no more difficult than studying the usual 2-dimensional matrices. Many exciting developments have to be omitted because they require too much background to describe.

Unless otherwise specified, everything discussed in this chapter applies to tensors or hypermatrices of arbitrary order $d \geq 2$ and all may be regarded as appropriate generalizations of properties of linear operators or matrices in the sense that they agree with the usual definitions when specialized to order $d = 2$. When notational simplicity is desired and when nothing essential is lost, we shall assume $d = 3$ and phrase our discussions in terms of 3-tensors. We will sometimes use the notation $\langle n \rangle := \{1, \ldots, n\}$ for any $n \in \mathbb{N}$. The bases in this chapter are always implicitly ordered according to their integer indices. All vector spaces in this chapter are finite dimensional.

We use standard notation for groups and modules. S_d is the symmetric group of permutations on d elements. An S_d-module means a $\mathbb{C}[S_d]$-module, where $\mathbb{C}[S_d]$ is the set of all formal linear combinations of elements in S_d with complex coefficients (see, e.g., [AW92]). The *general linear group* of the vector space V is the group $\mathrm{GL}(V)$ of linear isomorphisms from V onto itself with operation function composition. $\mathrm{GL}(n, F)$ is the general linear group of invertible $n \times n$ matrices over F. We will however introduce a shorthand for products of such classical groups, writing

$$\mathrm{GL}(n_1, \ldots, n_d, F) := \mathrm{GL}(n_1, F) \times \cdots \times \mathrm{GL}(n_d, F),$$

and likewise for $\mathrm{SL}(n_1, \ldots, n_d, F)$ (where $\mathrm{SL}(n, F)$ is the special linear group of $n \times n$ matrices over F having determinant one) and $\mathrm{U}(n_1, \ldots, n_d, \mathbb{C})$ (where $\mathrm{U}(n, \mathbb{C})$ is the group of $n \times n$ unitary matrices).

In this chapter, as in most other discussions of tensors in mathematics, we use $\otimes$ in multiple ways: (i) When applied to abstract vector spaces U, V, W, the notation $U \otimes V \otimes W$ is a tensor product space as defined in Section 15.2; (ii) when applied to vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ from abstract vector spaces U, V, W, the notation $\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$ is a symbol for a special element of $U \otimes V \otimes W$; (iii) when applied to l-, m-, n-tuples in F^l, F^m, F^n, it means the Segre outer product as defined in Section 15.1; (iv) when applied to F^l, F^m, F^n, $F^l \otimes F^m \otimes F^n$ means the set of all Segre outer products that can be obtained from linear combinations of terms like those in Eq. (15.3). Nonetheless, they are all consistent with each other.

15.1 Hypermatrices

What is the difference between an $m \times n$ matrix $A \in \mathbb{C}^{m \times n}$ and a mn-tuple $\mathbf{a} \in \mathbb{C}^{mn}$? The immediate difference is a superficial one: Both are lists of mn complex numbers except that we usually write A as a 2-dimensional array of numbers and $\mathbf{a}$ as a 1-dimensional array of numbers. The more important distinction comes from consideration of the natural group actions on $\mathbb{C}^{m \times n}$ and $\mathbb{C}^{mn}$. One may multiply $A \in \mathbb{C}^{m \times n}$ on "two sides" independently by an $m \times m$ matrix and an $n \times n$ matrix, whereas one may only multiply $\mathbf{a} \in \mathbb{C}^{mn}$ on "one side" by an $mn \times mn$ matrix. In algebraic parlance, $\mathbb{C}^{m \times n}$ is a $\mathbb{C}^{m \times m} \times \mathbb{C}^{n \times n}$-module whereas $\mathbb{C}^{mn}$ is a $\mathbb{C}^{mn \times mn}$-module. This extends to any order-d hypermatrices (i.e., d-dimensional matrices).

In Sections 15.3 to 15.9 we will be discussing various properties of hypermatrices and tensors. Most of these are generalizations of well-known notions for matrices or order-2 tensors. Since the multitude of indices when discussing an order-d hypermatrix can be distracting, for many of the discussions we assume that $d = 3$. The main differences between usual matrices and hypermatrices come from the transition from $d = 2$ to $d = 3$. An advantage of emphasizing 3-hypermatrices is that these may be conveniently written down on a 2-dimensional piece of paper as a list of usual matrices. This is illustrated in the examples.

Definitions:

For $n_1, \ldots, n_d \in \mathbb{N}$, a function $f : \langle n_1 \rangle \times \cdots \times \langle n_d \rangle \to F$ is a (complex) **hypermatrix**, also called an **order-d hypermatrix** or **d-hypermatrix**. We often just write $a_{k_1 \cdots k_d}$ to denote the value $f(k_1, \ldots, k_d)$ of f at $(k_1, \ldots, k_d)$ and think of f (renamed as A) as specified by a d-dimensional table of values, writing $A = [a_{k_1 \cdots k_d}]_{k_1, \ldots, k_d = 1}^{n_1, \ldots, n_d}$, or $A = [a_{k_1 \cdots k_d}]$.

The set of order-d hypermatrices (with domain $\langle n_1 \rangle \times \cdots \times \langle n_d \rangle$) is denoted by $F^{n_1 \times \cdots \times n_d}$, and we define **entrywise addition** and **scalar multiplication**: For any $[a_{k_1 \cdots k_d}], [b_{k_1 \cdots k_d}] \in F^{n_1 \times \cdots \times n_d}$ and $\gamma \in F$, $[a_{k_1 \cdots k_d}] + [b_{k_1 \cdots k_d}] := [a_{k_1 \cdots k_d} + b_{k_1 \cdots k_d}]$ and $\gamma [a_{k_1 \cdots k_d}] := [\gamma a_{k_1 \cdots k_d}]$.

The **standard basis** for $F^{n_1 \times \cdots \times n_d}$ is $\mathcal{E} := \{E_{k_1 k_2 \cdots k_d} : 1 \leq k_1 \leq n_1, \ldots, 1 \leq k_d \leq n_d\}$ where $E_{k_1 k_2 \cdots k_d}$ denotes the hypermatrix with 1 in the $(k_1, k_2, \ldots, k_d)$-coordinate and 0s everywhere else.

Let $X_1 = [x_{ij}^{(1)}] \in F^{m_1 \times n_1}, \ldots, X_d = [x_{ij}^{(d)}] \in F^{m_d \times n_d}$ and $A \in F^{n_1 \times \cdots \times n_d}$. Define **multilinear matrix multiplication** by $A' = (X_1, \ldots, X_d) \cdot A \in F^{m_1 \times \cdots \times m_d}$ where

$$a'_{j_1 \cdots j_d} = \sum\nolimits_{k_1, \ldots, k_d = 1}^{n_1, \ldots, n_d} x_{j_1 k_1}^{(1)} \cdots x_{j_d k_d}^{(d)} a_{k_1 \cdots k_d} \quad \text{for } j_1 \in \langle m_1 \rangle, \ldots, j_d \in \langle m_d \rangle. \tag{15.1}$$

For any $\pi \in S_d$, the **$\boldsymbol{\pi}$-transpose** of $A = [a_{j_1 \cdots j_d}] \in F^{n_1 \times \cdots \times n_d}$ is

$$A^\pi := [a_{j_{\pi(1)} \cdots j_{\pi(d)}}] \in F^{n_{\pi(1)} \times \cdots \times n_{\pi(d)}}. \tag{15.2}$$

If $n_1 = \cdots = n_d = n$, then a hypermatrix $A \in F^{n \times n \times \cdots \times n}$ is called **cubical** or **hypercubical** of dimension n.

A cubical hypermatrix $A = [a_{j_1 \cdots j_d}] \in F^{n \times n \times \cdots \times n}$ is said to be **symmetric** if $A^\pi = A$ for every $\pi \in S_d$ and **skew-symmetric** or **anti-symmetric** or **alternating** if $A^\pi = \operatorname{sgn}(\pi) A$ for every $\pi \in S_d$.

The **Segre outer product** of $\mathbf{a} = [a_i] \in F^\ell$, $\mathbf{b} = [b_j] \in F^m$, $\mathbf{c} = [c_k] \in F^n$ is

$$\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} := [a_i b_j c_k]_{i,j,k=1}^{\ell, m, n} \in F^{\ell \times m \times n}. \tag{15.3}$$

Let $A \in F^{n_1 \times \cdots \times n_d}$ and $B \in F^{m_1 \times \cdots \times m_e}$ be hypermatrices of orders d and e, respectively. Then the **outer product** of A and B is a hypermatrix C of order $d + e$ denoted

$$C = A \otimes B \in F^{n_1 \times \cdots \times n_d \times m_1 \times \cdots \times m_e} \tag{15.4}$$

with its $(i_1, \ldots, i_d, j_1 \ldots, j_e)$-entry given by

$$c_{i_1 \cdots i_d j_1 \cdots j_e} = a_{i_1 \cdots i_d} b_{j_1 \cdots j_e} \tag{15.5}$$

for all $i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle$ and $j_1 \in \langle m_1 \rangle, \ldots, j_e \in \langle m_e \rangle$.

Suppose $A \in F^{n_1 \times \cdots \times n_{d-1} \times n}$ and $B \in F^{n \times m_2 \times \cdots \times m_e}$ are an order-d and an order-e hypermatrix, respectively, where the last index i_d of A and the first index j_1 of B run over the same range, i.e., $i_d \in \langle n \rangle$ and $j_1 \in \langle n \rangle$. The **contraction product** of A and B is an order-$(d + e - 2)$ hypermatrix $C \in F^{n_1 \times \cdots \times n_{d-1} \times m_2 \times \cdots \times m_e}$ whose entries are

$$c_{i_1 \cdots i_{d-1} j_2 \cdots j_e} = \sum\nolimits_{k=1}^{n} a_{i_1 \cdots i_{d-1} k} b_{k j_2 \cdots j_e}, \tag{15.6}$$

for $i_1 \in \langle n_1 \rangle, \ldots, i_{d-1} \in \langle n_{d-1} \rangle$ and $j_2 \in \langle m_2 \rangle, \ldots, j_e \in \langle m_e \rangle$.

In general if $A \in F^{n_1\times\cdots\times n_d}$ and $B \in F^{m_1\times\cdots\times m_e}$ have an index with a common range, say, $n_p = m_q = n$, then $C \in F^{n_1\times\cdots\times\widehat{n}_p\times\cdots\times n_d\times m_1\times\cdots\times\widehat{m}_q\times\cdots\times m_e}$ is the hypermatrix with entries

$$c_{i_1\cdots\widehat{i}_p\cdots i_d j_1\cdots\widehat{j}_q\cdots j_e} = \sum\nolimits_{k=1}^{n} a_{i_1\cdots k\cdots i_d} b_{j_1\cdots k\cdots j_e}. \tag{15.7}$$

where by convention a caret over any entry means that the respective entry is omitted (e.g., $a_{i\widehat{j}k} = a_{ik}$ and $F^{\widehat{l}\times m\times n} = F^{m\times n}$).

Contractions are not restricted to one pair of indices at a time. For hypermatrices A and B, the hypermatrix

$$\langle A, B\rangle_{\alpha:\lambda,\beta:\mu,\ldots,\gamma:\nu}$$

is the hypermatrix obtained from contracting the αth index of A with the λth index of B, the βth index of A with the μth index of $B, \ldots,$ the γth index of A with the νth index of B (assuming that the indices that are contracted run over the same range and in the same order).

Facts:

Facts requiring proof for which no specific reference is given can be found in [Lim] and the references therein.

1. $F^{n_1\times\cdots\times n_d}$ with entrywise addition and scalar multiplication is a vector space.
2. The standard basis $\mathcal{E}$ is a basis for $F^{n_1\times n_2\times\cdots\times n_d}$, and $\dim F^{n_1\times n_2\times\cdots\times n_d} = n_1 n_2\cdots n_d$.
3. The elements of the standard basis of $F^{n_1\times\cdots\times n_d}$ may be written as

$$E_{k_1k_2\cdots k_d} = \mathbf{e}_{k_1}\otimes\mathbf{e}_{k_2}\otimes\cdots\otimes\mathbf{e}_{k_d},$$

   using the Segre outer product (15.3).
4. Let $A \in F^{n_1\times\cdots\times n_d}$ and $X_k \in F^{l_k\times m_k}$, $Y_k \in F^{m_k\times n_k}$ for $k = 1,\ldots,d$. Then

$$(X_1,\ldots,X_d)\cdot[(Y_1,\ldots,Y_d)\cdot A] = (X_1Y_1,\ldots,X_dY_d)\cdot A.$$

5. Let $A, B \in F^{n_1\times\cdots\times n_d}$, $\alpha,\beta\in F$, and $X_k \in F^{m_k\times n_k}$ for $k = 1,\ldots,d$. Then

$$(X_1,\ldots,X_d)\cdot[\alpha A + \beta B] = \alpha(X_1,\ldots,X_d)\cdot A + \beta(X_1,\ldots,X_d)\cdot B.$$

6. Let $A \in F^{n_1\times\cdots\times n_d}$, $\alpha,\beta\in F$, and $X_k, Y_k \in F^{m_k\times n_k}$ for $k = 1,\ldots,d$. Then

$$[\alpha(X_1,\ldots,X_d) + \beta(Y_1,\ldots,Y_d)]\cdot A = \alpha(X_1,\ldots,X_d)\cdot A + \beta(Y_1,\ldots,Y_d)\cdot A.$$

7. The Segre outer product interacts with multilinear matrix multiplication in the following manner

$$(X_1,\ldots,X_d)\cdot\left[\sum\nolimits_{p=1}^{r}\beta_p\mathbf{v}_p^{(1)}\otimes\cdots\otimes\mathbf{v}_p^{(d)}\right] = \sum\nolimits_{p=1}^{r}\beta_p(X_1\mathbf{v}_p^{(1)})\otimes\cdots\otimes(X_d\mathbf{v}_p^{(d)}).$$

8. For cubical hypermatrices, each $\pi\in S_d$ defines a linear operator

$$\pi : F^{n\times\cdots\times n}\to F^{n\times\cdots\times n},\ \pi(A) = A^{\pi}.$$

9. Any matrix $A \in F^{n\times n}$ may be written as a sum of a symmetric and a skew symmetric matrix, $A = \frac{1}{2}(A+A^T)+\frac{1}{2}(A-A^T)$. This is *not* true for hypermatrices of order $d > 2$. For example, for a 3-hypermatrix $A \in F^{n\times n\times n}$, the equivalent of the decomposition is

$$A = \frac{1}{6}(A+A^{(1,2,3)}+A^{(1,3,2)}+A^{(1,2)}+A^{(1,3)}+A^{(2,3)})+\frac{1}{3}(A+A^{(1,2)}-A^{(1,3)}-A^{(1,2,3)})$$
$$+\frac{1}{3}(A+A^{(1,3)}-A^{(1,2)}-A^{(1,3,2)})+\frac{1}{6}(A+A^{(1,2,3)}+A^{(1,3,2)}-A^{(1,2)}-A^{(1,3)}-A^{(2,3)}),$$

   where $S_3 = \{1,(1,2),(1,3),(2,3),(1,2,3),(1,3,2)\}$ (a 3-hypermatrix has five different "transposes" in addition to the original).

10. When $d = e = 2$, the contraction of two matrices is simply the matrix-matrix product: If $A \in F^{m\times n}$ and $B \in F^{n\times p}$, then $C \in F^{m\times p}$ is given by $C = AB$, $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$.
11. When $d = e = 1$, the contraction of two vectors $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^n$ is the scalar $c \in \mathbb{R}$ given by the Euclidean inner product $c = \sum_{k=1}^{n} a_k b_k$ (for complex vectors, we contract a with $\bar{b}$ to get the Hermitian inner product).
12. The contraction product respects expressions of hypermatrices written as a sum of decomposable hypermatrices. For example, if

$$A = \sum_{i=1}^{r} \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i \in F^{l\times m\times n}, \quad B = \sum_{j,k=1}^{s,t} \mathbf{w}_k \otimes \mathbf{x}_j \otimes \mathbf{y}_j \otimes \mathbf{z}_k \in F^{n\times p\times m\times q},$$

then

$$\Big\langle \sum_{i=1}^{r} \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i, \sum_{j,k=1}^{s,t} \mathbf{w}_k \otimes \mathbf{x}_j \otimes \mathbf{y}_j \otimes \mathbf{z}_k \Big\rangle_{2:3,3:1} = \sum_{i,j,k=1}^{r,s,t} \langle \mathbf{b}_i, \mathbf{y}_j\rangle\langle \mathbf{c}_i, \mathbf{w}_k\rangle \mathbf{a}_i \otimes \mathbf{x}_j \otimes \mathbf{z}_k \in F^{l\times p\times q},$$

where $\langle \mathbf{b}_i, \mathbf{y}_j\rangle = \mathbf{b}_i^T\mathbf{y}_j$ and $\langle \mathbf{c}_i, \mathbf{w}_k\rangle = \mathbf{c}_i^T\mathbf{w}_k$ as usual.
13. Multilinear matrix multiplication may also be expressed as the contraction of $A \in F^{n_1\times\cdots\times n_d}$ with matrices $X_1 \in F^{m_1\times n_1}, \dots, X_d \in F^{m_d\times n_d}$. Take $d = 3$ for example; $(X, Y, Z)\cdot A$ for $A \in F^{l\times m\times n}$ and $X \in F^{p\times l}, Y \in F^{q\times m}, Z \in F^{r\times n}$ is

$$(X, Y, Z)\cdot A = \langle X, \langle Y, \langle Z, A\rangle_{2:3}\rangle_{2:2}\rangle_{2:1}.$$

Note that the order of contractions does not matter, i.e.,

$$\langle X, \langle Y, \langle Z, A\rangle_{2:3}\rangle_{2:2}\rangle_{2:1} = \langle Y, \langle X, \langle Z, A\rangle_{2:3}\rangle_{2:1}\rangle_{2:2} = \cdots = \langle Z, \langle Y, \langle X, A\rangle_{2:1}\rangle_{2:2}\rangle_{2:3}.$$

14. For the special case where we contract two hypermatrices $A, B \in \mathbb{C}^{n_1\times\cdots\times n_d}$ in all indices to get a scalar in $\mathbb{C}$, we shall drop all indices and denote it by

$$\langle A, B\rangle = \sum_{j_1,\dots,j_d=1}^{n_1,\dots,n_d} a_{j_1\cdots j_d} b_{j_1\cdots j_d}.$$

If we replace B by its complex conjugate, this gives the usual Hermitian inner product.

Examples:

1. A 3-hypermatrix $A \in \mathbb{C}^{l\times m\times n}$ has "three sides" and may be multiplied by three matrices $X \in \mathbb{C}^{p\times l}$, $Y \in \mathbb{C}^{q\times m}$, $Z \in \mathbb{C}^{r\times n}$. This yields another 3-hypermatrix $A' \in \mathbb{C}^{p\times q\times r}$ where

$$A' = (X, Y, Z)\cdot A \in \mathbb{C}^{p\times q\times r}, \quad a'_{\alpha\beta\gamma} = \sum_{i,j,k=1}^{l,m,n} x_{\alpha i} y_{\beta j} z_{\gamma k} a_{ijk}.$$

2. A 3-hypermatrix may be conveniently written down on a (2-dimensional) piece of paper as a list of usual matrices, called *slices*. For example $A = [a_{ijk}]_{i,j,k=1}^{4,3,2} \in \mathbb{C}^{4\times 3\times 2}$ can be written down as two "slices" of 4×3 matrices

$$A = \left[\begin{array}{ccc|ccc} a_{111} & a_{121} & a_{131} & a_{112} & a_{122} & a_{132} \\ a_{211} & a_{221} & a_{231} & a_{212} & a_{222} & a_{232} \\ a_{311} & a_{321} & a_{331} & a_{312} & a_{322} & a_{332} \\ a_{411} & a_{421} & a_{431} & a_{412} & a_{422} & a_{432} \end{array}\right] \in \mathbb{C}^{4\times 3\times 2}$$

where i, j, k index the row, column, and slice, respectively.

3. More generally a 3-hypermatrix $A \in \mathbb{C}^{l\times m\times n}$ can be written down as n slices of $l \times m$ matrices $A_k \in \mathbb{C}^{l\times m}$, $k = 1,\ldots,n$, denoted
$$A = [A_1 \mid A_2 \mid \cdots \mid A_n] \in \mathbb{C}^{l\times m\times n}.$$
If $A = [a_{ijk}]_{i,j,k=1}^{l,m,n}$, then $A_k = [a_{ijk}]_{i,j=1}^{l,m}$.
4. A related alternative way is to introduce indeterminates $x_1,\ldots,x_n$ and represent $A \in \mathbb{C}^{l\times m\times n}$ as a matrix whose entries are linear polynomials in $x_1,\ldots,x_n$:
$$x_1A_1 + x_2A_2 + \cdots + x_nA_n \in \mathbb{C}[x_1,\ldots,x_n]^{l\times m}.$$
Clearly, we have a one-to-one correspondence between $l \times m \times n$ hypermatrices in $\mathbb{C}^{l\times m\times n}$ and $l \times m$ matrices in $\mathbb{C}[x_1,\ldots,x_n]^{l\times m}$.
5. Just like a matrix can be sliced up into rows or columns, we may of course also slice up a 3-hypermatrix in two other ways: as l slices of $m \times n$ matrices or m slices of $l \times n$ matrices. To avoid notational clutter, we shall not introduce additional notations for these but simply note that these correspond to looking at the slices of the π-tranposes of A (just like the rows of a matrix $A \in \mathbb{C}^{m\times n}$ are the columns of its transpose $A^T \in \mathbb{C}^{n\times m}$).

Applications:

1. [Bax78, Yan67] In statistical mechanics, the Yang–Baxter equation is given by
$$\sum_{\alpha,\beta,\gamma=1}^{N} R_{\ell\gamma i\alpha}R_{\alpha\beta jk}R_{mn\gamma\beta} = \sum_{\alpha,\beta,\gamma=1}^{N} R_{\alpha\beta jk}R_{\ell m\alpha\gamma}R_{\gamma n\beta k}$$
where $i,j,k,\ell,m,n = 1,\ldots,N$. This may be written in terms of contractions of hypermatrices. Let $R = (R_{ijkl}) \in \mathbb{C}^{N\times N\times N\times N}$, then we have
$$\langle\langle R,R\rangle_{4:1}, R\rangle_{2:3,4:4} = \langle R, \langle R,R\rangle_{4:1}\rangle_{1:3,2:4}.$$
2. Hooke's law in one spatial dimension, with $x =$ extension, $F =$ force, $c =$ the spring constant, is $F = -cx$. Hooke's law in three spatial dimensions is given by the linear elasticity equation:
$$\sigma_{ij} = \sum_{k,l=1}^{3} c_{ijkl}\gamma_{kl}.$$
where $\mathbf{x} = [x_1,x_2,x_3]$, $C = [c_{ijkl}] \in \mathbb{R}^{3\times3\times3\times3}$ is the elasticity tensor (also called stiffness tensor), $\Sigma \in \mathbb{R}^{3\times3}$ is the stress tensor, and $\Gamma \in \mathbb{R}^{3\times3}$ is strain tensor. Hooke's law may be expressed in terms of contraction product as
$$\Sigma = \langle C,\Gamma\rangle_{3:1,4:2}.$$
3. The observant reader might have noted that the word "tensor" was used to denote a tensor of order 2. The stress and strain tensors are all of order 2. This is in fact the most common use of the term "tensors" in physics, where order-2 tensors occur a lot more frequently than those of higher orders. There are authors (cf. [Bor90], for example) who use the term "tensor" to mean exclusively a tensor of order 2.

15.2 Tensors and Multilinear Functionals

There is a trend in modern mathematics where instead of defining a mathematical entity (like a tensor) directly, one first defines a whole space of such entities (like a space of tensors) and subsequently defines the entity as an element of this space. For example, a succinct answer to the question "What is a vector?" is, "It is an element of a vector space" [Hal85, p. 153]. One advantage of such an approach is that it allows us to examine the entity in the appropriate context. Depending on the context, a matrix can be an element of a vector space $F^{n\times n}$, of a ring, e.g., the endomorphism ring $L(F^n)$ of F^n, of a Lie algebra, e.g., $\mathfrak{gl}(n,F)$, etc. Depending on what properties of the matrix one is interested in studying, one chooses the space it lives in accordingly.

The same philosophy applies to tensors, where one first defines a tensor product of d vector spaces $V_1\otimes\cdots\otimes V_d$ and then subsequently defines an order-d tensor as an element of such a tensor product space. Since a tensor product space is defined via a universal factorization property (the definition used in Section 14.2), it can be interpreted in multiple ways, such as a quotient module (the equivalent definition used here) or a space of multilinear functionals. We will regard tensors as multilinear functionals. A perhaps unconventional aspect of our approach is that for clarity we isolate the notion of covariance and contravariance (see Section 15.10) from our definition of a tensor. We do not view this as an essential part of the definition but a source of obfuscation.

Tensors can also be represented as hypermatrices by choosing a basis. Given a set of bases, the essential information about a tensor T is captured by the coordinates $a_{j_1\cdots j_d}$'s (cf. Fact 3 below). We may view the coefficient $a_{j_1\cdots j_d}$ as the $(j_1,\dots,j_d)$-entry of the d-dimensional matrix $A=[a_{j_1\cdots j_d}]\in F^{n_1\times\cdots\times n_d}$, where A is a coordinate representation of T with respect to the specified bases.

See also Section 14.2 for more information on tensors and tensor product spaces.

Definitions:

Let F be a field and let $V_1,\dots,V_d$ be F-vector spaces.

The **tensor product space** $V_1\otimes\cdots\otimes V_d$ is the quotient module $F(V_1,\dots,V_d)/R$ where $F(V_1,\dots,V_d)$ is the free module generated by all n-tuples $(\mathbf{v}_1,\dots,\mathbf{v}_d)\in V_1\times\cdots\times V_d$ and R is the submodule of $F(V_1,\dots,V_d)$ generated by elements of the form

$$(\mathbf{v}_1,\dots,\alpha\mathbf{v}_k+\beta\mathbf{v}'_k,\dots,\mathbf{v}_d)-\alpha(\mathbf{v}_1,\dots,\mathbf{v}_k,\dots,\mathbf{v}_d)-\beta(\mathbf{v}_1,\dots,\mathbf{v}'_k,\dots,\mathbf{v}_d)$$

for all $\mathbf{v}_k,\mathbf{v}'_k\in V_k$, $\alpha,\beta\in F$, and $k\in\{1,\dots,d\}$. We write $\mathbf{v}_1\otimes\cdots\otimes\mathbf{v}_d$ for the element $(\mathbf{v}_1,\dots,\mathbf{v}_d)+R$ in the quotient space F/R.

An element of $V_1\otimes\cdots\otimes V_d$ that can be expressed in the form $\mathbf{v}_1\otimes\cdots\otimes\mathbf{v}_d$ is called **decomposable**. The symbol $\otimes$ is called the **tensor product** when applied to vectors from abstract vector spaces.

The elements of $V_1\otimes\cdots\otimes V_d$ are called **order-d tensors** or **d-tensors** and $n_k=\dim V_k$, $k=1,\dots,d$ are the **dimensions** of the tensors.

Let $\mathcal{B}_k=\{\mathbf{b}_1^{(k)},\dots,\mathbf{b}_{n_k}^{(k)}\}$ be a basis for $V_k, k=1,\dots,d$. For a tensor $T\in V_1\otimes\cdots\otimes V_d$, the **coordinate representation** of T with respect to the specified bases is $[T]_{\mathcal{B}_1,\dots,\mathcal{B}_d}=[a_{j_1\cdots j_d}]$. where

$$T=\sum\nolimits_{j_1,\dots,j_d=1}^{n_1,\dots,n_d}a_{j_1\cdots j_d}\mathbf{b}_{j_1}^{(1)}\otimes\cdots\otimes\mathbf{b}_{j_d}^{(d)}. \tag{15.8}$$

The special case where $V_1=\cdots=V_d=V$ is denoted $\mathsf{T}^d(V)$ or $V^{\otimes d}$, i.e., $\mathsf{T}^d(V)=V\otimes\cdots\otimes V$.

For any $\pi\in S_d$, the action of π on $\mathsf{T}^d(V)$ is defined by

$$\pi(\mathbf{v}_1\otimes\cdots\otimes\mathbf{v}_d):=\mathbf{v}_{\pi(1)}\otimes\cdots\otimes\mathbf{v}_{\pi(d)} \tag{15.9}$$

for decomposable elements, and then extended linearly to all elements of $\mathsf{T}^d(V)$.

A tensor $T\in\mathsf{T}^d(V)$ is **symmetric** if $\pi(T)=T$ for all $\pi\in S_d$ and is **alternating** if $\pi(T)=\operatorname{sgn}(\pi)T$ for all $\pi\in S_d$.

For a vector space V, V^* denotes the **dual space** of linear functionals of V (cf. Section 3.6).

A multilinear functional on $V_1,\dots,V_d$ is a function $T: V_1\times V_2\times\cdots\times V_d\to F$, i.e., for $\alpha,\beta\in F$,

$$T(\mathbf{x}_1, \ldots, \alpha \mathbf{y}_k + \beta \mathbf{z}_k, \ldots, \mathbf{x}_d) = \alpha T(\mathbf{x}_1, \ldots, \mathbf{y}_k, \ldots, \mathbf{x}_d) + \beta T(\mathbf{x}_1, \ldots, \mathbf{z}_k, \ldots, \mathbf{x}_d), \tag{15.10}$$

for every $k = 1, \ldots, d$.

The vector space of (F-valued) multilinear functionals on $V_1, \ldots, V_d$ is denoted by $L(V_1, \ldots, V_d; F)$.

A multilinear functional $T \in L(V_1, \ldots, V_d; F)$ is **decomposable** if $T = \theta_1 \cdots \theta_d$ where $\theta_i \in V_i^*$ and $\theta_1 \cdots \theta_d(\mathbf{v}_1, \ldots, \mathbf{v}_d) = \theta_1(\mathbf{v}_1) \cdots \theta_d(\mathbf{v}_d)$.

For any $\pi \in S_d$, the action of π on $L(V, \ldots, V, F)$ is defined by

$$\pi(T)(\mathbf{v}_1, \ldots, \mathbf{v}_d) = T(\mathbf{v}_{\pi(1)}, \ldots, \mathbf{v}_{\pi(d)}). \tag{15.11}$$

A multilinear functional $T \in L(V_1, \ldots, V_d; F)$ is **symmetric** if $\pi(T) = T$ for all $\pi \in S_d$ and is **alternating** if $\pi(T) = \operatorname{sgn}(\pi) T$ for all $\pi \in S_d$.

Facts:

Facts requiring proof for which no specific reference is given can be found in [Bou98, Chap. II], [KM97, Chap. 4], [Lan02, Chap. XVI], and [Yok92, Chap. 1]. Additional facts about tensors can be found in Section 14.2.

1. The tensor product space $V_1 \otimes \cdots \otimes V_d$ with $\nu : V_1 \times \cdots \times V_m \to V_1 \otimes \cdots \otimes V_d$ defined by
$$\nu(\mathbf{v}_1, \ldots, \mathbf{v}_d) = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_d = (\mathbf{v}_1, \ldots, \mathbf{v}_d) + R \in F(V_1, \ldots, V_d)/R$$
and extended linearly satisfies the *Universal Factorization Property* that can be used to define tensor product spaces (cf. Section 14.2), namely:

   If φ is a multilinear map from $V_1 \times \cdots \times V_d$ into the vector space U, then there exists a unique linear map ψ from $V_1 \otimes \cdots \otimes V_d$ into U, that makes the following diagram commutative:

$$\begin{tikzcd} V_1 \times \cdots \times V_d \arrow[r, "\nu"] \arrow[dr, "\varphi"'] & V_1 \otimes \cdots \otimes V_d \arrow[d, "\psi"] \\ & U \end{tikzcd}$$

   i.e., $\psi\nu = \varphi$.

2. If $U = F^l$, $V = F^m$, $W = F^n$, we may identify
$$F^l \otimes F^m \otimes F^n = F^{l \times m \times n}$$
through the interpretation of the tensor product of vectors as a hypermatrix via the Segre outer product (cf. Eq. (15.3)),
$$[a_1, \ldots, a_l]^T \otimes [b_1, \ldots, b_m]^T \otimes [c_1, \ldots, c_n]^T = [a_i b_j c_k]_{i,j,k=1}^{l,m,n}.$$
This is a model of the universal definition of $\otimes$ given in Section 14.2.
3. Given bases $\mathcal{B}_k = \{\mathbf{b}_1^{(k)}, \ldots, \mathbf{b}_{n_k}^{(k)}\}$ for $V_k, k = 1, \ldots, d$, any tensor T in $V_1 \otimes \cdots \otimes V_d$, can be expressed as a linear combination
$$T = \sum\nolimits_{j_1, \ldots, j_d = 1}^{n_1, \ldots, n_d} a_{j_1 \cdots j_d} \mathbf{b}_{j_1}^{(1)} \otimes \cdots \otimes \mathbf{b}_{j_d}^{(d)}.$$
In older literature, the $a_{j_1 \cdots j_d}$'s are often called the *components* of T.
4. One loses information when going from the tensor to its hypermatrix representation, in the sense that the bases $\mathcal{B}_1, \ldots, \mathcal{B}_d$ must be specified in addition to the hypermatrix A in order to recover the tensor T.

5. Every choice of bases on $V_1, \ldots, V_d$ gives a (usually) different hypermatrix representation of the same tensor in $V_1 \otimes \cdots \otimes V_d$.
6. Given two sets of bases $\mathcal{B}_1, \ldots, \mathcal{B}_d$ and $\mathcal{B}'_1, \ldots, \mathcal{B}'_d$ for $V_1, \ldots, V_d$, the same tensor T has two coordinate representations as a hypermatrix,

$$A = [T]_{\mathcal{B}_1,\ldots,\mathcal{B}_d} \qquad \text{and} \qquad A' = [T]_{\mathcal{B}'_1,\ldots,\mathcal{B}'_d}$$

where $A = [a_{k_1\cdots k_d}], A' = [a'_{k_1\cdots k_d}] \in F^{n_1\times\cdots\times n_d}$. The relationship between A and A' is given by the multilinear matrix multiplication

$$A' = (X_1, \ldots, X_d) \cdot A$$

where $X_k \in \mathrm{GL}(n_k, F)$ is the change-of-basis matrix transforming $\mathcal{B}'_k$ to $\mathcal{B}_k$ for $k = 1, \ldots, d$. We shall call this the *change-of-basis* rule. Explicitly, the entries of A' and A are related by

$$a'_{j_1\cdots j_d} = \sum\nolimits_{k_1,\ldots,k_d=1}^{n_1,\ldots,n_d} x_{j_1k_1}\cdots x_{j_dk_d} a_{k_1\cdots k_d} \quad \text{for } j_1 \in \langle n_1\rangle, \ldots, j_d \in \langle n_d\rangle,$$

where $X_1 = [x_{j_1k_1}] \in \mathrm{GL}(n_1, F), \ldots, X_d = [x_{j_dk_d}] \in \mathrm{GL}(n_d, F)$.
7. Cubical hypermatrices arise from a natural coordinate representation of tensors $T \in \mathsf{T}^d(V)$, i.e.,

$$T : V \times \cdots \times V \to F$$

where by "natural" we mean that we make the same choice of basis $\mathcal{B}$ for every copy of V, i.e.,

$$[T]_{\mathcal{B},\ldots,\mathcal{B}} = A.$$

8. $\mathsf{T}^d(V)$ is an S_d-module.
9. Symmetric and alternating hypermatrices are natural coordinate representations of symmetric and alternating tensors.
10. For finite-dimensional vector spaces $V_1, \ldots, V_d$, the space $L(V_1^*, \ldots, V_d^*; F)$ of multilinear functionals on the dual spaces V_i^* is naturally isomorphic to the tensor product space $V_1 \otimes \cdots \otimes V_d$, with $\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_d \leftrightarrow \hat{\mathbf{v}}_1 \cdots \hat{\mathbf{v}}_d$ extended by linearity, where $\hat{\mathbf{v}} \in V^{**}$ is defined by $\hat{\mathbf{v}}(f) = f(\mathbf{v})$ for $f \in V^*$. Since V is naturally isomorphic to V^{**} via $\mathbf{v} \leftrightarrow \hat{\mathbf{v}}$ (cf. Section 3.6), every vector space of multilinear functionals is a tensor product space (of the dual spaces), and a multilinear functional is a tensor: A decomposable multilinear functional $T = \theta_1 \cdots \theta_d \in L(V_1, \ldots, V_d; F)$ with $\theta_i \in V_i^*$ is naturally associated with $\theta_1 \otimes \cdots \otimes \theta_d \in V_1^* \otimes \cdots \otimes V_d^*$, and this relationship is extended by linearity.
11. A multilinear functional is decomposable as a multilinear functional if and only if it is a decomposable tensor.
12. Let $\mathcal{B}_k = \{\mathbf{b}_1^{(k)}, \ldots, \mathbf{b}_{n_k}^{(k)}\}$ be a basis for the vector space V_k for $k = 1, \ldots, d$, so $V_k \cong F^{n_k}$ where $n_k = \dim V_k$, with the isomorphism $\mathbf{x} \mapsto [\mathbf{x}]$ where $[\mathbf{x}]$ denotes the coordinate vector with respect to basis $\mathcal{B}_k$. For a multilinear functional $T : V_1 \times \cdots \times V_d \to F$, define $a_{j_1\cdots j_d} := T(\mathbf{b}_{j_1}^{(1)}, \ldots, \mathbf{b}_{j_d}^{(d)})$ for $j_i \in \langle n_i\rangle$. Then T has the explicit formula

$$T(\mathbf{x}_1, \ldots, \mathbf{x}_d) = \sum\nolimits_{j_1,\ldots,j_d=1}^{n_1,\ldots,n_d} a_{j_1\cdots j_d} x_{j_1}^{(1)} \cdots x_{j_d}^{(d)},$$

in terms of the coordinates of the coordinate vector $[\mathbf{x}_k] = [x_1^{(k)}, \ldots, x_{n_k}^{(k)}]^T \in F^{n_k}$ for $k = 1, \ldots, d$. In older literature, the $a_{j_1\cdots j_d}$'s are also often called the *components* of T as in Fact 3.

13. Let $\mathcal{B}_k = \{\mathbf{b}_1^{(k)}, \dots, \mathbf{b}_{n_k}^{(k)}\}$ be a basis for V_k for $k = 1, \dots, d$.

   (a) For each $i_1 \in \langle n_1 \rangle, \dots, i_d \in \langle n_d \rangle$, define the multilinear functional $\varphi_{i_1 \cdots i_d} : V_1 \times \cdots \times V_d \to F$ by
   $$\varphi_{i_1 \cdots i_d}(\mathbf{b}_{j_1}^{(1)}, \dots, \mathbf{b}_{j_d}^{(d)}) = \begin{cases} 1 & \text{if } i_1 = j_1, \dots, i_d = j_d, \\ 0 & \text{otherwise,} \end{cases}$$
   and extend the definition to all of $V_1 \times \cdots \times V_d$ via (15.10). The set
   $$\mathcal{B}^* := \{\varphi_{i_1 \cdots i_d} : i_1 \in \langle n_1 \rangle, \dots, i_d \in \langle n_d \rangle\}$$
   is a basis for $L(V_1, \dots, V_d; F)$.

   (b) The set
   $$\mathcal{B} := \left\{ \mathbf{b}_{j_1}^{(1)} \otimes \cdots \otimes \mathbf{b}_{j_d}^{(d)} : j_1 \in \langle n_1 \rangle, \dots, j_d \in \langle n_d \rangle \right\}$$
   is a basis for $V_1 \otimes \cdots \otimes V_d$.

   (c) For a multilinear functional $T : V_1 \times \cdots \times V_d \to F$ with $a_{j_1 \cdots j_d}$ as defined in Fact 12, $T = \sum_{j_1,\dots,j_d=1}^{n_1,\dots,n_d} a_{j_1 \cdots j_d} \varphi_{j_1 \cdots j_d}$.

   (d) $\dim V_1 \otimes \cdots \otimes V_d = \dim V_1 \cdots \dim V_d$ since $|\mathcal{B}^*| = |\mathcal{B}| = n_1 \cdots n_d$.

14. $\mathsf{T}^d(V)$ is an $\operatorname{End}(V)$-module (where $\operatorname{End}(V)$ is the algebra of linear operators on V) with the natural action defined on decomposable elements via
   $$g(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_d) = g(\mathbf{v}_1) \otimes \cdots \otimes g(\mathbf{v}_d)$$
   for any $g \in \operatorname{End}(V)$ and then extended linearly to all of $\mathsf{T}^d(V)$.

Examples:

For notational convenience, let $d = 3$.

1. Explicitly, the definition of a tensor product space above simply means that
   $$U \otimes V \otimes W := \left\{ \sum\nolimits_{i=1}^{n} \alpha_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i : \mathbf{u}_i \in U, \mathbf{v}_i \in V, \mathbf{w}_i \in W, n \in \mathbb{N} \right\}$$
   where $\otimes$ satisfies
   $$\begin{aligned} (\alpha \mathbf{v}_1 + \beta \mathbf{v}_1') \otimes \mathbf{v}_2 \otimes \mathbf{v}_3 &= \alpha \mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \mathbf{v}_3 + \beta \mathbf{v}_1' \otimes \mathbf{v}_2 \otimes \mathbf{v}_3, \\ \mathbf{v}_1 \otimes (\alpha \mathbf{v}_2 + \beta \mathbf{v}_2') \otimes \mathbf{v}_3 &= \alpha \mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \mathbf{v}_3 + \beta \mathbf{v}_1 \otimes \mathbf{v}_2' \otimes \mathbf{v}_3, \\ \mathbf{v}_1 \otimes \mathbf{v}_2 \otimes (\alpha \mathbf{v}_3 + \beta \mathbf{v}_3') &= \alpha \mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \mathbf{v}_3 + \beta \mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \mathbf{v}_3'. \end{aligned}$$
   The "modulo relation" simply means that $\otimes$ obeys these rules. The statement that $\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$ are generators of $U \otimes V \otimes W$ simply means that $U \otimes V \otimes W$ is the set of all possible linear combinations of the form $\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$ where $\mathbf{u} \in U$, $\mathbf{v} \in V$, $\mathbf{w} \in W$.
2. We emphasize here that a tensor and a hypermatrix are quite different. To specify a tensor $T \in V_1 \otimes \cdots \otimes V_d$, we need both the hypermatrix $[T]_{\mathcal{B}_1,\dots,\mathcal{B}_d} \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ and the bases $\mathcal{B}_1, \dots, \mathcal{B}_d$ that we chose for $V_1, \dots, V_d$.
3. Each hypermatrix in $F^{n_1 \times \cdots \times n_d}$ has a unique, natural tensor associated with it: the tensor in the standard basis of $F^{n_1} \otimes \cdots \otimes F^{n_d}$. (The same is true for matrices and linear operators.)

Applications:

1. In physics parlance, a decomposable tensor represents *factorizable* or *pure states*. In general, a tensor in $U \otimes V \otimes W$ will not be decomposable.

2. [Cor84] In the standard model of particle physics, a proton is made up of two up quarks and one down quark. A more precise statement is that the state of a proton is a 3-tensor

$$\frac{1}{\sqrt{2}}(\mathbf{e}_1 \otimes \mathbf{e}_1 \otimes \mathbf{e}_2 - \mathbf{e}_2 \otimes \mathbf{e}_1 \otimes \mathbf{e}_1) \in V \otimes V \otimes V$$

where V is a 3-dimensional inner product space spanned by orthonormal vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ that have the following interpretation:

$$\mathbf{e}_1 = \text{state of the up quark,}$$
$$\mathbf{e}_2 = \text{state of the down quark,}$$
$$\mathbf{e}_3 = \text{state of the strange quark,}$$

and thus

$$\mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k = \text{composite state of the three quark states } e_i, e_j, e_k.$$

3. In physics, the question "What is a tensor?" is often taken to mean "What kind of physical quantities should be represented by tensors?" It is often cast in the form of questions such as "Is elasticity a tensor?", "Is gravity a tensor?", etc. The answer is that the physical quantity in question is a tensor if it obeys the change-of-bases rule in Fact 15.2.6: A d-tensor is an object represented by a list of numbers $a_{j_1\cdots j_d} \in \mathbb{C}$, $j_k = 1, \ldots, n_k$, $k = 1, \ldots, d$, once a basis is chosen, but only if these numbers transform themselves as expected when one changes the basis.
4. Elasticity is an order-4 tensor and may be represented by a hypermatrix $C \in \mathbb{R}^{3\times3\times3\times3}$. If we measure stress using a different choice of coordinates (i.e., different basis), then the new hypermatrix representation $C' \in \mathbb{R}^{3\times3\times3\times3}$ must be related to C via

$$C' = (X, X, X, X) \cdot C \tag{15.12}$$

where $X \in \mathrm{GL}(3, \mathbb{R})$ is the change-of-basis matrix, and Eq. (15.12) is defined according to Fact 6

$$c'_{pqrs} = \sum\nolimits_{i,j,k,l=1}^{3} x_{pi}x_{qj}x_{rk}x_{sl}c_{ijkl}, \qquad p, q, r, s = 1, 2, 3. \tag{15.13}$$

5. Let $\mathfrak{A}$ be an algebra over a field $\mathbb{F}$, i.e., a vector space on which a notion of vector multiplication $\cdot : \mathfrak{A} \times \mathfrak{A} \to \mathfrak{A}$, $(a, b) \mapsto a \cdot b$ is defined. Let $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for $\mathfrak{A}$. Then $\mathfrak{A}$ is completely determined by the hypermatrix $C = [c_{ijk}] \in \mathbb{F}^{n\times n\times n}$ where

$$\mathbf{e}_i \cdot \mathbf{e}_j = \sum_{k=1}^{n} c_{ijk}\mathbf{e}_k.$$

The n^3 entries of C are often called the *structure constants* of $\mathfrak{A}$. This hypermatrix is the coordinate representation of a tensor in $\mathfrak{A} \otimes \mathfrak{A} \otimes \mathfrak{A}$ with respect to the basis $\mathcal{B}$. If we had chosen a new basis $\mathcal{B}'$, then the new coordinate representation C' would be related to C as in Fact 6 — in this case $C' = (X, X, X) \cdot C$ where X is the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}'$. Note that this says that the entries of the hypermatrix of structure constants are coordinates of a tensor with respect to a basis.
6. For an explicit example, the Lie algebra $\mathfrak{so}(3)$ is the set of all skew-symmetric matrices in $\mathbb{R}^{3\times3}$. A basis is given by

$$Z_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad Z_2 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad Z_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The product $\cdot$ for $\mathfrak{so}(3)$ is the commutator product $[X,Y] = XY - YX$. Note that $[Y,X] = -[X,Y]$ and $[X,X] = 0$. Since $[Z_1,Z_2] = Z_3$, $[Z_2,Z_3] = Z_1$, $[Z_3,Z_1] = Z_2$, the structure constants of $\mathfrak{so}(3)$ are given by the hypermatrix $\varepsilon = (\varepsilon_{ijk}) \in \mathbb{R}^{3\times3\times3}$ defined by

$$\begin{aligned}\varepsilon_{ijk} &= \begin{cases} +1 & \text{if } (i,j,k) = (1,2,3),(2,3,1),(3,1,2), \\ -1 & \text{if } (i,j,k) = (1,3,2),(2,1,3),(3,2,1), \\ 0 & \text{if } i = j, j = k, k = i, \end{cases} \\ &= \frac{(i-j)(j-k)(k-i)}{2}.\end{aligned}$$

ε is often called the *Levi-Civita symbol.*

15.3 Tensor Rank

There are several equivalent ways to define the rank of a matrix which yield non-equivalent definitions on hypermatrices of higher order. We will examine two of the most common ones in this chapter: tensor rank as defined below and multilinear rank as defined in Section 15.7. Both notions are due to Frank L. Hitchcock [Hit27a, Hit27b].

Definitions:

Let F be a field.

A hypermatrix $A \in F^{n_1\times\cdots\times n_d}$ has **rank one** or **rank-1** if there exist non-zero $\mathbf{v}^{(i)} \in F^n$, $i = 1,\dots,d$, so that $A = \mathbf{v}^{(1)} \otimes \cdots \otimes \mathbf{v}^{(d)}$ and $\mathbf{v}^{(1)} \otimes \cdots \otimes \mathbf{v}^{(d)}$ is the Segre outer product defined in Eq. (15.3).

The **rank** of a hypermatrix $A \in F^{n_1\times\cdots\times n_d}$ is defined to be the smallest r such that it may be written as a sum of r rank-1 hypermatrices, i.e.,

$$\operatorname{rank}(A) := \min\left\{r : A = \sum\nolimits_{p=1}^{r} \mathbf{v}_p^{(1)} \otimes \cdots \otimes \mathbf{v}_p^{(d)}\right\}. \tag{15.14}$$

For vector spaces $V_1,\dots,V_d$, the **rank** or **tensor rank** of $T \in V_1 \otimes \cdots \otimes V_d$ is

$$\operatorname{rank}(T) = \min\left\{r : T = \sum\nolimits_{p=1}^{r} \mathbf{v}_p^{(1)} \otimes \cdots \otimes \mathbf{v}_p^{(d)}\right\}. \tag{15.15}$$

Here $\mathbf{v}_p^{(k)}$ is a vector in the abstract vector space V_k and $\otimes$ denotes tensor product as defined in Section 15.2.

A hypermatrix or a tensor has **rank zero** if and only if it is zero (in accordance with the convention that the sum over the empty set is zero).

A minimum length decomposition of a tensor or hypermatrix, i.e.,

$$T = \sum\nolimits_{p=1}^{\operatorname{rank}(T)} \mathbf{v}_p^{(1)} \otimes \cdots \otimes \mathbf{v}_p^{(d)}, \tag{15.16}$$

is called a **rank-retaining decomposition** or simply **rank decomposition**.

Facts:

Facts requiring proof for which no specific reference is given can be found in [BCS96, Chap. 19], [Lan12, Chap. 3], [Lim], and the references therein.

1. Let $A = [a_{ijk}]_{i,j,k=1}^{l,m,n}$ and $A_k = [a_{ijk}]_{i,j=1}^{l,m}$. The following are equivalent statements characterize $\operatorname{rank}(A) \le r$.

(a) there exist $\mathbf{x}_1, \ldots, \mathbf{x}_r \in F^l$, $\mathbf{y}_1, \ldots, \mathbf{y}_r \in F^m$, and $\mathbf{z}_1, \ldots, \mathbf{z}_r \in F^m$,

$$A = \mathbf{x}_1 \otimes \mathbf{y}_1 \otimes \mathbf{z}_1 + \cdots + \mathbf{x}_r \otimes \mathbf{y}_r \otimes \mathbf{z}_r;$$

(b) there exist $\mathbf{x}_1, \ldots, \mathbf{x}_r \in F^l$ and $\mathbf{y}_1, \ldots, \mathbf{y}_r \in F^m$ with

$$\operatorname{span}\{A_1, \ldots, A_n\} \subseteq \operatorname{span}\{\mathbf{x}_1\mathbf{y}_1^T, \ldots, \mathbf{x}_r\mathbf{y}_r^T\};$$

(c) there exist diagonal $D_1, \ldots, D_n \in F^{r\times r}$ and $X \in F^{l\times r}, Y \in F^{m\times r}$,

$$A_k = XD_kY^T, \quad k = 1, \ldots, n.$$

The statements analogous to (1a), (1b), and (1c) with r required to be minimal characterize $\operatorname{rank}(A) = r$. A more general form of this fact is valid for for d-tensors.

2. For $A \in F^{n_1\times\cdots\times n_d}$ and $(X_1, \ldots, X_d) \in \mathrm{GL}(n_1, \ldots, n_d, F)$,
$$\operatorname{rank}((X_1, \ldots, X_d) \cdot A) = \operatorname{rank}(A).$$
3. If $T \in V_1 \otimes \cdots \otimes V_d$, $\mathcal{B}_1, \ldots, \mathcal{B}_d$ are bases for $V_1, \ldots, V_d$ and $A = [T]_{\mathcal{B}_1,\ldots,\mathcal{B}_d} \in F^{n_1\times\cdots\times n_d}$ (where $n_k = \dim V_k$, $k = 1, \ldots, d$), then $\operatorname{rank}(A) = \operatorname{rank}(T)$.
4. Since $\mathbb{R}^{l\times m\times n} \subseteq \mathbb{C}^{l\times m\times n}$, given $A \in \mathbb{R}^{l\times m\times n}$, we may consider its rank over $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$,
$$\operatorname{rank}_{\mathbb{F}}(A) = \left\{r : A = \sum\nolimits_{i=1}^{r} \mathbf{x}_i \otimes \mathbf{y}_i \otimes \mathbf{z}_i,\ \mathbf{x}_i \in \mathbb{F}^l, \mathbf{y}_i \in \mathbb{F}^m, \mathbf{z}_i \in \mathbb{F}^n\right\}.$$
Clearly $\operatorname{rank}_{\mathbb{C}}(A) \le \operatorname{rank}_{\mathbb{R}}(A)$. However, strict inequality can occur (see Example 1 next). Note for a matrix $A \in \mathbb{R}^{m\times n}$, this does not happen; we always have $\operatorname{rank}_{\mathbb{C}}(A) = \operatorname{rank}_{\mathbb{R}}(A)$.
5. When $d = 2$, Eq. (15.14) agrees with the usual definition of matrix rank and Eq. (15.15) agrees with the usual definition of rank for linear operators and bilinear forms on finite-dimensional vector spaces.
6. In certain literature, the term "rank" is often used to mean what we have called "order" in Section 15.2. We avoid such usage for several reasons, among which the fact that it does not agree with the usual meaning of rank for linear operators or matrices.

Examples:

1. The phenomenon of rank dependence on field was first observed by Bergman [Ber69]. Take linearly independent pairs of vectors $\mathbf{x}_1, \mathbf{y}_1 \in \mathbb{R}^l$, $\mathbf{x}_2, \mathbf{y}_2 \in \mathbb{R}^m$, $\mathbf{x}_3, \mathbf{y}_3 \in \mathbb{R}^n$ and set $\mathbf{z}_k = \mathbf{x}_k + i\mathbf{y}_k$ and $\bar{\mathbf{z}}_k = \mathbf{x}_k - i\mathbf{y}_k$, then
$$\begin{aligned} A &= \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3 + \mathbf{x}_1 \otimes \mathbf{y}_2 \otimes \mathbf{y}_3 - \mathbf{y}_1 \otimes \mathbf{x}_2 \otimes \mathbf{y}_3 + \mathbf{y}_1 \otimes \mathbf{y}_2 \otimes \mathbf{x}_3 \\ &= \frac{1}{2}(\bar{\mathbf{z}}_1 \otimes \mathbf{z}_2 \otimes \bar{\mathbf{z}}_3 + \bar{\mathbf{z}}_1 \otimes \bar{\mathbf{z}}_2 \otimes \mathbf{z}_3). \end{aligned} \tag{15.17}$$
One may in fact show that $\operatorname{rank}_{\mathbb{C}}(A) = 2 < 3 = \operatorname{rank}_{\mathbb{R}}(A)$.
2. An example where $\operatorname{rank}_{\mathbb{R}}(A) \le 2 < \operatorname{rank}_{\mathbb{Q}}(A)$ is given by
$$\bar{\mathbf{z}} \otimes \mathbf{z} \otimes \bar{\mathbf{z}} + \mathbf{z} \otimes \bar{\mathbf{z}} \otimes \mathbf{z} = 2\mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} - 4\mathbf{y} \otimes \mathbf{y} \otimes \mathbf{x} + 4\mathbf{y} \otimes \mathbf{x} \otimes \mathbf{y} - 4\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{y} \in \mathbb{Q}^{2\times 2\times 2},$$
where $\mathbf{z} = \mathbf{x} + \sqrt{2}\mathbf{y}$ and $\bar{\mathbf{z}} = \mathbf{x} - \sqrt{2}\mathbf{y}$.

Applications:

1. Let $M, C, K \in \mathbb{R}^{n\times n}$ be the mass, damping, and stiffness matrices of a **viscously damped linear system** in free vibration
$$M\ddot{\mathbf{x}}(t) + C\dot{\mathbf{x}}(t) + K\mathbf{x}(t) = 0.$$
where M, C, K are all symmetric positive definite. The system may be decoupled using classical modal analysis [CO65] if and only if
$$CM^{-1}K = KM^{-1}C.$$
Formulated in hypermatrix language, this asks when $A = [M \mid C \mid K] \in \mathbb{R}^{n\times n\times 3}$ has $\operatorname{rank}(A) \le 3$.

2. The notion of tensor rank arises in several areas and a well-known one is algebraic computational complexity [BCS96], notably the complexity of matrix multiplications. This is surprisingly easy to explain. For matrices $X = [x_{ij}], Y = [y_{jk}] \in \mathbb{C}^{n\times n}$, observe that the product may be expressed as

$$XY = \sum\nolimits_{i,j,k=1}^{n} x_{ik}y_{kj}E_{ij} = \sum\nolimits_{i,j,k=1}^{n} \varphi_{ik}(X)\varphi_{kj}(Y)E_{ij} \tag{15.18}$$

where $E_{ij} = \mathbf{e}_i\mathbf{e}_j^* \in \mathbb{C}^{n\times n}$ has all entries 0 except 1 in the (i,j)-entry and $\varphi_{ij}(X) = \operatorname{tr}(E_{ij}^*X) = x_{ij}$ is the linear functional $\varphi_{ij} : \mathbb{C}^{n\times n} \to \mathbb{C}$ dual to E_{ij}. Let $T_n : \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ be the map that takes a pair of matrices $(X,Y) \in \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times n}$ to their product $T(X,Y) = XY \in \mathbb{C}^{n\times n}$. Then by Eq. (15.18), T_n is given by the tensor

$$T_n = \sum\nolimits_{i,j,k=1}^{n} \varphi_{ik} \otimes \varphi_{kj} \otimes E_{ij} \in (\mathbb{C}^{n\times n})^* \otimes (\mathbb{C}^{n\times n})^* \otimes \mathbb{C}^{n\times n}. \tag{15.19}$$

The *exponent of matrix multiplication* is then a positive number ω defined in terms of tensor rank,

$$\omega := \inf\{\alpha : \operatorname{rank}(T_n) = O(n^\alpha),\ n \in \mathbb{N}\}.$$

It is not hard to see that whatever the value of $\omega > 0$, there must exist $O(n^\omega)$ algorithms for multiplying $n\times n$ matrices. In fact, every r-term decomposition (15.16) of T_n yields an explicit algorithm for multiplying two $n \times n$ matrices in $O(n^{\log_2 r})$ complexity. Via elementary row operations, we may deduce that any $O(n^\omega)$ algorithm for computing matrix multiplications $(A,B) \mapsto AB$ would also yield a corresponding $O(n^\omega)$ algorithm for matrix inversion $A \mapsto A^{-1}$ (and thus for solving linear systems $Ax = b$). If we choose the standard bases $\{E_{ij} : i,j \in \langle n\rangle\}$ on $\mathbb{C}^{n\times n}$ and its dual bases $\{\varphi_{ij} : i,j \in \langle n\rangle\}$ on the dual space $(\mathbb{C}^{n\times n})^*$, then the 3-tensor T_n may be represented by a 3-hypermatrix

$$M_n \in \mathbb{C}^{n^2\times n^2\times n^2}.$$

The connection between the exponent of matrix multiplication and tensor rank was first noted by Strassen in [Str73]. We refer the reader to Chapter 61 and [Knu98] for very readable accounts and to [BCS96] for an extensive in-depth discussion.

3. The special case $n = 2$ is behind *Strassen's algorithm for matrix multiplication and inversion* with $O(n^{\log_2 7})$ time complexity [Str69]. We shall present it in the modern language of hypermatrices. We write $M_2 = [A_1 \mid A_2 \mid A_3 \mid A_4] \in \mathbb{C}^{4\times4\times4}$ where the matrix slices of M_2 are

$$A_1 = \begin{bmatrix} I & O \\ O & O \end{bmatrix}, \quad A_2 = \begin{bmatrix} O & O \\ I & O \end{bmatrix}, \quad A_3 = \begin{bmatrix} O & I \\ O & O \end{bmatrix}, \quad A_4 = \begin{bmatrix} O & O \\ O & I \end{bmatrix},$$

I and O are 2×2 identity and zero matrices. Define

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & -1 & 0 & 1 \end{bmatrix},$$

$$D_1 = \operatorname{diag}(1,0,1,1,0,0,-1), \quad D_2 = \operatorname{diag}(0,0,-1,0,0,1,0),$$
$$D_3 = \operatorname{diag}(0,0,0,0,1,0,1), \quad D_4 = \operatorname{diag}(-1,1,0,0,-1,-1,0).$$

One may check that $XD_jY^T = A_j$ for $j = 1,2,3,4$. In other words, this is a simultaneous diagaonlization of A_1, A_2, A_3, A_4 by X and Y in the sense of Fact 1 and so we conclude that $\operatorname{rank}(M_2) \le 7$. In fact, it has been shown that $\operatorname{rank}(M_2) = 7$ [HK71, Win71] and much more recently, the border rank (see Section 15.4) of M_2 is 7 [Lan06].

4. In quantum computing, a *pure state*, also known as a completely separable state, corresponds to a rank-1 tensor. A quantum state that is not pure is called *entangled*. A natural, but not the most commonly used, measure of the degree of *entanglement* is therefore tensor rank [Bry02], i.e., the minimal number of pure states it can be written as a sum of. For example, the well-known Greenberger–Horne–Zeilinger state [GHZ89] may be regarded as a $2 \times 2 \times 2$ hypermatrix of rank 2:

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle \otimes |1\rangle) \in \mathbb{C}^{2\times 2\times 2},$$

while the W state [DVC00] may be regarded as a $2 \times 2 \times 2$ hypermatrix of rank 3:

$$|W\rangle = \frac{1}{\sqrt{3}}(|0\rangle \otimes |0\rangle \otimes |1\rangle + |0\rangle \otimes |1\rangle \otimes |0\rangle + |1\rangle \otimes |0\rangle \otimes |0\rangle) \in \mathbb{C}^{2\times 2\times 2}.$$

15.4 Border Rank

We now discuss a phenomenon that may appear peculiar at first since one does not encounter this for usual matrices or tensors of order 2. As one will see from the following examples, we may get a sequence of 3-hypermatrices of rank not more than 2 converging to a limit that has rank 3, which is somewhat surprising since for matrices, this can never happen. Another way to say this is that the set $\{A \in \mathbb{C}^{m\times n} : \mathrm{rank}(A) \le r\}$ is closed. What we deduce from this example is the same does not hold for hypermatrices, the set $\{A \in \mathbb{C}^{l\times m\times n} : \mathrm{rank}(A) \le r\}$ is not a closed set in general.

Definitions:

The **border rank** of a hypermatrix $A \in \mathbb{C}^{n_1\times\cdots\times n_d}$ is

$$\overline{\mathrm{rank}}(A) = \min\left\{ r : \inf_{\mathrm{rank}(B)\le r} \|A - B\| = 0 \right\},$$

where $\|\cdot\|$ is any norm on hypermatrices (including one obtained by identifying $A \in \mathbb{C}^{n_1\times\cdots\times n_d}$ with $\mathbb{C}^{n_1\cdots n_d}$; see Section 15.8).

Facts:

Facts requiring proof for which no specific reference is given can be found in [BCS96, Chap. 19], [Lan12, Chap. 3], [Lim], and the references therein.

1. For $A \in \mathbb{C}^{n_1\times\cdots\times n_d}$, $\overline{\mathrm{rank}}(A) \le \mathrm{rank}(A)$.
2. For a matrix $A \in \mathbb{C}^{m\times n}$, $\overline{\mathrm{rank}}(A) = \mathrm{rank}(A)$.
3. There exist examples of a sequence of 3-hypermatrices of rank not more than 2 that converges to a limit that has rank 3 [BLR80] (see Examples 1 and 2). In fact, the gap between rank and border rank can be arbitrarily large.
4. Let $A = [A_1 \mid A_2 \mid A_3] \in \mathbb{C}^{n\times n\times 3}$ where A_2 is invertible. Then
$$\overline{\mathrm{rank}}(A) \ge n + \frac{1}{2}\mathrm{rank}(A_1A_2^{-1}A_3 - A_3A_2^{-1}A_1).$$
5. Let $A \in \mathbb{C}^{n_1\times\cdots\times n_d}$ and $(X_1, \ldots, X_d) \in \mathrm{GL}(n_1, \ldots, n_d, \mathbb{C})$. Then
$$\overline{\mathrm{rank}}((X_1, \ldots, X_d) \cdot A) = \overline{\mathrm{rank}}(A).$$
6. While border rank is defined here for hypermatrices, the definition of border rank extends to tensors via coordinate representation because of the invariance of rank under change of basis (cf. Fact 15.3.2).

7. Suppose $\operatorname{rank}(T) = r$ and $\overline{\operatorname{rank}}(T) = s$. If $s < r$, then T has no best rank-s approximation.

Examples:

1. [BLR80] The original example of a hypermatrix whose border rank is less than its rank is a $2 \times 2 \times 2$ hypermatrix. In the notation of Section 15.1, if we choose $\mathbf{x}_p = (1, 0), \mathbf{y}_p = (0, 1) \in \mathbb{C}^2$ for $p = 1, 2, 3$, then

$$A = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \left[\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right] \in \mathbb{C}^{2\times 2\times 2}$$

and

$$\begin{aligned} A_\varepsilon &= \frac{1}{\varepsilon} \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix} - \frac{1}{\varepsilon} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{1}{\varepsilon} \left[\begin{array}{cc|cc} 1 & \varepsilon & \varepsilon & \varepsilon^2 \\ \varepsilon & \varepsilon^2 & \varepsilon^2 & \varepsilon^3 \end{array}\right] - \frac{1}{\varepsilon} \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] = \left[\begin{array}{cc|cc} 0 & 1 & 1 & \varepsilon \\ 1 & \varepsilon & \varepsilon & \varepsilon^2 \end{array}\right] \in \mathbb{C}^{2\times 2\times 2}, \end{aligned}$$

from which it is clear that $\lim_{\varepsilon\to 0} A_\varepsilon = A$. We note here that the hypermatrix A is actually the same hypermatrix that represents the W state in quantum computing (cf. Example 15.3.4).

2. Choose linearly independent pairs of vectors $\mathbf{x}_1, \mathbf{y}_1 \in \mathbb{C}^l$, $\mathbf{x}_2, \mathbf{y}_2 \in \mathbb{C}^m$, $\mathbf{x}_3, \mathbf{y}_3 \in \mathbb{C}^n$ (so $l, m, n \geq 2$ necessarily). Define the 3-hypermatrix,

$$A := \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{y}_3 + \mathbf{x}_1 \otimes \mathbf{y}_2 \otimes \mathbf{x}_3 + \mathbf{y}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3 \in \mathbb{C}^{l\times m\times n},$$

and, for any $\varepsilon \neq 0$, a family of 3-hypermatrices parameterized by ε,

$$A_\varepsilon := \frac{(\mathbf{x}_1 + \varepsilon\mathbf{y}_1) \otimes (\mathbf{x}_2 + \varepsilon\mathbf{y}_2) \otimes (\mathbf{x}_3 + \varepsilon\mathbf{y}_3) - \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3}{\varepsilon}.$$

Now it is easy to verify that

$$A - A_\varepsilon = \varepsilon(\mathbf{y}_1 \otimes \mathbf{y}_2 \otimes \mathbf{x}_3 + \mathbf{y}_1 \otimes \mathbf{x}_2 \otimes \mathbf{y}_3 + \mathbf{x}_1 \otimes \mathbf{y}_2 \otimes \mathbf{y}_3) \tag{15.20}$$

and so

$$\|A - A_\varepsilon\| = O(\varepsilon)$$

In other words, A can be approximated arbitrarily closely by A_ε. As a result,

$$\lim_{\varepsilon\to 0} A_\varepsilon = A.$$

One may also check that $\operatorname{rank}(A) = 3$ while it is clear that $\operatorname{rank}(A_\varepsilon) \leq 2$. By definition, the border rank of A is not more than 2. In fact, $\overline{\operatorname{rank}}(A) = 2$ because a rank-1 hypermatrix always has border rank equals to 1 and as such the border rank of a hypermatrix whose rank and border rank differ must have border rank at least 2 and rank at least 3.

3. There are various results regarding the border rank of hypermatrices of specific dimensions. If either l, m, or $n = 2$, then a hypermatrix $A = [A_1, A_2] \in \mathbb{C}^{m\times n\times 2}$ may be viewed as a matrix pencil and we may depend on the Kronecker-Weierstraß canonical form to deduce results about rank and border rank.

4. One may define border rank algebraically over arbitrary fields without involving norm by considering expressions like those in Eq. (15.20) modulo ε. We refer the reader to [BCS96, Knu98] for more details on this.

15.5 Generic and Maximal Rank

Unlike tensor rank, border rank, and multilinear rank, the notions of rank discussed in this section apply to the whole space rather than an individual hypermatrix or tensor.

Definitions:

Let F be a field.

The **maximum rank** of $F^{n_1 \times \cdots \times n_d}$ is

$$\operatorname{maxrank}_F(n_1, \dots, n_d) = \max\{\operatorname{rank}(A) : A \in F^{n_1 \times \cdots \times n_d}\}.$$

The **generic rank** of $\mathbb{C}^{n_1 \times \cdots \times n_d}$ as

$$\operatorname{genrank}_{\mathbb{C}}(n_1, \dots, n_d) = \max\{\overline{\operatorname{rank}}(A) : A \in \mathbb{C}^{n_1 \times \cdots \times n_d}\}.$$

Facts:

Facts requiring proof for which no specific reference is given can be found in [Lan12, Chap. 3].

1. The concepts of maxrank and genrank are uninteresting for matrices (i.e., 2-hypermatrices) since
$$\operatorname{genrank}_{\mathbb{C}}(m, n) = \operatorname{maxrank}_{\mathbb{C}}(m, n) = \max\{m, n\},$$
but for $d > 2$, the exact values of $\operatorname{genrank}_{\mathbb{C}}(l, m, n)$ and $\operatorname{maxrank}_{\mathbb{C}}(l, m, n)$ are mostly unknown.
2. Since $\overline{\operatorname{rank}}(A) \le \operatorname{rank}(A)$, $\operatorname{genrank}_{\mathbb{C}}(n_1, \dots, n_d) \le \operatorname{maxrank}_{\mathbb{C}}(n_1, \dots, n_d)$.
3. A simple dimension count yields the following lower bound
$$\operatorname{genrank}_{\mathbb{C}}(n_1, \dots, n_d) \ge \left\lceil \frac{n_1 \cdots n_d}{n_1 + \cdots + n_d - d + 1} \right\rceil.$$
Note that strict inequaltiy can occur when $d > 2$. For example, for $d = 3$,
$$\operatorname{genrank}_{\mathbb{C}}(2, 2, 2) = 2 \text{ whereas } \operatorname{maxrank}_{\mathbb{C}}(2, 2, 2) = 3$$
and $\operatorname{genrank}_{\mathbb{C}}(3, 3, 3) = 5$ [Str83] while $\lceil 3^3/(3 \times 3 - 3 + 1) \rceil = 4$.
4. The generic rank is a *generic* property in $\mathbb{C}^{n_1 \times \cdots \times n_d}$, i.e., every hypermatrix would have rank equal to $\operatorname{genrank}_{\mathbb{C}}(n_1, \dots, n_d)$ except for those in some proper Zariski-closed subset. In particular, hypermatrices that do not have rank $\operatorname{genrank}_{\mathbb{C}}(n_1, \dots, n_d)$ are contained in a measure-zero subset of $\mathbb{C}^{n_1 \times \cdots \times n_d}$ (with respect to, say, the Lebesgue or Gaussian measure).
5. There is a notion of *typical rank* that is analogous to generic rank for fields that are not algebraically closed.

15.6 Rank-Retaining Decomposition

Matrices have a notion of rank-retaining decomposition, i.e., given $A \in \mathbb{C}^{m \times n}$, there are various ways to write it down as a product of two matrices $A = GH$ where $G \in \mathbb{C}^{m \times r}$ and $H \in \mathbb{C}^{r \times n}$ where $r = \operatorname{rank}(A)$. Examples include the LU, QR, SVD, and others.

There is a rank-retaining decomposition that may be stated in a form that is almost identical to the singular value decomposition of a matrix, with one caveat — the analogue of the singular vectors would in general not be orthogonal (cf. Fact 1). This decomposition is often unique, a result due to Kruskal that depends on the notion of *Kruskal rank*, a concept that has been reinvented multiple times under different names (originally as *girth* in matroid theory [Oxl11], as *krank* in tensor decompositions [SB00], and most recently as *spark* in compressive sensing [DE03]). The term "Kruskal rank" was not coined by Kruskal himself who simply denoted it as k_S. It is an unfortunate nomenclature, since it is not a notion of rank. "Kruskal dimension" would have been more appropriate.

Definitions:

Let V be a vector space, $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$, and $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$.

The **Kruskal rank** or **krank** of S, denoted $\operatorname{krank}(S)$ or $\operatorname{krank}(\mathbf{v}_1, \dots, \mathbf{v}_r)$, is the largest $k \in \mathbb{N}$ so that every k-element subset of S is linearly independent.

The **girth** or **spark** of S is the smallest $s \in \mathbb{N}$ so that there exists an s-element subset of S that is linear dependent.

We say that a decomposition of the form (15.21) is **essentially unique** if given another such decomposition,

$$\sum\nolimits_{p=1}^{r} \alpha_p \mathbf{v}_p^{(1)} \otimes \cdots \otimes \mathbf{v}_p^{(d)} = \sum\nolimits_{p=1}^{r} \beta_p \mathbf{w}_p^{(1)} \otimes \cdots \otimes \mathbf{w}_p^{(d)},$$

we must have (i) the coefficients $\alpha_p = \beta_p$ for all $p = 1, \dots, r$; and (ii) the factors $\mathbf{v}_p^{(1)}, \dots, \mathbf{v}_p^{(d)}$ and $\mathbf{w}_p^{(1)}, \dots, \mathbf{w}_p^{(d)}$ differ at most via unimodulus scaling, i.e.,

$$\mathbf{v}_p^{(1)} = e^{i\theta_{1p}} \mathbf{w}_p^{(1)}, \dots, \mathbf{v}_p^{(d)} = e^{i\theta_{dp}} \mathbf{w}_p^{(d)}$$

where $\theta_{1p} + \cdots + \theta_{dp} \equiv 0 \bmod 2\pi$, for all $p = 1, \dots, r$. In the event when successive coefficients are equal, $\sigma_{p-1} > \sigma_p = \sigma_{p+1} = \cdots = \sigma_{p+q} > \sigma_{p+q+1}$, the uniqueness of the factors in (ii) is only up to relabeling of indices $p, \dots, p+q$.

Facts:

Facts requiring proof for which no specific reference is given can be found in [Lim] and the references therein.

1. Let $A \in \mathbb{C}^{n_1 \times \cdots \times n_d}$. Then A has a rank-retaining decomposition

$$A = \sum\nolimits_{p=1}^{r} \sigma_p \mathbf{v}_p^{(1)} \otimes \cdots \otimes \mathbf{v}_p^{(d)}, \tag{15.21}$$

with $r = \operatorname{rank}(A)$ and $\sigma_1, \dots, \sigma_r \in \mathbb{R}$, which can be chosen to be all positive, and

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r,$$

and $\mathbf{v}_p^{(k)} \in \mathbb{C}^{n_k}$ has unit norm $\|\mathbf{v}_p^{(k)}\|_2 = 1$ all $k = 1, \dots, d$ and all $p = 1, \dots, r$.

2. Fact 1 has a coordinate-free counterpart: If $T \in V_1 \otimes \cdots \otimes V_d$, then we may also write

$$T = \sum\nolimits_{p=1}^{r} \sigma_p \mathbf{v}_p^{(1)} \otimes \cdots \otimes \mathbf{v}_p^{(d)},$$

where $r = \operatorname{rank}(T)$, $\mathbf{v}_p^{(k)} \in V_k$ are vectors in an abstract $\mathbb{C}$-vector space. Furthermore if the V_k's are all equipped with norms, then $\mathbf{v}_p^{(k)}$ may all be chosen to be unit vectors and σ_p may all be chosen to be positive real numbers. As such, we may rightly call a decomposition of the form (15.21) a *tensor decomposition*, with Fact 1 being the special case when $V_k = \mathbb{C}^{n_k}$.

3. For the special case $d = 2$, the unit-norm vectors $\{\mathbf{v}_1^{(1)}, \dots, \mathbf{v}_r^{(1)}\}$ and $\{\mathbf{v}_1^{(2)}, \dots, \mathbf{v}_r^{(2)}\}$ may in fact be chosen to be orthonormal. If we write $U := [\mathbf{v}_1^{(1)}, \dots, \mathbf{v}_r^{(1)}] \in \mathbb{C}^{n_1 \times r}$ and $V := [\mathbf{v}_1^{(2)}, \dots, \mathbf{v}_r^{(2)}] \in \mathbb{C}^{n_2 \times r}$, Fact 1 yields the singular value decomposition

$A = U\Sigma V^*$. Hence, Fact 1 may be viewed as a generalization of matrix singular value decomposition. The one thing that we lost in going from order $d = 2$ to $d \geq 3$ is the orthonormality of the "singular vectors" $\mathbf{v}_1^{(k)}, \dots, \mathbf{v}_r^{(k)}$. In fact, this is entirely to be expected because when $d \geq 2$, it is often the case that $\operatorname{rank}(A) > \max\{n_1, \dots, n_d\}$ and so it is impossible for $\mathbf{v}_1^{(k)}, \dots, \mathbf{v}_r^{(k)}$ to be orthonormal or even linearly independent.

4. A decomposition of the form in Fact 1 is in general not unique if we do not impose orthogonality on the factors $\{\mathbf{v}_p^{(k)} : p = 1, \dots, r\}$ but this is easily seen to be impossible via a simple dimension count: Suppose $n_1 = \cdots = n_d = n$, an orthonormal collection of n vectors has dimension $n(n-1)/2$ and so the right-hand side of Eq. (15.21) has dimension at most $n + dn(n-1)/2$ but the left-hand side has dimension n^d. Fortunately, there is an amazing result due to Kruskal that guarantees a slightly weaker form of uniqueness of Eq. (15.21) for $d \geq 3$ without requiring orthogonality.
5. It is clear that girth (= spark) and krank are one and the same notion, related by $\operatorname{girth}(S) = \operatorname{krank}(S) + 1$.
6. The krank of $\mathbf{v}_1, \dots, \mathbf{v}_r \in V$ is $\mathrm{GL}(V)$-invariant: If $X \in \mathrm{GL}(V)$, then

$$\operatorname{krank}(X\mathbf{v}_1, \dots, X\mathbf{v}_r) = \operatorname{krank}(\mathbf{v}_1, \dots, \mathbf{v}_r).$$

7. $\operatorname{krank}(\mathbf{v}_1, \dots, \mathbf{v}_r) \leq \dim \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$.
8. [Kru77] Let $A \in \mathbb{C}^{l \times m \times n}$. Then a decomposition of the form

$$A = \sum\nolimits_{p=1}^{r} \sigma_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p$$

is both rank-retaining, i.e., $r = \operatorname{rank}(A)$, and essentially unique if the following combinatorial condition is met:

$$\operatorname{krank}(\mathbf{u}_1, \dots, \mathbf{u}_r) + \operatorname{krank}(\mathbf{v}_1, \dots, \mathbf{v}_r) + \operatorname{krank}(\mathbf{w}_1, \dots, \mathbf{w}_r) \geq 2r + 2.$$

This is the *Kruskal uniqueness theorem.*

9. [SB00] Fact 8 has been generalized to arbitrary order $d \geq 3$. Let $A \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ where $d \geq 3$. Then a decomposition of the form

$$A = \sum\nolimits_{p=1}^{r} \sigma_p \mathbf{v}_p^{(1)} \otimes \cdots \otimes \mathbf{v}_p^{(d)}$$

is both rank-retaining, i.e., $r = \operatorname{rank} A$, and essentially unique if the following condition is satisfied:

$$\sum\nolimits_{k=1}^{d} \operatorname{krank}(\mathbf{v}_1^{(k)}, \dots, \mathbf{v}_r^{(k)}) \geq 2r + d - 1. \tag{15.22}$$

10. A result analogous to the Kruskal uniqueness theorem (Fact 8) does not hold for $d = 2$, since a decomposition of a matrix $A \in \mathbb{C}^{n \times n}$ can be written in infinitely many different ways $A = UV^\top = (UX^{-1})(XV)$ for any $X \in \mathrm{GL}(n, \mathbb{C})$. This is not surprising since Eq. (15.22) can never be true for $d = 2$ because of Fact 7.
11. [Der13] For $d \geq 3$, the condition (15.22) is sharp in the sense that the right-hand side cannot be further reduced.
12. One may also write the decomposition in Fact 1 in the form of multilinear matrix multiplication,

$$A = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{x}_i \otimes \mathbf{y}_i \otimes \mathbf{z}_i = (X, Y, Z) \cdot \Lambda, \quad a_{\alpha\beta\gamma} = \sum\nolimits_{i=1}^{r} \lambda_i x_{\alpha i} y_{\beta i} z_{\gamma i},$$

of three matrices $X = [\mathbf{x}_1, \dots, \mathbf{x}_r] \in \mathbb{C}^{l \times r}$, $Y = [\mathbf{y}_1, \dots, \mathbf{y}_r] \in \mathbb{C}^{m \times r}$, $Z = [\mathbf{z}_1, \dots, \mathbf{z}_r] \in \mathbb{C}^{n \times r}$ with a diagonal hypermatrix $\Lambda = \operatorname{diag}[\lambda_1, \dots, \lambda_r] \in \mathbb{C}^{r \times r \times r}$, i.e.,

$$\lambda_{ijk} = \begin{cases} \lambda_i & i = j = k, \\ 0 & \text{otherwise.} \end{cases}$$

Note however that this is not a multilinear rank-retaining decomposition in the sense of the next section, since $\mu\,\text{rank}(A) \neq (r, r, r)$ in general.

15.7 Multilinear Rank

For a matrix $A = [a_{ij}] \in F^{m\times n}$ we have a nice equality between three numerical invariants associated with A:

$$\begin{aligned}\text{rank}(A) &= \dim \text{span}_F\{A_{\bullet 1}, \ldots, A_{\bullet n}\} &(15.23)\\ &= \dim \text{span}_F\{A_{1\bullet}, \ldots, A_{m\bullet}\} &(15.24)\\ &= \min\{r : A = \mathbf{x}_1\mathbf{y}_1^\top + \cdots + \mathbf{x}_r\mathbf{y}_r^\top\}. &(15.25)\end{aligned}$$

Here we let $A_{i\bullet} = [a_{i1}, \ldots, a_{im}]^T \in F^m$ and $A_{\bullet j} = [a_{1j}, \ldots, a_{nj}]^T \in F^n$ denote the ith row and jth column vectors of A. The numbers in Eqs. (15.24) and (15.23) are the row and column ranks of A. Their equality is a standard fact in linear algebra and the common value is called the rank of A. The number in Eq. (15.25) is also easily seen to be equal to rank(A).

For a hypermatrix $A = [a_{ijk}] \in F^{l\times m\times n}$, one may also define analogous numbers. However, we generally have four distinct numbers, with the first three, called *multilinear rank* associated with the three different ways to slice A, and the analog of Eq. (15.25) being the tensor rank of A (cf. Section 15.3). The multilinear rank is essentially matrix rank and so inherits many of the latter's properties, so we do not see the sort of anomalies discussed in Sections 15.4 and 15.5.

Like the tensor rank, the notion of multilinear rank was due to Hitchcock [Hit27b], as a special case (2-plex rank) of his *multiplex rank*.

Definitions:

For a hypermatrix $A = [a_{ijk}] \in F^{l\times m\times n}$,

$$\begin{aligned}r_1 &= \dim \text{span}_F\{A_{1\bullet\bullet}, \ldots, A_{l\bullet\bullet}\}, &(15.26)\\ r_2 &= \dim \text{span}_F\{A_{\bullet 1\bullet}, \ldots, A_{\bullet m\bullet}\}, &(15.27)\\ r_3 &= \dim \text{span}_F\{A_{\bullet\bullet 1}, \ldots, A_{\bullet\bullet n}\}. &(15.28)\end{aligned}$$

Here $A_{i\bullet\bullet} = [a_{ijk}]_{j,k=1}^{m,n} \in F^{m\times n}$, $A_{\bullet j\bullet} = [a_{ijk}]_{i,k=1}^{l,n} \in F^{l\times n}$, $A_{\bullet\bullet k} = [a_{ijk}]_{i,j=1}^{l,m} \in F^{l\times m}$. The **multilinear rank** of $A \in F^{l\times m\times n}$ is $\mu\,\text{rank}(A) := (r_1, r_2, r_3)$, with r_1, r_2, r_3 defined in Eqs. (15.26), (15.27), and (15.28). The definition for a d-hypermatrix with $d > 3$ is analogous.

The k**th flattening map** on $F^{n_1\times\cdots\times n_d}$ is the function

$$\flat_k : F^{n_1\times\cdots\times n_d} \to F^{n_k\times(n_1\ldots\widehat{n}_k\ldots n_d)}$$

defined by

$$(\flat_k(A))_{ij} = (A)_{s_k(i,j)}$$

where $s_k(i, j)$ is the jth element in lexicographic order in the subset of $\langle n_1\rangle \times \cdots \times \langle n_d\rangle$ consisting of elements that have kth coordinate equal to i, and by convention a caret over any entry of a d-tuple means that the respective entry is omitted. See Example 1.

The **multilinear kernels** or **nullspaces** of $T \in U \otimes V \otimes W$ are

$$\begin{aligned}\ker_1(T) &= \{\mathbf{u} \in U : T(\mathbf{u}, \mathbf{v}, \mathbf{w}) = 0 \text{ for all } \mathbf{v} \in V, \mathbf{w} \in W\},\\ \ker_2(T) &= \{\mathbf{v} \in V : T(\mathbf{u}, \mathbf{v}, \mathbf{w}) = 0 \text{ for all } \mathbf{u} \in U, \mathbf{w} \in W\},\\ \ker_3(T) &= \{\mathbf{w} \in W : T(\mathbf{u}, \mathbf{v}, \mathbf{w}) = 0 \text{ for all } \mathbf{u} \in U, \mathbf{v} \in V\}.\end{aligned}$$

The **multilinear images** or **range spaces** of $T \in U \otimes V \otimes W$ are

$$\begin{aligned}
\mathrm{im}_1(T) &= \{A \in V \otimes W : A(\mathbf{v}, \mathbf{w}) = T(\mathbf{u}_0, \mathbf{v}, \mathbf{w}) \text{ for some } \mathbf{u}_0 \in U\},\\
\mathrm{im}_2(T) &= \{B \in U \otimes W : B(\mathbf{u}, \mathbf{w}) = T(\mathbf{u}, \mathbf{v}_0, \mathbf{w}) \text{ for some } \mathbf{v}_0 \in V\},\\
\mathrm{im}_3(T) &= \{C \in U \otimes V : C(\mathbf{u}, \mathbf{v}) = T(\mathbf{u}, \mathbf{v}, \mathbf{w}_0) \text{ for some } \mathbf{w}_0 \in W\}.
\end{aligned}$$

The **multilinear nullity** of $T \in U \otimes V \otimes W$ is

$$\mu\,\mathrm{nullity}(T) = (\dim \ker_1(T), \dim \ker_2(T), \dim \ker_3(T)).$$

The **multilinear rank** of $T \in U \otimes V \otimes W$ is

$$\mu\,\mathrm{rank}(T) = (\dim \mathrm{im}_1(T), \dim \mathrm{im}_2(T), \dim \mathrm{im}_3(T)).$$

Facts:

Facts requiring proof for which no specific reference is given can be found in [Lan12, Chap. 2] as well as [Lim] and the references therein.

1. For $A \in F^{l \times m \times n}$, the multilinear rank $\mu\,\mathrm{rank}(A) = (r_1, r_2, r_3)$ is given by

$$r_1 = \mathrm{rank}(\flat_1(A)), \quad r_2 = \mathrm{rank}(\flat_2(A)), \quad r_3 = \mathrm{rank}(\flat_3(A)),$$

where rank here is of course the usual matrix rank of the matrices $\flat_1(A), \flat_2(A), \flat_3(A)$.
2. Let $A \in F^{l \times m \times n}$. For any $X \in \mathrm{GL}(l, F), Y \in \mathrm{GL}(m, F), Z \in \mathrm{GL}(n, F)$,

$$\mu\,\mathrm{rank}((X, Y, Z) \cdot A) = \mu\,\mathrm{rank}(A).$$

3. Let $T \in V_1 \otimes \cdots \otimes V_d$. For bases $\mathcal{B}_k$ for $V_k, k = 1, \ldots, d$,

$$\mu\,\mathrm{rank}([T]_{\mathcal{B}_1, \ldots, \mathcal{B}_d}) = \mu\,\mathrm{rank}(T).$$

4. Let $T \in U \otimes V \otimes W$. The *rank-nullity theorem* for multilinear rank is:

$$\mu\,\mathrm{nullity}(T) + \mu\,\mathrm{rank}(T) = (\dim U, \dim V, \dim W).$$

5. If $A \in F^{l \times m \times n}$ has $\mu\,\mathrm{rank}(A) = (p, q, r)$, then there exist matrices $X \in F^{l \times p}, Y \in F^{m \times q}, Z \in F^{n \times r}$ of full column rank and $C \in \mathbb{C}^{p \times q \times r}$ such that

$$A = (X, Y, Z) \cdot C,$$

i.e., A may be expressed as a multilinear matrix product of C by X, Y, Z, defined according to Eq. (15.1). We shall call this a multilinear rank-retaining decomposition or just *multilinear decomposition* for short.
6. One may also write Fact 5 in the form of a multiply indexed sum of decomposable hypermatrices

$$A = \sum\nolimits_{i,j,k=1}^{p,q,r} c_{ijk} \mathbf{x}_i \otimes \mathbf{y}_j \otimes \mathbf{z}_k$$

where $\mathbf{x}_i \in F^l, \mathbf{y}_j \in F^m, \mathbf{z}_k \in F^n$ are the respective columns of the matrices X, Y, Z.
7. Let $A \in \mathbb{C}^{l \times m \times n}$ and $\mu\,\mathrm{rank}(A) = (p, q, r)$. There exist $L_1 \in \mathbb{C}^{l \times p}$, $L_2 \in \mathbb{C}^{m \times q}$, $L_3 \in \mathbb{C}^{n \times r}$ unit lower triangular and $U \in \mathbb{C}^{p \times q \times r}$ such that

$$A = (L_1, L_2, L_3) \cdot U.$$

This may be viewed as the analogues of the LU decomposition for hypermatrices with respect to the notion of multilinear rank.

8. Let $A \in \mathbb{C}^{l\times m\times n}$ and $\mu\,\mathrm{rank}(A) = (p, q, r)$. There exist $Q_1 \in \mathbb{C}^{l\times p}$, $Q_2 \in \mathbb{C}^{m\times q}$, $Q_3 \in \mathbb{C}^{n\times r}$ with orthonormal columns and $R \in \mathbb{C}^{p\times q\times r}$ such that
$$A = (Q_1, Q_2, Q_3) \cdot R.$$
This may be viewed as the analogues of the QR decomposition for hypermatrices with respect to the notion of multilinear rank.
9. The LU and QR decompositions in Facts 7 and 8 are among the few things that can be computed for hypermatrices, primarily because multilinear rank is essentially a matrix notion. For example, one may apply usual Gaussian elimination or Householder/Givens QR to the flattenings of $A \in F^{l\times m\times n}$ successively, reducing all computations to standard matrix computations.

Examples:

1. Let
$$A = \left[\begin{array}{ccc|ccc} a_{111} & a_{121} & a_{131} & a_{112} & a_{122} & a_{132} \\ a_{211} & a_{221} & a_{231} & a_{212} & a_{222} & a_{232} \\ a_{311} & a_{321} & a_{331} & a_{312} & a_{322} & a_{332} \\ a_{411} & a_{421} & a_{431} & a_{412} & a_{422} & a_{432} \end{array}\right] \in \mathbb{C}^{4\times 3\times 2},$$
then
$$\flat_1(A) = \begin{bmatrix} a_{111} & a_{112} & a_{121} & a_{122} & a_{131} & a_{132} \\ a_{211} & a_{212} & a_{221} & a_{222} & a_{231} & a_{232} \\ a_{311} & a_{312} & a_{321} & a_{322} & a_{331} & a_{332} \\ a_{411} & a_{412} & a_{421} & a_{422} & a_{431} & a_{432} \end{bmatrix} \in \mathbb{C}^{4\times 6},$$
$$\flat_2(A) = \begin{bmatrix} a_{111} & a_{112} & a_{211} & a_{212} & a_{311} & a_{312} & a_{411} & a_{412} \\ a_{121} & a_{122} & a_{221} & a_{222} & a_{321} & a_{322} & a_{421} & a_{422} \\ a_{131} & a_{132} & a_{231} & a_{232} & a_{331} & a_{332} & a_{431} & a_{432} \end{bmatrix} \in \mathbb{C}^{3\times 8},$$
$$\flat_3(A) = \begin{bmatrix} a_{111} & a_{121} & a_{131} & a_{211} & a_{221} & a_{231} & a_{311} & a_{321} & a_{331} & a_{411} & a_{421} & a_{431} \\ a_{112} & a_{122} & a_{132} & a_{212} & a_{222} & a_{232} & a_{312} & a_{322} & a_{332} & a_{412} & a_{422} & a_{432} \end{bmatrix} \in \mathbb{C}^{2\times 12}.$$
2. Note that if we had a bilinear form represented by a matrix $M \in \mathbb{C}^{m\times n}$, the analogues of these subspaces would be the four fundamental subspaces of the matrix: $\ker_1(M) = \ker(M)$, $\mathrm{im}_1(M) = \mathrm{im}(M)$, $\ker_2(M) = \ker(M^T)$, $\mathrm{im}_2(M) = \mathrm{im}(M^T)$. The rank-nullity theorem reduces to the usual one:
$$(\mathrm{nullity}(M) + \mathrm{rank}(M), \mathrm{nullity}(M^T) + \mathrm{rank}(M^T)) = (n, m).$$

15.8 Norms

In this section we will discuss the Hölder, induced, and nuclear norms of hypermatrices. When discussing multiple variety of norms, one has to introduce different notation to distinguish them and here we follow essentially the notation and terminology for matrix norms in Chapter 24, adapted as needed for hypermatrices. For the induced and nuclear norms, we assume $d = 3$ to avoid notational clutter.

Definitions:

For $A = [a_{j_1\cdots j_d}]_{j_1,\ldots,j_d=1}^{n_1,\ldots,n_d} \in \mathbb{C}^{n_1\times\cdots\times n_d}$ and $p \in [1, \infty]$, the **Hölder p-norm** is defined by
$$\|A\|_{H,p} := \left(\sum\nolimits_{j_1,\ldots,j_d=1}^{n_1,\ldots,n_d} |a_{j_1\cdots j_d}|^p\right)^{1/p}, \tag{15.29}$$

with the usual alternate definition for $p = \infty$,

$$\|A\|_{H,\infty} := \max\{|a_{j_1\cdots j_d}| : j_1 = 1, \ldots, n_1; \ldots; j_d = 1, \ldots, n_d\}.$$

$\|A\|_{H,2}$ is often denoted $\|A\|_F$ and called the **Frobenius norm** or **Hilbert–Schmidt norm** of the hypermatrix $A \in \mathbb{C}^{n_1 \times \cdots \times n_d}$.

For hypermatrices, the **induced norm**, **natural**, or **operator norm** is defined for $p, q, r \in [1, \infty]$, by the quotient

$$\|A\|_{p,q,r} := \max_{\mathbf{x},\mathbf{y},\mathbf{z} \neq \mathbf{0}} \frac{|A(\mathbf{x}, \mathbf{y}, \mathbf{z})|}{\|\mathbf{x}\|_p \|\mathbf{y}\|_q \|\mathbf{z}\|_r} \tag{15.30}$$

where

$$A(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum\nolimits_{i,j,k=1}^{l,m,n} a_{ijk} x_i y_j z_{j_k}.$$

The special case $p = q = r = 2$, i.e., $\|\cdot\|_{2,2,2}$ is the **spectral norm**.

Let $A \in \mathbb{C}^{l \times m \times n}$. The **nuclear norm** or **Schatten 1-norm** of A is defined as

$$\|A\|_{S,1} := \min\Big\{\sum\nolimits_{i=1}^{r} |\lambda_i| : A = \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i, \\ \|\mathbf{u}_i\|_2 = \|\mathbf{v}_i\|_2 = \|\mathbf{w}_i\|_2 = 1,\ r \in \mathbb{N}\Big\}. \tag{15.31}$$

For hypermatrices $A \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ of arbitrary order, use the obvious generalizations, and of course for $p = \infty$, replace the sum by $\max\{|\lambda_1|, \ldots, |\lambda_r|\}$.

The inner product on $\mathbb{C}^{n_1 \times \cdots \times n_d}$ obtained from the dot product on $\mathbb{C}^{n_1 \cdots n_d}$ by viewing hypermatrices as complex vectors of dimension $n_1 \cdots n_d$, the dot product will be denoted by

$$\langle A, B \rangle := \sum\nolimits_{j_1,\ldots,j_d=1}^{n_1,\ldots,n_d} a_{j_1\cdots j_d} \bar{b}_{j_1\cdots j_d}. \tag{15.32}$$

Facts:

Facts requiring proof for which no specific reference is given can be found in [DF93, Chap. I] as well as [Lim] and the references therein.

1. The hypermatrix Hölder p-norm $\|A\|_{H,p}$ of $A \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ is the (vector Hölder) p-norm of A when regarded as a vector of dimension $n_1 \cdots n_d$ (see Section 50.1).
2. The Hölder p-norm has the property of being multiplicative on rank-1 hypermatrices in the following sense:

$$\|\mathbf{u} \otimes \mathbf{v} \otimes \cdots \otimes \mathbf{z}\|_{H,p} = \|\mathbf{u}\|_p \|\mathbf{v}\|_p \cdots \|\mathbf{z}\|_p,$$

   where $\mathbf{u} \in \mathbb{C}^{n_1}, \mathbf{v} \in \mathbb{C}^{n_2}, \ldots, \mathbf{z} \in \mathbb{C}^{n_d}$ and the norms on these are the usual p-norms on vectors.
3. The Frobenius norm is the norm (length) in the inner product space $\mathbb{C}^{n_1 \times \cdots \times n_d}$ with Eq. (15.32).
4. The inner product (15.32) (and thus the Frobenius norm) is invariant under multilinear matrix multiplication by unitary matrices,

$$\langle (Q_1, \ldots, Q_d) \cdot A, (Q_1, \ldots, Q_d) \cdot B \rangle = \langle A, B \rangle, \quad \|(Q_1, \ldots, Q_d) \cdot A\|_F = \|A\|_F$$

   for any $Q_1 \in \mathrm{U}(n_1), \ldots, Q_d \in \mathrm{U}(n_d)$.
5. The inner product (15.32) satisfies a Cauchy-Schwarz inequality

$$|\langle A, B \rangle| \leq \|A\|_F \|B\|_F,$$

and more generally a Hölder inequality

$$|\langle A, B\rangle| \leq \|A\|_{H,p}\|B\|_{H,q}, \quad \frac{1}{p}+\frac{1}{q}=1, \quad p,q \in [1,\infty].$$

6. The induced (p,q,r)-norm is a norm on $\mathbb{C}^{l\times m\times n}$ (with *norm* as defined in Section 50.1).
7. Alternate expressions for the induced norms include

$$\begin{aligned}\|A\|_{p,q,r} &= \max\{|A(\mathbf{u},\mathbf{v},\mathbf{w})| : \|\mathbf{u}\|_p = \|\mathbf{v}\|_q = \|\mathbf{w}\|_r = 1\}\\ &= \max\{|A(\mathbf{x},\mathbf{y},\mathbf{z})| : \|\mathbf{x}\|_p \leq 1,\ \|\mathbf{y}\|_q \leq 1,\ \|\mathbf{z}\|_r \leq 1\}.\end{aligned}$$

8. The induced norm is multiplicative on rank-1 hypermatrices

$$\|\mathbf{u}\otimes\mathbf{v}\otimes\mathbf{w}\|_{p,q,r} = \|\mathbf{u}\|_p\|\mathbf{v}\|_q\|\mathbf{w}\|_r$$

for $\mathbf{u}\in\mathbb{C}^l, \mathbf{v}\in\mathbb{C}^m, \mathbf{w}\in\mathbb{C}^n$.
9. For square matrices $A\in\mathbb{C}^{n\times n}$ and $\frac{1}{p}+\frac{1}{q}=1$,

$$\|A\|_{p,q} = \max_{\mathbf{x},\mathbf{y}\neq\mathbf{0}} \frac{|A(\mathbf{x},\mathbf{y})|}{\|\mathbf{x}\|_p\|\mathbf{y}\|_q} = \max_{\mathbf{x},\mathbf{y}\neq\mathbf{0}} \frac{|\mathbf{x}^T A\mathbf{y}|}{\|\mathbf{x}\|_p\|\mathbf{y}\|_q} = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_q}{\|\mathbf{x}\|_p},$$

the matrix (p,q)-norm.
10. The spectral norm is invariant under mutilinear matrix multiplication

$$\|(Q_1,\dots,Q_d)\cdot A\|_{2,\dots,2} = \|A\|_{2,\dots,2}$$

for any $Q_1\in\mathrm{U}(n_1),\dots,Q_d\in\mathrm{U}(n_d)$.
11. When $d=2$, the nuclear norm defined above reduces to the nuclear norm (i.e., Schatten 1-norm) of a matrix $A\in\mathbb{C}^{m\times n}$ defined more commonly by

$$\|A\|_{S,1} = \sum_{i=1}^{\max\{m,n\}} \sigma_i(A).$$

12. [DF93] The nuclear norm defines a norm on $\mathbb{C}^{n_1\times\cdots\times n_d}$ (with *norm* as defined in Section 50.1).
13. The nuclear norm is invariant under mutilinear unitary matrix multiplication, i.e.,

$$\|(Q_1,\dots,Q_d)\cdot A\|_{S,1} = \|A\|_{S,1}$$

for any $Q_1\in\mathrm{U}(n_1),\dots,Q_d\in\mathrm{U}(n_d)$.
14. The nuclear norm and the spectral norm are dual norms to each other, i.e.,

$$\begin{aligned}\|A\|_{S,1} &= \max\{|\langle A,B\rangle| : \|B\|_{2,\dots,2}=1\},\\ \|A\|_{2,\dots,2} &= \max\{|\langle A,B\rangle| : \|B\|_{S,1}=1\}.\end{aligned}$$

15. Since $\mathbb{C}^{n_1\times\cdots\times n_d}$ is finite dimensional, all norms are necessarily equivalent (see Section 50.1). Given that all norms induce the same topology, questions involving convergence of sequences of hypermatrices, whether a set of hypermatrices is closed, etc., are independent of the choice of norms.

Examples:

1. [Der13a] Let T_n be the matrix multiplication tensor in Eq. (15.19). Then $\|T_n\|_F = n^{3/2}$, $\|T_n\|_{S,1} = n^3$, and $\|T_n\|_{2,2,2} = 1$.

15.9 Hyperdeterminants

There are two ways to extend the determinant of a matrix to hypermatrices of higher order. One is to simply extend the usual expression of an $n \times n$ matrix determinant as a sum of $n!$ monomials in the entries of the matrix, which we will call the *combinatorial hyperdeterminant*, and the other, which we will call the *geometric hyperdeterminant*, is by using the characterization that a matrix has $\det A = 0$ if and only if $A\mathbf{x} = 0$ has nonzero solutions. Both approaches were proposed by Cayley [Cay49, Cay45], who also gave the explicit expression of a $2 \times 2 \times 2$ geometric hyperdeterminant. Gelfand, Kapranov, and Zelevinsky [GKZ94, GKZ92] have a vast generalization of Cayley's result describing the dimensions of hypermatrices for which geometric hyperdeterminants exist.

Definitions:

The **combinatorial hyperdeterminant** of a cubical d-hypermatrix $A = [a_{i_1 i_2 \cdots i_d}] \in F^{n\times\cdots\times n}$ is defined as

$$\det(A) = \frac{1}{n!} \sum\nolimits_{\pi_1,\ldots,\pi_d \in S_n} \operatorname{sgn}\pi_1 \cdots \operatorname{sgn}\pi_d \prod_{i=1}^{n} a_{\pi_1(i)\cdots\pi_d(i)}. \tag{15.33}$$

If it exists for a given set of dimensions $(n_1, \ldots, n_d)$, a **geometric hyperdeterminant** or **hyperdeterminant** $\operatorname{Det}_{n_1,\ldots,n_d}(A)$ of $A \in \mathbb{C}^{n_1\times\cdots\times n_d}$ is a homogeneous polynomial in the entries of A such that $\operatorname{Det}_{n_1,\ldots,n_d}(A) = 0$ if and only if the system of multilinear equations $\nabla A(\mathbf{x}_1, \ldots, \mathbf{x}_d) = 0$ has a nontrivial solution, i.e., $\mathbf{x}_1, \ldots, \mathbf{x}_d$ all nonzero. Here $\nabla A(\mathbf{x}_1, \ldots, \mathbf{x}_d)$ is the gradient of $A(\mathbf{x}_1, \ldots, \mathbf{x}_d)$ viewed as a function of the coordinates of the vector variables $\mathbf{x}_1, \ldots, \mathbf{x}_d$ (see Preliminaries for the definition of gradient). See Fact 4 for conditions that describe the $n_1, \ldots, n_d$ for which $\operatorname{Det}_{n_1,\ldots,n_d}$ exists.

Facts:

Facts requiring proof for which no specific reference is given may be found in [GKZ94, Chap. 13].

1. For odd order d, the combinatorial hyperdeterminant of a cubical d-hypermatrix is identically zero.
2. For even order d, the combinatorial hyperdeterminant of a cubical d-hypermatrix $A = [a_{i_1 i_2 \cdots i_d}]$ is

$$\det(A) = \sum\nolimits_{\pi_2,\ldots,\pi_d \in S_n} \operatorname{sgn}(\pi_2 \cdots \pi_d) \prod_{i=1}^{n} a_{i\pi_2(i)\cdots\pi_d(i)}.$$

   For $d = 2$, this reduces to the usual expression for the determinant of an $n \times n$ matrix.
3. Taking the π-transpose of an order d-hypermatrix $A \in F^{n\times\cdots\times n}$ leaves the combinatorial hyperdetermiant invariant, i.e., $\det A^\pi = \det A$ for and $\pi \in S_d$.
4. (Gelfand–Kapranov–Zelevinsky) A geometric hyperdeterminant exists for $\mathbb{C}^{n_1\times\cdots\times n_d}$ if and only if for all $k = 1, \ldots, d$, the following *dimension condition* is satisfied:

$$n_k - 1 \leq \sum\nolimits_{j \neq k} (n_j - 1).$$

5. When the dimension condition in Fact 4 is met, a geometric hyperdeterminant $\operatorname{Det}_{n_1,\ldots,n_d}(A)$ is a multivariate homogeneous polynomial in the entries of A that is unique up to a scalar multiple, and can be scaled to have integer coefficients with greatest common divisor one (the latter is referred to as *the* geometric hyperdeterminant and is unique up to ± 1 multiple).

6. The geometric hyperdeterminant exists for all cubical hypermatrices (and by definition is nontrivial when it exists, including for odd order cubical hypermatrices). The geometric hyperdeterminant also exists for some non-cubical hypermatrices (see Fact 3).
7. For $\mathbb{C}^{m\times n}$, the dimension condition in Fact 4 is $m \leq n$ and $n \leq m$, which may be viewed as a reason why matrix determinants are only defined for square matrices.
8. It is in general nontrivial to find an explicit expression for $\mathrm{Det}_{n_1,\dots,n_d}(A)$. While systematic methods for finding it exists, notably one due to Schläfli [GKZ94], an expression may nonetheless contain a large number of terms when expressed as a sum of monomials. For example, $\mathrm{Det}_{2,2,2,2}(A)$ has more than 2.8 million monomial terms [GHS08] even though a hypermatrix $A \in \mathbb{C}^{2\times 2\times 2\times 2}$ has only 16 entries.
9. Unlike rank and border rank (cf. Facts 15.3.2 and 15.4.5), the geometric hyperdeterminant is not invariant under multilinear matrix multiplication by nonsingular matrices (which is expected since ordinary matrix determinant is not invariant under left and right multiplications by nonsingular matrices either). However, it is *relatively invariant* in the following sense [GKZ94].

   Let $n_1,\dots,n_d$ satisfy the dimension condition in Fact 4. Then for any $A \in \mathbb{C}^{n_1\times\cdots\times n_d}$ and any $X_1 \in \mathrm{GL}(n_1,\mathbb{C}),\dots,X_d \in \mathrm{GL}(n_d,\mathbb{C})$,

   $$\mathrm{Det}_{n_1,\dots,n_d}((X_1,\dots,X_d)\cdot A) = \det(X_1)^{m/n_1}\cdots\det(X_d)^{m/n_d}\,\mathrm{Det}_{n_1,\dots,n_d}(A) \quad (15.34)$$

   where m is the degree of $\mathrm{Det}_{n_1,\dots,n_d}$. Hence for $X_1 \in \mathrm{SL}(n_1,\mathbb{C}),\dots,X_d \in \mathrm{SL}(n_d,\mathbb{C})$, we get

   $$\mathrm{Det}_{n_1,\dots,n_d}((X_1,\dots,X_d)\cdot A) = \mathrm{Det}_{n_1,\dots,n_d}(A).$$

10. A consequence of Fact 9 is the properties of the usual matrix determinant under row/column interchanges, addition of a scalar multiple of row/column to another, etc., are also true for the hyperdeterminant. For notational convenience, we shall just state the following corollary for a 3-hypermatrix although it is true in general for any d-hypermatrix [GKZ94]. We use the notation in Example 15.1.3 where $A_1,\dots,A_n$ denote the slices of the 3-hypermatrix A, analogues of rows or columns of a matrix. Let l, m, n satisfy the dimension condition in Fact 4 and let $A = [A_1|\dots|A_n] \in \mathbb{C}^{l\times m\times n}$. Then

    (a) interchanging two slices leaves the hyperdeterminant invariant up to sign:

    $$\mathrm{Det}_{l,m,n}([A_1|...|A_i|...|A_j|...|A_n]) = \pm\,\mathrm{Det}_{l,m,n}([A_1|...|A_j|...|A_i|...|A_n]);$$

    (b) adding a scalar multiple of a slice to another leaves the hyperdeterminant invariant:

    $$\mathrm{Det}_{l,m,n}([A_1|...|A_i|...|A_j|...|A_n]) = \mathrm{Det}_{l,m,n}([A_1|...|A_i+\alpha A_j|...|A_j|...|A_n])$$

    for any $\alpha \in \mathbb{C}$;

    (c) having two proportional slices implies that the hyperdetermiant vanishes:

    $$\mathrm{Det}_{l,m,n}([A_1|...|A_i|...|\alpha A_i|...|A_n]) = 0 \qquad \text{for any } \alpha \in \mathbb{C};$$

    (d) taking π-transpose leaves the hyperdetermiant invariant

    $$\mathrm{Det}_{l,m,n}(A^\pi) = \mathrm{Det}_{l,m,n}(A) \qquad \text{for any } \pi \in S_3.$$

    In particular, the last property implies that the other three properties hold for slices of A in any other fixed index.
11. $\mathrm{Det}_{n_1,\dots,n_d}(A) = 0$ if and only if 0 is a singular value of A.

Examples:

1. For $A = [a_{ijk\ell}] \in F^{2\times 2\times 2\times 2}$, the combinatorial hyperdeterminant of A is

$$\det A = a_{1111}a_{2222} - a_{1112}a_{2221} - a_{1121}a_{2212} + a_{1122}a_{2211} \\ - a_{1211}a_{2122} + a_{1122}a_{2211} + a_{1221}a_{2112} + a_{1212}a_{2121}.$$

This is immediate from Fact 2, or by applying the definition (15.33) and simplifying.

2. [Cay45] The geometric hyperdeterminant of $A = [A_1|A_2] = [a_{ijk}] \in \mathbb{C}^{2\times 2\times 2}$ (where A_1 and A_2 slices of A, cf. Section 15.1) is

$$\begin{aligned}\mathrm{Det}_{2,2,2}(A) &= \frac{1}{4}[\det(A_1 + A_2) - \det(A_1 - A_2)]^2 - 4\det[A_1]\det[A_2] \\ &= \frac{1}{4}\left[\det\left(\begin{bmatrix} a_{111} & a_{112} \\ a_{121} & a_{122}\end{bmatrix} + \begin{bmatrix} a_{211} & a_{221} \\ a_{212} & a_{222}\end{bmatrix}\right) - \det\left(\begin{bmatrix} a_{111} & a_{121} \\ a_{112} & a_{122}\end{bmatrix} - \begin{bmatrix} a_{211} & a_{221} \\ a_{212} & a_{222}\end{bmatrix}\right)\right]^2 \\ &\qquad - 4\det\begin{bmatrix} a_{111} & a_{121} \\ a_{112} & a_{122}\end{bmatrix}\det\begin{bmatrix} a_{211} & a_{221} \\ a_{212} & a_{222}\end{bmatrix}.\end{aligned}$$

The statement that the gradient $\nabla A(\mathbf{x}, \mathbf{y}, \mathbf{z})$ of the trilinear functional defined by A vanishes for some nonzero vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ is the statement that the system of bilinear equations,

$$\begin{aligned} a_{111}x_1y_1 + a_{121}x_1y_2 + a_{211}x_2y_1 + a_{221}x_2y_2 &= 0, \\ a_{112}x_1y_1 + a_{122}x_1y_2 + a_{212}x_2y_1 + a_{222}x_2y_2 &= 0, \\ a_{111}x_1z_1 + a_{112}x_1z_2 + a_{211}x_2z_1 + a_{212}x_2z_2 &= 0, \\ a_{121}x_1z_1 + a_{122}x_1z_2 + a_{221}x_2z_1 + a_{222}x_2z_2 &= 0, \\ a_{111}y_1z_1 + a_{112}y_1z_2 + a_{121}y_2z_1 + a_{122}y_2z_2 &= 0, \\ a_{211}y_1z_1 + a_{212}y_1z_2 + a_{221}y_2z_1 + a_{222}y_2z_2 &= 0, \end{aligned}$$

has a non-trivial solution ($\mathbf{x}, \mathbf{y}, \mathbf{z}$ all non-zero).

3. The geometric hyperdeterminant of $A = [a_{ijk}] \in \mathbb{C}^{2\times 2\times 3}$ is

$$\begin{aligned}\mathrm{Det}_{2,2,3}(A) &= \det\begin{bmatrix} a_{111} & a_{112} & a_{113} \\ a_{211} & a_{212} & a_{213} \\ a_{121} & a_{122} & a_{123}\end{bmatrix}\det\begin{bmatrix} a_{211} & a_{212} & a_{213} \\ a_{121} & a_{122} & a_{123} \\ a_{221} & a_{222} & a_{223}\end{bmatrix} \\ &\qquad - \det\begin{bmatrix} a_{111} & a_{112} & a_{113} \\ a_{211} & a_{212} & a_{213} \\ a_{221} & a_{222} & a_{223}\end{bmatrix}\det\begin{bmatrix} a_{111} & a_{112} & a_{113} \\ a_{121} & a_{122} & a_{123} \\ a_{221} & a_{222} & a_{223}\end{bmatrix}.\end{aligned}$$

The statement that the gradient $\nabla A(\mathbf{x}, \mathbf{y}, \mathbf{z})$ of the trilinear functional defined by A vanishes for some nonzero vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ is the statement that the system of bilinear equations,

$$\begin{aligned} a_{111}x_1y_1 + a_{121}x_1y_2 + a_{211}x_2y_1 + a_{221}x_2y_2 &= 0, \\ a_{112}x_1y_1 + a_{122}x_1y_2 + a_{212}x_2y_1 + a_{222}x_2y_2 &= 0, \\ a_{113}x_1y_1 + a_{123}x_1y_2 + a_{213}x_2y_1 + a_{223}x_2y_2 &= 0, \\ a_{111}x_1z_1 + a_{112}x_1z_2 + a_{113}x_1z_3 + a_{211}x_2z_1 + a_{212}x_2z_2 + a_{213}x_2z_3 &= 0, \\ a_{121}x_1z_1 + a_{122}x_1z_2 + a_{123}x_1z_3 + a_{221}x_2z_1 + a_{222}x_2z_2 + a_{223}x_2z_3 &= 0, \\ a_{111}y_1z_1 + a_{112}y_1z_2 + a_{113}y_1z_3 + a_{121}y_2z_1 + a_{122}y_2z_2 + a_{123}y_2z_3 &= 0, \\ a_{211}y_1z_1 + a_{212}y_1z_2 + a_{213}y_1z_3 + a_{221}y_2z_1 + a_{222}y_2z_2 + a_{223}y_2z_3 &= 0, \end{aligned}$$

has a non-trivial solution ($\mathbf{x}, \mathbf{y}, \mathbf{z}$ all non-zero).

Applications:

1. In Example 15.3.4, we saw how one may characterize the notion of a pure state in terms of the rank of a hypermatrix, namely $\operatorname{rank}(A) = 1$. The hyperdeterminant characterizes the dual notion — the *degenerate entangled states* are precisely the hypermatrices $A \in \mathbb{C}^{2\times 2\times 2}$ with $\operatorname{Det}_{2,2,2}(A) = 0$ [MW02]. More generally, when $\operatorname{Det}_{2,2,2}(A) \neq 0$, the magnitude of the hyperdeterminant $|\operatorname{Det}_{2,2,2}(A)|$ gives another way to measure the amount of entanglement called the *concurrence* [HW97].

15.10 Odds and Ends

Because of page constraints, many topics have to be omitted from this chapter: covariance and contravariance, tensor fields and hypermatrix-valued functions, symmetric tensors and symmetric hypermatrices, alternating tensors and skew-symmetric hypermatrices, eigenvalues and singular values of hypermatrices, positive definite hypermatrices and Cholesky decomposition, nonnegative hypermatrices and the Perron-Frobenius theorem, the Dirac and Einstein notations, tensor products of modules, of Hilbert and Banach spaces, of functions, of operators, of representations, etc.

Also, topics that are best presented in a usual narrative format (as opposed to a handbook format) are not included in the chapter. These include discussions of the difference between hypermatrices and tensors, the existence of canonical forms for hypermatrices, the duality between the geometric hyperdeterminant and tensor rank, questions relating to computability and complexity, as well as connections to algebraic geometry and representation theory.

As a consequence, there are many additonal interesting applications and examples that are not included in this chapter because they require one or more of these omitted topics. These include self-concordance and higher optimality conditions in optimization theory, Riemann curvature tensor and the Einstein field equations in general relativity, electromagnetic field tensor in gauge theory, linear piezoelectric equations in continuum mechanics, Fourier coefficients of triply-periodic functions in X-ray crystallography, other forms of the Yang–Baxter equation, density matrix renormalization group (DMRG) techniques in quantum chemistry, the Salmon conjecture (now a theorem) in phylogenetics, moments and cumulants in statistics, polynomial Mercer kernels and naïve Bayes model in machine learning, and blind source separation and independent components analysis in signal processing.

We refer the reader to [Lim] for an expanded treatment that covers these and other topics.

Acknowledgment

The author would like to express his heartfelt gratitude to Leslie Hogben for her tireless editorial efforts, transforming a manuscript that is an earlier version of [Lim] into the HLA format that one sees here. It is an arduous undertaking without which this chapter would not have made it in time for inclusion in this volume. The author would also like to thank Harm Derksen, Shenglong Hu, J. M. Landsberg, Peter McCullagh, and Ke Ye for their very helpful comments.

References

[AW92] W. A. Adkins and S. H. Weintraub. *Algebra: An Approach Via Module Theory.* Springer-Verlag, New York, 1992.

[Bax78] R. J. Baxter. Solvable eight-vertex model on an arbitrary planar lattice. *Philos. Trans.*

Roy. Soc. London Ser. A, 289 (1359): 315–346, 1978.

[Ber69] G. M. Bergman, Ranks of tensors and change of base field, *J. Algebra*, 11: 613–621, 1969.

[BLR80] D. Bini, G. Lotti, and F. Romani. Approximate solutions for the bilinear form computational problem. *SIAM J. Comput.*, 9 (4): 692–697, 1980.

[Bor90] S. F. Borg. *Matrix-Tensor Methods in Continuum Mechanics*, 2nd ed. World Scientific, Singapore, 1990.

[Bou98] N. Bourbaki. *Algebra I: Chapters 1–3*. Springer-Verlag, Berlin, 1998.

[Bry02] J.-L. Brylinski. Algebraic measures of entanglement. In R. K. Brylinski and G. Chen (Eds.), *Mathematics of Quantum Computation*, pp. 3–23, CRC Press, Boca Raton, FL, 2002.

[BCS96] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic Complexity Theory*, Springer-Verlag, Berlin, 1996.

[CO65] T. K. Caughey, M. E. J. O'Kelly. Classical normal modes in damped linear dynamic systems. *ASME J. Applied Mechanics*, 32: 583—588, 1965.

[Cay45] A. Cayley. On the theory of linear transformations. *Cambridge Math. J.*, 4: 193–209, 1854.

[Cay49] A. Cayley. On the theory of determinants. *Trans. Cambridge Philos. Soc.*, 8, no. 7: 75–88, 1849.

[Cor84] J. F. Cornwell. *Group Theory in Physics*, Vol. II. Academic Press, London, 1984.

[DF93] A. Defant and K. Floret. *Tensor Norms and Operator Ideals*. North-Holland, Amsterdam, 1993.

[Der13] H. Derksen. Sharpness of Kruskal's theorem. *Linear Algebra Appl.*, 438(2): 708–712, 2013.

[Der13a] H. Derksen. Personal communication, 2013.

[DE03] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proc. Nat. Acad. Sci.*, 100(5): 2197–2202, 2003.

[DVC00] W. Dür, G. Vidal, and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A*, 62(6), 062314, 12 pp., 2000.

[GKZ94] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky. *Discriminants, Resultants, and Multidimensional Determinants*. Birkhäuser Publishing, Boston, MA, 1994.

[GKZ92] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky. Hyperdeterminants. *Adv. Math.*, 96(2): 226–263, 1992.

[GHZ89] D. M. Greenberger, M. A. Horne, and A. Zeilinger. Going beyond Bell's theorem, In M. Kafatos (Ed.), *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, pp. 69–72, Kluwer, Dordrecht, Netherlands, 1989.

[Gre78] W. Greub. *Multilinear Algebra*, 2nd ed. Springer-Verlag, New York, 1978.

[GHS08] D. Grier, P. Huggins, B. Sturmfels, and J. Yu. The hyperdeterminant and triangulations of the 4-cube. *Math. Comp.*, 77(263): 1653–1679, 2008.

[Hal85] P. Halmos. *I Want to Be a Mathematician... An Automathography*. Springer-Verlag, New York, 1985.

[HW97] S. Hill and W. K. Wootters. Entanglement of a pair of quantum bits. *Phys. Rev. Lett.*, 78(26): 5022–5025, 1997.

[Hit27a] F. L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys.*, 6(1): 164–189, 1927.

[Hit27b] F. L. Hitchcock. Multiple invariants and generalized rank of a p-way matrix or tensor. *J. Math. Phys.*, 7(1): 39–79, 1927.

[HK71] J. E. Hopcroft and L. R. Kerr. On minimizing the number of multiplications necessary for matrix multiplication. *SIAM J. Appl. Math.*, 20(1): 30–36, 1971.

[Knu98] D. Knuth. *The Art of Computer Programming 2: Seminumerical Algorithms*, 3rd ed. Addison–Wesley, Reading, MA, 1998.

[KM97] A. I. Kostrikin and Y. I. Manin. *Linear Algebra and Geometry.* Gordon and Breach, Amsterdam, 1997.

[Kru77] J. B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Lin. Alg. Appl.*, 18(2): 95–138, 1977.

[Lan06] J. M. Landsberg. The border rank of the multiplication of 2×2 matrices is seven. *J. Amer. Math. Soc.*, 19(2): 447–459, 2006.

[Lan12] J. M. Landsberg. *Tensors: Geometry and Applications.* AMS, Providence, RI, 2012.

[Lan02] S. Lang. *Algebra*, Rev. 3rd ed. Springer-Verlag, New York, 2002.

[Lim] L.-H. Lim. Tensors and hypermatrices. Preprint, `http://www.stat.uchicago.edu/~lekheng/work/tensors.pdf`.

[Mar23] M. Marcus. *Finite Dimensional Multilinear Algebra*, Parts I and II. Marcel Dekker, New York, 1973 and 1975.

[MW02] A. Miyake and M. Wadati. Multipartite entanglement and hyperdeterminant. *Quantum Inf. Comput.*, 2: 540–555, 2002.

[Nor84] D. G. Northcott. *Multilinear Algebra.* Cambridge University Press, Cambridge, 1984.

[Oxl11] J. G. Oxley. *Matroid Theory*, 2nd ed. Oxford University Press, Oxford, 2011.

[SB00] N. D. Sidiropoulos and R. Bro. On the uniqueness of multilinear decomposition of N-way arrays. *J. Chemometrics*, 14(3): 229–239, 2000.

[Str69] V. Strassen. Gaussian elimination is not optimal. *Numer. Math.*, 13(4): 354–356, 1969.

[Str83] V. Strassen. Rank and optimal computation of generic tensors. *Lin. Alg. Appl.*, 52/53: 645–685, 1983.

[Str73] V. Strassen. Vermeidung von Divisionen. *J. Reine Angew. Math.*, 264: 184–202, 1973.

[Win71] S. Winograd. On multiplication of 2×2 matrices, *Lin. Alg. Appl.*, 4(4): 381–388, 1971.

[Yan67] C. N. Yang. Some exact results for the many-body problem in one dimension with repulsive delta-function interaction. *Phys. Rev. Lett.*, 19(23): 1312–1315, 1967.

[Yok92] T. Yokonuma. *Tensor Spaces and Exterior Algebra.* AMS, Providence, RI, 1992.