Tensors and Some Applications to Networks

Ariel Chenet
Supervisors: B.Mossé - É. Remy (I2M)

June 2020

Contents

Introduction

This memoir first develops the mathematical framework of tensors, including several original responses to problems posed in [3], and then uses this framework to attempt to study the types of large networks that are commonly found in biological data.

Chapter 1

Basic algebra

In this chapter will introduce a few of the key concepts which we'll use to define tensors.

1.1 Modules

1.1.1 Generalities

In this section we will define modules and describe some of their properties. Intuitively, a module is like a vector space, but the scalars are from a ring that isn't necessarily a field. We will now proceed to formally define a module, starting with the definition of some more basic algebraic structures.

Definition 1. A **group** is a set of elements G and an internal composition law between those elements '.' such that

1. *the law . is associative*: for any three elements x, y, z in G, we have $(x.y).z = x.(y.z)$;
2. *the law . admits a neutral element*: there is a element e in G such that for all x in G, $e.x = x.e = x$;
3. *each x in G is invertible*: for every element x in G there is a element x^{-1} in G, called the inverse of x, such that $x.x^{-1} = x^{-1}.x = e$.

A group is called *commutative* or *abelian* if $x.y = y.x$ for all x and y in G. Please note that the internal composition law is not always noted by '.' - other symbols are often used. In particular, the internal composition law of an abelian group is sometimes noted '+'; in this case the neutral element is noted 0, and the inverse of x becomes its opposite $-x$.

Definition 2. Let N be a set of n elements. Then the group of bijective functions from N to N with the composition of functions as internal composition law, will be called the **permutation group** or the **symmetric group** of n elements, and noted $\mathfrak{S}_n$.

We consider that there is only one symmetric group on n elements as we can create a bijection between any set of n elements. As such, we usually consider that $N = \{1, \ldots, n\}$.

Definition 3. A **ring** $(A, +, \times)$ is an abelian group that has a second internal composition law, generally called multiplication and frequently noted '$\times$', associative and also distributive over the first one, which means that for all a, b and c in A we have that

$$a \times (b + c) = (a \times b) + (a \times c) \ ,$$

$$(b + c) \times a = (b \times a) + (c \times a) \ .$$

If the ring has a neutral element for its second law, this element is generally noted 1 and the ring is said to be *unitary*. We will almost exclusively consider unitary rings in the rest of this text, and a ring should be assumed to be unitary unless it is explicitly said otherwise.

A ring with a commutative multiplication group is called a *commutative* ring.

Definition 4. An element a of a ring A is called **invertible** if it has a multiplicative inverse, in other words, if there is an a^{-1} in A such that $a \times a^{-1} = a^{-1} \times a = 1$.

Definition 5. A **field** is a unitary ring with $1 \neq 0$ such that all non zero elements are invertible.

Now we can define modules. Once again, the general intuition of "a module is a vector space with the scalars coming from a ring that is not necessarily a field" may be useful to keep in mind. The switch from a field of scalars to a ring of scalars can lead to the loss of some of the nice properties of vector spaces.

In all of the following definitions, we will be using a unitary ring $(A, +, \times)$.

We can now define left and right modules.

Definition 6. A **left A-module** $(M, +, .)$ is an abelian group on which is defined an operation $A \times M \longrightarrow M$ called scalar multiplication, that respects the following properties for all a, b in A and for all m, n in M :

- $a(m + n) = am + an$,
- $(a + b)m = am + bm$,
- $a(bm) = (ab)m$,
- $1_A m = m$.

If we instead define scalar multiplication on the right, then we have a right A-module. If $(A, \times)$ is commutative, both are the same. We will exclusively consider left modules in the rest of this text.

From this point on in the text, we may occasionally refer to an A-module as simply a module, and omit the reference to the scalar ring.

Definition 7. If M is an A-module and N is a subset of M, we call N a **submodule** of M if it is a subset of M that is an A-module for the addition and scalar multiplication induced by the ones of M.

If N is a submodule of M, then N is non empty, and we have that for all n, n' in N, and for all scalars a in A, $an \in N$ and and $n + n' \in N$. This is actually a characterisation of submodules: if we have a module M and we want to verify that a subset $N \subseteq M$ is a submodule, we simply verify that it is non empty and closed under addition and scalar multiplication.

Proposition 1. *If M is an A-module and N a submodule of M, the binary relation $\sim_N$ on M defined by $x \sim_N y$ if there exists an element n of N such that $x + n = y$ is an equivalence relation. Let us define the* ***quotient*** *of M by N, usually noted M/N, as the set of equivalence classes under this relation. Then the operations of M induce operations on the quotient that give to M/N a structure of A-module. In particular, elements of N are sent to zero by the projection from M to M/N.*

Definition 8. A **linear combination** of $m_1, ..., m_n$, where all the m_i are elements of an A-module M is simply an element of M that can be written as $\lambda_1 m_1 + ... + \lambda_n m_n$, with $\lambda_1, .., \lambda_n$ being elements of the scalar ring A.

Definition 9. Some elements m_i, with $i \in I$, of M are said to be **linearly independent** if the only possible finite linear combination of these elements equal to zero is the one where all scalar coefficients are equal to zero. In other words (or rather, symbols), for any finite subset $\{i_1, \ldots, i_r\}$ of I,

$$\sum_{1 \leq k \leq r} \lambda_{i_k} m_{i_k} = 0 \implies \forall k \in \{1, \ldots, r\}, \quad \lambda_{i_k} = 0 \ .$$

Definition 10. The **direct product** of a family of modules $(M_i)_{i\in I}$ is the module with the underlying set $\prod_i M_i$ being the cartesian product of the underlying sets, the group law being determined termwise, ie $(m_i)_i + (n_i)_i = (m_i + n_i)_i$, and scalar multiplication being defined as $a(m_i)_i = (am_i)_i$.

Definition 11. The **direct sum** of a family $(M_i)_{i\in I}$ of modules, noted $\bigoplus_i M_i$ is the submodule of their direct product composed of elements who have only a finite number of components with non-zero values.

Definition 12. Let M an A-module and $(m_i)_{i\in I}$ a family of elements of M.

- If every element of M can be written as a finite linear combination of the m_i, then we call $\{m_i \ ; \ i \in I\}$ a **generating set** of M.
- If $\{m_i \ ; \ i \in I\}$ is generating and the m_i are linearly independent, the family $(m_i)_{i\in I}$ is called a **basis** of M. Or to put it another way, $(m_i)_{i\in I}$ is a basis if $M = \bigoplus_i Am_i$, where A is the ring of scalars.
- If M admits a basis indexed on a finite set I, then M is said **finitely generated**, or **of finite type**.

Definition 13. A module over a field is called a **vector space**.

Definition 14. Let V be a vector space over a field K. If V is equipped with a binary operation $* : V \times V \to V$ such that for all vectors x, y, z and all scalars α, β

$$\begin{aligned}
&(\alpha x + \beta y) * z = \alpha(x * z) + \beta(y * z),\\
&x * (\alpha y + \beta z) = \alpha(x * y) + \beta(y * z),\\
&(\alpha x) * (\beta y) = (\alpha\beta)(x * y),
\end{aligned}$$

then V becomes an **algebra over the field** K.

1.1.2 Free modules

In this section we define the free A-module over a set X, using a method detailed in [3].

We will let X be any non empty set, and A be a ring of scalars, and we will use the standard set-theoric notation A^X to denote the set of functions from X to A. Moreover, we will denote $A^{(X)}$ the subset of A^X consisting of functions from X to A with only a finite number of non-zero values, ie

$$A^{(X)} = \{u \in A^X \ ; \ \#\{x \in X \ : \ u(x) \neq 0\} < \infty\} \ .$$

We will show briefly that A^X is a module, then that $A^{(X)}$ is a submodule of A^X, give a basis of $A^{(X)}$ and show that it has a universal property.

Proposition 2. *The set A^X is a module under the operations given for $f, g \in A^X$ and $\lambda \in A$, by $\lambda f : x \mapsto \lambda f(x)$ and $f + g : x \mapsto f(x) + g(x)$.*

Proof. Because any ring is a module over itself, we have that $(A^X, +)$ is an abelian group with neutral element $0_{A^X} : x \mapsto 0_A$, and for all x in X,

$$\begin{aligned}
&\lambda(f + g)(x) = \lambda f(x) + \lambda g(x),\\
&(\lambda + \mu)f(x) = \lambda f(x) + \mu f(x),\\
&\lambda(\mu f(x)) = (\lambda\mu)f(x),\\
&1f(x) = f(x),
\end{aligned}$$

and hence the module axioms are verified. □

We now define for each x in X a corresponding function $e_x \in A^{(X)}$ which sends x to 1_A and every other element of X to 0_A.

Proposition 3. *$A^{(X)}$ is a submodule of A^X.*

Proof. It suffices to verify that $A^{(X)}$ is non empty and closed under scalar multiplication and addition. The family $(e_x)_{x\in X}$ is in $A^{(X)}$ as every element of that family has exactly one non-zero value. Furthermore, the sum of two functions with a finite number of non-zero values will also have a finite number of non-zero values, and multiplying by a scalar does not change the zeros of a function. Hence, $A^{(X)}$ is indeed a submodule of A^X. □

Proposition 4. *The $(e_x)_{x\in X}$ form a basis of $A^{(X)}$.*

Proof. We will first show that this family is linearly independent. Suppose that there exists a finite subset Y of X and $(\lambda_x)_{x\in Y}$ such that

$$\sum_{x\in Y} \lambda_x e_x = 0 \ .$$

This means that for every y in X, $\sum_{x\in Y} \lambda_x e_x(y) = 0$. But if we take $y = x$, then this must mean that λ_x is nul for every x in Y, and therefore the family $(e_x)_{x\in X}$ is indeed linearly independent.

To show that it is a generating family of $A^{(X)}$, take any u in the aforementioned set. It can be written as

$$u = \sum_{x\in X} u(x)e_x \ ,$$

(recall that $u(x) = 0$ apart from a finite numbers of values of x), and so the $(e_x)_{x\in X}$ do indeed form a basis of $A^{(X)}$. □

We now give a universal property of $A^{(X)}$.

Proposition 5. *If M is any A-module and f is any function from X to M, there exists a unique linear map $\hat{f}$ from $A^{(X)}$ to M such that for all x in X, $\hat{f}(e_x) = f(x)$.*

Proof. As $u(x)$ is a scalar of A for each x, we can simply define

$$\hat{f}(u) = \sum_{x\in X} u(x)e_x$$

for every u in $A^{(X)}$. We have existence by construction and uniqueness by linearity. This is illustrated in the following commutative diagram.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d] & M \\ A^{(X)} \arrow[ur, "\hat{f}"] & \end{tikzcd}$$

Figure 1.1: Commutatif diagram

□

We can now define the **free A-module on** X to be $A^{(X)}$. If we identify each x in X with the corresponding e_x in $A^{(X)}$, the free A-module on X can be thought of as the set of finite linear combinations of elements of X, and in particular X as a subset of $A^{(X)}$.

With this construction, if X and X' are two finite sets with the same cardinal, then they give isomorphic free modules.

Notation The free module over X will also sometimes be noted F_X.

Notice From now on, we will consider modules over a commutative unitary ring K.

1.2 Linear and multi-linear maps

1.2.1 Definitions and notations

Definition 15. A **K-linear map**, or more simply a **linear map** is a map l from one K-module M to another K-module N such that for all m, m' in M and for all a in K we have $l(m+m') = l(m) + l(m')$ and $l(am) = al(m)$.

Definition 16. A **K-bilinear map**, or more simply a **bilinear map** is a map ϕ from a cartesian product of K-modules $M \times N$ into another K-module R such that for all $(m,n), (m', n')$ in $M \times N$ and for all a in K, we have that $\phi(am, n) = a\phi(m,n) = \phi(m, an)$, $\phi(m+m', n) = \phi(m,n) + \phi(m', n)$ and $\phi(m, n+n') = \phi(m,n) + \phi(m,n')$.

We can generalize this to a cartesian product of n modules with multi-linear maps.

Definition 17. A **multi-linear** or **n-linear** map is a function with n variables that is linear in each one of it's variables. In other words, if $X_1, \dots X_n$ and Y are K-modules then $f : X_1 \times \dots X_n \longrightarrow Y$ is n-linear, if for all i in $\{1, \dots, n\}$, and for any set of vectors $v_1, \dots v_{i-1}, v_{i+1}, \dots, v_n$ in $X_1 \times \dots X_{i-1} \times X_{i+1} \times \dots \times X_n$ the function $x_i \mapsto f(v_1, \dots, v_{i-1}, x_i, v_{i+1}, \dots, v_n)$ is linear.

This definition is a generalization of the two previous ones, linear maps are n-linear maps for $n=1$, and bilinear maps are n-linear maps for $n=2$.

Like with linear maps and multi-linear maps, we can also define linear and multi-linear forms.

Definition 18. A linear map from an K-module M to its scalar ring K is called a **linear form**.

Definition 19. A **multi-linear form** or an **n-linear form** is a multi-linear or n-linear map from the cartesian product of K-modules $M_1 \times \dots \times M_n$ to their scalar ring K.

Notation Let $X, Y, X_1, \dots, X_n$ be modules over a commutative ring K. We will use the notation $\mathcal{L}(X;Y)$ to designate the set of linear maps from X to Y, $\mathcal{L}(X)$ to designate the set of linear maps from X to X and $\mathcal{L}(X_1, \dots, X_n; Y)$ to designate the set of n-linear maps from $X_1 \times \dots \times X_n$ to Y.

We note that all of the sets mentioned above are modules for the operations

$$f + g = x \mapsto f(x) + g(x)$$
$$\lambda f = x \mapsto \lambda f(x)$$

by virtue of X and Y being modules.

Proposition 6. *If M, N and Z are K-modules, then $\mathcal{L}(M,N;Z)$ is isomorphic to $\mathcal{L}(M, \mathcal{L}(N;Z))$.*

Proof. Let us consider the map

$$\tau : \mathcal{L}(M,N;Z) \longrightarrow \mathcal{L}(M;\mathcal{L}(N;Z)) \ ,$$

$$\phi \longmapsto u_\phi$$

where $u_\phi : M \longrightarrow \mathcal{L}(N;Z) \qquad m \longmapsto u_{\phi,m}$, and $u_{\phi,m} : N \longrightarrow Z$ is the map that takes $n \in N$ and sends it to $\phi(m,n)$: we have $u_{\phi,m}(n) = \phi(m,n)$.
The map $u_{\phi,m}$ is linear because ϕ is bilinear, and the linearities of u_ϕ and τ are immediate. Moreover, the equalities $u_{\phi,m}(n) = \phi(m,n)$ imply that u_ϕ determines ϕ, thus τ is an isomorphism. $\square$

1.2.2 A characterization of multi-linear forms on finitely generated free modules

One of the advantages of multi-linear forms on finitely generated free modules is that they are completely and uniquely determined by what the basis elements map to.

Theorem 1. *Let $X_1, \ldots, X_r$ (with $r \geq 0$) be finitely generated free modules over a commutative ring K, with $X_1, \ldots, X_r$ having for respective bases $(a_{1,i})_{1\leq i\leq n_1}, \ldots, (a_{r,i})_{1\leq i\leq n_r}$.*

A function $f : X_1 \times \cdots \times X_r \to K$ is r-linear if and only if there exists a family $(c_{i_1\ldots i_r})_{1\leq i_1\leq n_1,\ldots,1\leq i_r\leq n_r}$ of elements of K such that, for any tuple of vectors $(x_1, \ldots, x_r) \in X_1 \times \cdots \times X_r$, where each x_k in X_k is written as $\sum_{i=1}^{n_k} x_{k,i}a_{k,i}$, we have

$$f(x_1, \ldots, x_r) = \sum_{i_1,\ldots,i_r} x_{1,i_1} \ldots x_{r,i_r} c_{i_1\ldots i_r} ,$$

$$\text{and } c_{i_1\ldots i_r} = f(a_{1,i_1}, \ldots, a_{r,i_r}).$$

Proof. If f is multi-linear, then by definition we must have that

$$\begin{aligned} f(x_1, \ldots, x_r) &= f(\sum_{i_1} x_{1,i_1}a_{1,i_1}, \ldots, \sum_{i_r} x_{r,i_r}a_{r,i_r}) \\ &= \sum_{i_1,\ldots,i_r} x_{1,i_1} \ldots x_{r,i_r} f(a_{1,i_1}, \ldots, a_{r,i_r}) \\ &= \sum_{i_1,\ldots,i_r} x_{1,i_1} \ldots x_{r,i_r} c_{i_1\ldots i_r} , \end{aligned}$$

where $c_{i_1\ldots i_r} = f(a_{1,i_1}, \ldots, a_{r,i_r})$.

Conversely, suppose the existence of constants as described in the theorem. For any index $j \in \{1, \ldots, r\}$, we have for addition that

$$\begin{aligned} f(x_1, \ldots, x_j + x'_j, \ldots, x_r) &= \sum_{i_1,\ldots,i_r} x_{i_1} \ldots (x_{i_j} + x'_{i_j}) \ldots x_{i_r} c_{i_1\ldots i_r} \\ &= \sum_{i_1,\ldots,i_r} x_{i_1} \ldots x_{i_j} \ldots x_{i_r} c_{i_1\ldots i_r} + \sum_{i_1,\ldots,i_r} x_{i_1} \ldots x'_{i_j} \ldots x_{i_r} c_{i_1\ldots i_r} \\ &= f(x_1, \ldots, x_j, \ldots, x_r) + f(x_1, \ldots, x'_j, \ldots, x_r), \end{aligned}$$

and for scalar multiplication, we have that

$$\begin{aligned} f(x_1, \ldots, \lambda x_j, \ldots, x_r) &= \sum_{i_1,\ldots,i_r} x_{i_1} \ldots \lambda x_{i_j} \ldots x_{i_r} c_{i_1\ldots i_r} \\ &= \lambda \sum_{i_1,\ldots,i_r} x_{i_1} \ldots x_{i_j} \ldots x_{i_r} c_{i_1\ldots i_r} \\ &= \lambda f(x_1, \ldots, x_j, \ldots, x_r) , \end{aligned}$$

and therefore f is indeed multi-linear. □

The scalars $c_{i_1\ldots i_r}$ are sometimes referred to [1] as **structure constants** of f with regards to the bases $(a_{i_1})_{i_1}, \ldots, (a_{i_r})_{i_r}$. They are also sometimes referred to [3] as **coefficients** or **components** or **coordinates**.

This leads directly to the following result:

Theorem 2. *Let $X_1, \ldots, X_p$ be finitely generated free modules with respective bases $(a_{1,i})_{1\leq i\leq n_1}, \ldots, (a_{r,i})_{1\leq i\leq n_r}$. The functions*

$$\begin{aligned} v_{i_1\ldots i_p} : X_1 \times \cdots \times X_p &\to K \\ (x_1, \ldots, x_p) &\mapsto x_{1,i_1} \ldots x_{p,i_p} \end{aligned} ,$$

where each x_k in X_k is written as $\sum_{i=1}^{n_k} x_{k,i}a_{k,i}$, form a basis of $\mathcal{L}(X_1, \ldots, X_p; K)$.

Proof. By the previous theorem, any element of $\mathcal{L}(X_1, \ldots, X_p; K)$ can be written as a linear combination of the $v_{i_1 \ldots i_p}$:

$$f : X_1 \times \cdots \times X_p \to K$$
$$(x_1, \ldots, x_p) \mapsto \sum_{i_1, \ldots, i_p} c_{i_1 \ldots i_p} v_{i_1 \ldots i_p}(x_1, \ldots, x_p) \, ,$$

and so the family generates $\mathcal{L}(X_1, \ldots, X_p; K)$.

Suppose that there exists a family of coefficents $(\lambda_{i_1 \ldots i_p})_{1 \le i_1 \le n_1, \ldots, 1 \le i_p \le n_1}$ such that

$$\sum_{i_1 \ldots i_p} \lambda_{i_1 \ldots i_p} v_{i_1 \ldots i_p} = 0_{\mathcal{L}(X_1, \ldots, X_p; K)}.$$

Then for any p-tuple of indices $(j_1, \ldots, j_p)$, with $1 \le j_1 \le n_1, \ldots, 1 \le j_p \le n_p$, we have that

$$\sum_{i_1 \ldots i_p} \lambda_{i_1 \ldots i_p} v_{i_1 \ldots i_p}(a_{1,j_1}, \ldots, a_{p,j_p}) = 0_K.$$

But $v_{i_1 \ldots i_p}(a_{1,j_1}, \ldots, a_{p,j_p}) = \delta_{i_1, j_1} \ldots \delta_{i_p, j_p}$, that gives

$$\sum_{i_1 \ldots i_p} \lambda_{i_1 \ldots i_p} v_{i_1 \ldots i_p}(a_{1,j_1}, \ldots, a_{p,j_p}) = \lambda_{j_1 \ldots j_p} = 0_K \; .$$

As such, all the coefficients $\lambda_{j_1, \ldots, j_p}$ must be null, and therefore the family $(v_{i_1 \ldots i_p})_{1 \le i_1 \le n_1, \ldots, 1 \le i_p \le n_1}$ is linearly independent. □

Remark 1. Let us denote $(a^{k,i_k})_{1 \le i_k \le n_k}$ the dual basis of X_k^* associated with the basis $(a_{k,i_k})_{1 \le i_k \le n_k}$ of X_k. Then, for any p-tuple $(x_1, \ldots, x_p)$ of $X_1 \times \cdots \times X_p$, we have

$$v_{i_1 \ldots i_p}(x_1, \ldots, x_p) = \prod_{k=1}^{p} a^{k,i_k}(x_k) \; .$$

1.3 Matrices

1.3.1 Matrices as arrays of scalars

A matrix can be seen simply as a double-indexed family of elements of the commutative unitary ring K, for example $A = (\alpha_{ij})_{1 \le i \le n; 1 \le j \le m}$, which when represented visually would take the form

$$A = \begin{pmatrix} \alpha_{11} & \ldots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \ldots & \alpha_{nm} \end{pmatrix}.$$

We note $\mathcal{M}_{n,m}(K)$ the set of matrices with n lines and m columns with coefficients in K, and more simply $\mathcal{M}_n(K)$ the set $\mathcal{M}_{n,n}(K)$.

It is clear that $\mathcal{M}_{n,m}(K)$ can be identified to K^{nm}, so that it has a natural structure of K-module. Moreover, a multiplication is defined between elements of $\mathcal{M}_{n,m}(K)$ and $\mathcal{M}_{m,p}(K)$:

for $A = (\alpha_{ij})_{1 \le i \le n; 1 \le j \le m} \in \mathcal{M}_{n,m}(K)$ and $B = (\beta_{jk})_{1 \le j \le m; 1 \le k \le p} \in \mathcal{M}_{m,p}(K)$, the matrix AB is the element $C = (\gamma_{ik})_{1 \le i \le n; 1 \le k \le p}$ of $\mathcal{M}_{n,p}(K)$ whose coefficients are given by

$$\gamma_{ik} = \sum_{j=1}^{m} \alpha_{ij} \beta_{jk} \; .$$

1.3.2 Transpose of a matrix

Definition 20. If $A = (a_{ij})_{1\leq i\leq m;1\leq j\leq n}$ is an $m \times n$ matrix, written

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix},$$

then we define the **transpose** of A as the matrix $A^t = (a_{ji})_{1\leq j\leq n:1\leq i\leq m}$, written

$$A^t = \begin{pmatrix} a_{11} & \dots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{mn} \end{pmatrix}.$$

The following property can be easily verified:

Proposition 7. *If A is a matrix of $\mathcal{M}_{n,m}(K)$ and B a matrix of $\mathcal{M}_{m,p}(K)$, then $(AB)^t = B^t A^t$.*

Definition 21. A square matrix $A = (a_{ij})_{1\leq i,j\leq n}$ is said **symmetric** if $A = A^t$, that is if $a_{ij} = a_{ji}$ for each (i, j) such that $1 \leq i, j \leq n$ and $i \neq j$.

1.3.3 Matrices and linear maps

Linear maps between two finitely generated free K-modules can be represented by matrices.

Consider a linear map $f : M \to N$, where M is a finitely generated free K-module with basis $\mathcal{B}_1 = (a_j)_{1\leq j\leq m}$, and N is another finitely generated free K-module with basis $\mathcal{B}_2 = (b_i)_{1\leq i\leq n}$. As f is linear, the image by f of any element $x \in M$ is uniquely determined by the images by f of the elements of the basis $(a_j)_{1\leq j\leq m}$:

$$f(a_j) = \sum_{i=1}^{n} \alpha_{ij} b_i.$$

Indeed, let $x = \sum_{j=1}^{m} x_j a_j$ be a vector of M. Then we can write

$$f(x) = \sum_{j=1}^{m} x_j f(a_j) = \sum_{j=1}^{m} x_j \sum_{i=1}^{n} \alpha_{ij} b_i \ .$$

We can visually represent the vector x as the $m \times 1$ dimensional matrix of its coordinates with regards to the basis $\mathcal{B}_1$ of M:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix},$$

and let us denote by A the matrix

$$A = \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \dots & \alpha_{nm} \end{pmatrix}.$$

Then $f(x)$ is represented with regards to the basis $\mathcal{B}_2$ as

$$Y = AX = \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \dots & \alpha_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^m x_j \alpha_{1j} \\ \vdots \\ \sum_{j=1}^m x_j \alpha_{nj} \end{pmatrix} .$$

We say that A is the matrix of f with regards to the bases $\mathcal{B}_1$ and $\mathcal{B}_2$, that we write $A = Mat_{\mathcal{B}_1,\mathcal{B}_2}(f)$.

In the particular case where f is a linear form, that is $N = K$, we will always choose $\mathcal{B}_2 = (1)$, and we will note $A = Mat_{\mathcal{B}_1}(f)$

In the previous notation, the operations on matrices are defined in such a way that we get easily the following equalities for $f_1 \in \mathcal{L}(M;N)$, $f_2 \in \mathcal{L}(M;N)$ and $\lambda \in K$:

$$Mat_{\mathcal{B}_1,\mathcal{B}_2}(f_1 + f_2) = Mat_{\mathcal{B}_1,\mathcal{B}_2}(f_1) + Mat_{\mathcal{B}_1,\mathcal{B}_2}(f_2) \ ,$$

$$Mat_{\mathcal{B}_1,\mathcal{B}_2}(\lambda f_1) = \lambda Mat_{\mathcal{B}_1,\mathcal{B}_2}(f_1) \ .$$

Finally, if P is a third finitely generated free K-module with basis $\mathcal{B}_3$, and if $f \in \mathcal{L}(M;N)$ and $g \in \mathcal{L}(N;P)$, then

$$Mat_{\mathcal{B}_1,\mathcal{B}_3}(g \circ f) = Mat_{\mathcal{B}_2,\mathcal{B}_3}(g) Mat_{\mathcal{B}_1,\mathcal{B}_2}(f) \ ,$$

and this is the main justification for the definition of the product of matrices.

1.4 Covariance and contravariance

We suppose that M is a finitely generated free K-module.

1. Contravariant coordinates

   It can be proved that all the bases of M involve the same number n of vectors, called rank of M (dimension if K is a field). Let $\mathcal{B} = (a_1, \dots, a_n)$ and $\mathcal{B}' = (b_1, \dots, b_n)$ be two bases of M.

   For the sake of convenience, we represent these bases as row matrices whose components are vectors. Then there exists an invertible matrix P of $\mathcal{M}_n(K)$ such that

   $$(b_1 \dots b_n) = (a_1 \dots a_n) \ P \ .$$

   Let x be a vector of M, decomposed in these bases in $x = \sum_{i=1}^n x_i a_i = \sum_{i=1}^n x'_i b_i$. In terms of matrices, this gives

   $$(a_1 \dots a_n) \begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{pmatrix} = (b_1 \dots b_n) \begin{pmatrix} x'_1 \\ . \\ . \\ . \\ x'_n \end{pmatrix} = (a_1 \dots a_n) \ P \begin{pmatrix} x'_1 \\ . \\ . \\ . \\ x'_n \end{pmatrix} \ ,$$

   which implies that

   $$\begin{pmatrix} x'_1 \\ . \\ . \\ . \\ x'_n \end{pmatrix} = P^{-1} \begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{pmatrix} \ .$$

   Hence, the matrix P gives $\mathcal{B}'$ from $\mathcal{B}$, but P^{-1} gives the coordinates x'_i from the coordinates x_i. For this reason, the coordinates of a vector x in a basis of M are said *contravariant*.

2. Covariant coordinates

   Let us now consider the dual of M, that is the space $M^* = \mathcal{L}(M; K)$ of linear forms from M to the ring of scalars K.

   To the basis $\mathcal{B} = (a_1, \ldots, a_n)$ of M are associated n elements a^j of its dual, defined by $a^j(a_i) = \delta_{i,j}$. That means that $a^j(\sum_{i=1}^n x_i a_i) = x_j$ and the linear form a^j gives the j-th coordinate of the vectors in basis $\mathcal{B}$. It is easily seen that $(a^1, \ldots, a^n)$ is a basis of M^*, called dual basis of $\mathcal{B}$, and therefore M^* is also of finite type. Remark in particular that M and M^* have the same rank n.

   Moreover, the K-module $M^{**} = (M^*)^*$ is called the bi-dual of M, and the map ϕ defined on M by

   $$x \in M \mapsto \phi_x \in M^{**} ,$$

   where $\phi_x(u) = u(x)$ for any linear form $u \in M^*$, is a linear map from M to M^{**}. The image under ϕ of the basis $\mathcal{B}$ is the dual basis of the dual basis of $\mathcal{B}$, and it turns out that ϕ realizes an isomorphism of K-modules between M and M^{**}.

   Remark 2. In the more general case - that is without the hypothesis that M is a finitely generated free K-module - the map ϕ realizes a natural morphism from M to M^{**}, which can be neither injective nor sujective.

   Let u be an element of M^*. The matrix of u in the basis $\mathcal{B}$ is the row matrix $Mat_{\mathcal{B}}(u) = \begin{pmatrix}\alpha_1 & \ldots & \alpha_n\end{pmatrix}$, where $u(a_i) = \alpha_i$. The map $\psi_{\mathcal{B}}$ from M^* to M defined by

   $$u \in M^* \mapsto \psi_{\mathcal{B}}(u) = \alpha_1 a_1 + \cdots + \alpha_n a_n \in M$$

   is clearly a non canonical isomorphism between M^* and M (depending on the choice of $\mathcal{B}$).

   In the notation of the previous paragraph, it is well known that

   $$Mat_{\mathcal{B}'}(u) = Mat_{\mathcal{B}}(u)\ P .$$

   Accordingly, setting $Mat_{\mathcal{B}'}(u) = \begin{pmatrix}\beta_1 & \ldots & \beta_n\end{pmatrix}$, the vectors $\psi_{\mathcal{B}}(u) = \sum_{i=1}^n \alpha_i a_i$ and $\psi_{\mathcal{B}'}(u) = \sum_{i=1}^n \beta_i b_i$ of M are related by the matrix equality

   $$\begin{pmatrix}\beta_1 & \ldots & \beta_n\end{pmatrix} = \begin{pmatrix}\alpha_1 & \ldots & \alpha_n\end{pmatrix}\ P ,$$

   that is

   $$\begin{pmatrix}\beta_1\\ \cdot\\ \cdot\\ \cdot\\ \beta_n\end{pmatrix} = P^t \begin{pmatrix}\alpha_1\\ \cdot\\ \cdot\\ \cdot\\ \alpha_n\end{pmatrix} .$$

   Hence, the matrix P gives $\mathcal{B}'$ from $\mathcal{B}$, but P^t gives the coordinates β_i from the coordinates α_i. For this reason, the coordinates of the vectors associated to linear forms on M are said *covariant*.

1.5 Singular value decomposition of matrices

We will first introduce the notion of orthogonal matrices, which will be used in the singular value decomposition. We restrain our studies to the case of matrices over the field of real numbers. For the complex case, the transpose of a matrix should be replaced with its conjugate transpose, and orthogonal matrices with unitary matrices.

1.5.1 Orthogonal matrices

For this subsection, we will place ourselves in $\mathcal{M}_n(\mathbb{R})$, which we will also sometimes note $\mathbb{R}^{n\times n}$.

Definition 22. An **real inner product** on a finite dimensional vector space over the field $\mathbb{R}$ is a mapping

$$\begin{aligned} V \times V &\to \mathbb{R} \\ (x, y) &\mapsto \langle x, y\ \rangle \end{aligned}$$

such that, for all $x, x', y \in V$ and $\lambda \in \mathbb{R}$,

- $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ only if $x = 0$,
- $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$,
- $\langle x, y\rangle = \langle y, x\rangle$,
- $\langle x + x', y\rangle = \langle x, y\rangle + \langle x', y\rangle$.

Definition 23. A **real inner product space** is a real vector space equiped with an inner product.

As a consequence of the so called Cauchy-Schwarz inequality, it can be proved that a **norm associated with an inner product** is given by $\|x\| = \sqrt{\langle x, x\rangle}$.

The essential notion of orthogonality is closely linked to the scalar product.

Definition 24. Let V be a real inner product space.

- Two vectors x and y of V are said to be **orthogonal** if $\langle x, x\rangle = 0$.
- A set or a family of vectors of V is said **orthonormal** if all vectors are of norm one, and any two distinct vectors are orthogonal.

The most common inner product on $\mathbb{R}^n$ is given for $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^n$ by

$$\langle X, Y\rangle = X^t\, Y,$$

identifying elements of $\mathbb{R}^n$ with column matrices. We will only be considering this inner product from now on.

Definition 25. Let A be a square matrix of $\mathcal{M}_n(\mathbb{R})$.

- The matrix A is said **column orthogonal** if its set of columns consitutes an orthogonal set of vectors of $\mathbb{R}^n$.
- The matrix A is said **row orthogonal** if its set of rows consitutes an orthogonal set of vectors of $\mathbb{R}^n$.

The following proposition enables to introduce orthogonal matrices.

Proposition 8.
- *A matrix A of $\mathcal{M}_n(\mathbb{R})$ is row orthogonal if and only if it is column orthogonal.*
- *Furthermore, the matrix is orthogonal (that is row orthogonal, or in an equivalent manner column orthogonal) if and only if its transpose is orthogonal.*

- *The inverse of an orthogonal matrix is equal to its transpose.*

Proof. Suppose that A is an $n \times n$ column orthogonal real matrix, with $A^{(1)}, \dots, A^{(n)}$ being the columns of A. We must then have that A^t is a row orthogonal matrix. When we mutliply the two, we find that the ij index of A^tA is simply $\langle A^{(i)}, A^{(j)}\rangle$. As each $A^{(i)}$ is of norm one, and they are all orthogonal, $\langle A^{(i)}, A^{(j)}\rangle = \delta_{i,j}$ and so $A^tA = I_n$. This means that $A^{-1} = A^t$, and as such, since $A\ A^t = I_n$, the matrix A must also be row orthogonal. □

1.5.2 Singular value decomposition

Singular value decomposition can be seen as a generalization of eigenvalue decomposition to all matrices, not just non-defective square matrices.

We will first recall some definitions of eigenvectors and eigenvalues.

Definition 26. Let A be a matrix of $\mathcal{M}_n(\mathbb{R})$. An **eigenvector** of A is a nonzero vector v of $\mathbb{R}^n$ such that $Av = \lambda v$, with λ being a scalar which is referred to as the **eigenvalue** of A associated with v.

In order to prove the main result (Theorem 3), we shall admit the following spectral result on symmetric real matrices.

Proposition 9 (Spectral Theorem)**.** *If A is a symmetric matrix of $\mathcal{M}_n(\mathbb{R})$, then there exists an orthogonal matrix P and a diagonal matrix D such that*

$$A = PDP^{-1}, \tag{1.1}$$

the diagonal elements of D being the eigenvalues of A.

We note that because P is orthogonal, the equality 1.1 is equivalent to $A = PDP^t$.

This spectral theorem says that there exists an orthonormal basis of $\mathbb{R}^n$ made up of eigenvectors of A. Moreover, if the eigenvalues of A are ≥ 0, the symmetric matrix A is commonly referred to as a **semi-definite positive symmetric matrix**.

Definition 27. Let A be a real $m \times n$ matrix. A **singular value decomposition** of A is any factorization

$$A = U\ \Sigma\ V^t\ ,$$

where

- U is an orthogonal $m \times m$ matrix,
- Σ is an $m \times n$ matrix with non zero entries only on the diagonal (ie $\Sigma_{ij} = 0$ for all (i, j) not in $\{(1, 1), (2, 2), \dots, (m, m)\}$),
- V is an $n \times n$ orthogonal matrix.

Theorem 3. *Every real $m \times n$ matrix admits a singular value decomposition.*

Proof. Let A be a real $m \times n$ matrix.

The matrix A^tA is a symmetric $n \times n$ matrix, and as such there exists an orthogonal matrix V and a diagonal matrix Λ such that $A^tA = V\Lambda V^t$. Let $V^{(1)}, \dots, V^{(n)}$ be the columns of V, which are eigenvectors of A^tA, and $\lambda_1, \dots, \lambda_n$ be the diagonal elements of Λ, which are the eigenvalues of A^tA.

First of all, let us remark that these eigenvalues are ≥ 0. Indeed, for each column vector X of $\mathbb{R}^n$, eigenvector of A^tA related to an eigenvalue λ, we have $X^tA^tAX = (AX)^tAX = \langle AX, AX\rangle = X^t\lambda X = \lambda\langle X, X\rangle \geq 0$, that gives $\lambda \geq 0$: the matrix A^tA is semi-definite positive.

Since the columns of V are eigenvectors of A^tA, we have that $A^tAV^{(i)} = \lambda_i V^{(i)}$, for each $i \in \{1, \ldots, n\}$. Furthermore, since these columns are orthogonal, we have

$$(V^{(i)})^tA^tAV^{(j)} = (V^{(i)})^t\lambda_j V^{(j)} = \lambda_j\delta_{i,j}.$$

Define $\sigma_i = \sqrt{\lambda_i}$ and suppose that r of the σ_i, let us say $\sigma_1, \ldots, \sigma_r$, are nonzero. For $i = 1, \ldots, r$ let

$$U^{(i)} = AV^{(i)}/\sigma_i.$$

The $U^{(i)}$ form an orthonormal family of r vectors of $\mathbb{R}^m$. We prolong this family into an orthonormal basis of $\mathbb{R}^m$, and put those vectors as columns of a matrix U. We then have that the ij coordinate of the matrix U^tAV is

$$(U^{(i)})^tAV^{(j)} = (V^{(j)})^tA^tU^{(i)} = (V^{(j)})^tA^tAV^{(i)}/\sigma_i = \sigma_i\delta_{i,j}.$$

As such, setting $\Sigma = U^tAV$, we have $A = U\Sigma V^t$. □

1.5.3 The Moore-Penrose generalized inverse

Not every real matrix is invertible; however it is possible to define for every real matrix a generalized inverse or pseudoinverse that has some useful properties. Its existence is a nice application of the singular value decomposition.

Definition 28. Let $A \in \mathbb{R}^{m\times n}$ be a matrix. A **generalized inverse**, or **pseudoinverse** of A is any $n \times m$ matrix, noted A^+, that satisfies the following conditions, sometimes called the Moore-Penrose conditions:

- $AA^+A = A$,
- $A^+AA^+ = A^+$,
- $(AA^+)^t = AA^+$,
- $(A^+A)^t = A^+A$.

Remark 3. It can be proved that the four Moore-Penrose conditions are also equivalent to the following ones, which are usefull in calculations:

- $A^+AA^t = A^t$,
- $A^+(A^+)^tA = A^+$,
- $A^t(A^+)^tA^+ = A^+$,
- $A^tAA^+ = A^t$.

The following result enables to show the existence of a generalized inverse for any real matrix ; it can be easily verified by a simple calculation.

Proposition 10. *Let $A \in \mathbb{R}^{m\times n}$ be a real matrix with the singular value decomposition*

$$A = U\Sigma V^t.$$

Then the matrix

$$V\Sigma^+U^t\ ,$$

where Σ^+ is obtained from Σ by inversing each non-zero element of its diagonal, satisfies the Moore-Penrose conditions, and is therefore a generalized inverse of A.

We then prove the uniqueness.

Proposition 11. *Let $A \in \mathbb{R}^{m\times n}$ be a real matrix. Then A admits a unique pseudoinverse.*

Proof. If two matrices X and Y satisfy these conditions, then

$$X = XAX = X(AX)^t = XX^tA^t = XX^tA^tAY = XAY,$$

and in an similar way,

$$Y = YAY = XAY \ ,$$

thus $X = Y$. □

Chapter 2

An introduction to tensors

Throughout the chapter, K will refer to a commutative unitary ring.

We will first use an approach detailed in [3], where tensors are defined in the context of multi-linear forms - this approach is often used by physicists. We will then use a more abstract approach detailed in [10] and [9], and finally link the two approaches in the final section.

2.1 Tensor product of multi-linear forms

2.1.1 General definition

Definition 29. Let r be a positive integer, $X_{11}, \dots, X_{1p_1}, \dots, X_{r1}, \dots, X_{rp_r}$ K-modules and

$$\begin{aligned} &u_1 : X_{11} \times \dots \times X_{1p_1} \to K, \\ &\vdots \\ &u_r : X_{r1} \times \dots \times X_{rp_r} \to K \end{aligned}$$

multi-linear forms on said modules. We define the **tensor product** of u_1, ..., u_r as the map

$$u_1 \otimes \dots \otimes u_r \ : \ (X_{11} \times \dots \times X_{1p_1}) \times \dots \times (X_{r1} \times \dots \times X_{rp_r}) \ \longrightarrow \ K$$

such that for all $(x_{11}, \dots, x_{1p_1}, \dots, x_{r1}, \dots, x_{rp_r}) \ \in \ (X_{11} \times \dots \times X_{1p_1}) \times \dots \times (X_{r1} \times \dots \times X_{rp_r})$,

$$u_1 \otimes \dots \otimes u_r \ (x_{11}, \dots, x_{1p_1}, \dots, x_{r1}, \dots, x_{rp_r}) = u_1(x_{11}, \dots, x_{1p_1}) \dots u_r(x_{r1}, \dots, x_{rp_r}).$$

The multi-linearity of $u_1, \dots, u_r$ allows us to immediately deduce the following property.

Proposition 12. *If $u_1, \dots, u_r$ are multi-linear forms, then $u_1 \otimes \dots \otimes u_r$ is a multi-linear form.*

Proposition 13. *The tensor product mapping*

$$\begin{aligned} \mathcal{L}(X_{11}, \dots, X_{1p_1}; K) \times \dots \times \mathcal{L}(X_{r1}, \dots, X_{rp_r}; K) &\to \mathcal{L}(X_{11}, \dots, X_{1p_1}, \dots, X_{r1}, \dots, X_{rp_r}; K) \\ (u_1, \dots, u_r) &\mapsto u_1 \otimes \dots \otimes u_r \end{aligned}$$

is multi-linear.

Proof. We will prove the result for $r = 2$ in order to simply notation; moreover, we restrict ourselves to the case of the left-handside, but the right-handside case is very similar. Let $u, u' \in \mathcal{L}(X_1, \ldots, X_p; K)$ and $v \in \mathcal{L}(Y_1, \ldots, Y_q; K)$ be multi-linear forms, and λ, μ be scalars in K. We have that

$$\begin{aligned}
&[(\lambda u + \mu u') \otimes v](x_1, \ldots, x_p, y_1, \ldots, y_q) \\
&\quad = (\lambda u + \mu u')(x_1, \ldots, x_p) v(y_1, \ldots, y_q) \\
&\quad = \lambda u(x_1, \ldots, x_p) v(y_1, \ldots, y_q) + \mu u'(x_1, \ldots, x_p) v(y_1, \ldots, y_q) \\
&\quad = [\lambda u \otimes v](x_1, \ldots, x_p, y_1, \ldots, y_q) + [\mu u' \otimes v](x_1, \ldots, x_p, y_1, \ldots, y_q) \ ,
\end{aligned}$$

for all $x_1, \ldots, x_p, y_1, \ldots, y_q$ in $X_1, \ldots, X_p, Y_1, \ldots, Y_q$, hence the result. □

2.1.2 Case of finitely generated free modules

We suppose that $X_1, \ldots, X_p$ are finitely generated free modules with respective bases $(a_{1,i})_{1 \le i \le n_1}, \ldots, (a_{r,i})_{1 \le i \le n_r}$.

We can reformulate Theorem 2 using this latter definition, in the following manner:
If $f : X_1 \times \cdots \times X_p \to K$ is a multi-linear form, then there exist constants $c_{i_1 \ldots i_p}$ in the scalar ring K such that

$$f = \sum_{i_1 \ldots i_p} c_{i_1 \ldots i_p} a^{1,i_1} \otimes \cdots \otimes a^{1,i_p} \ ,$$

where $a^{1,i_1}, \ldots, a^{p,i_p}$ are respectively elements of the dual bases $(a^{1,j_1})_{1 \le j_1 \le n_1}, \ldots, (a^{p,j_p})_{1 \le j_p \le n_p}$.
In other words, the functions

$$v_{i_1, \ldots, i_p} : X_1 \times \cdots \times X_p \to K \quad (x_1, \ldots, x_p) \mapsto x_{1,i_1} \ldots x_{p,i_p}$$

can be written

$$v_{i_1 \ldots i_p} = a^{1,i_1} \otimes \cdots \otimes a^{p,i_p}.$$

We can also reformulate Theorem 1: a map $f : X_1 \times \cdots \times X_p \to K$ is multi-linear if and only if there exist constants $c_{i_1, \ldots i_p}$ in K such that, for any p-tuple $(x_1, \ldots, x_p)$,

$$f(x_1, \ldots, x_p) = \sum_{i_1, \ldots, i_p} c_{i_1 \ldots i_p} a^{1,i_1} \otimes \cdots \otimes a^{p,i_p}(x_1, \ldots, x_p).$$

2.2 Tensors

We will now introduce tensors as a particular type of multi-linear form, in the way commonly used by physicists. The following definition will refer to the terminology of covariance and contravariance introduced in section 1.4.

2.2.1 General definition

Definition 30. A **tensor** of **class** $\binom{p}{q}$ or tensor of **covariant index** p and **contravariant index** q on the module M is a $(p+q)$-linear form on $(M^*)^p \times M^q$, in other words an element of the module $\mathcal{L}(M^*, \ldots, M^*, M, \ldots, M; K)$ (with p occurrences of M^* and q occurrences of M). We note $T^p_q(M)$ the module of tensors of class $\binom{p}{q}$ on M.

Remark 4. 1. If $u \in T^p_q(M)$ and $v \in T^r_s(M)$, then $u \otimes v$ can be seen as a tensor of class $\binom{p+r}{q+s}$. Indeed, $u \otimes v$ is a multi-linear form on $(M^*)^p \times M^q \times (M^*)^r \times M^s$, which is isomorphic to $(M^*)^{p+r} \times M^{q+s}$.

2. The tensor product is generally **not** commutative: if f and g are two multi-linear forms, then $f \otimes g(x,y) = f(x)g(y)$, while $g \otimes f(x,y) = g(x)f(y)$, when these two expressions make sense (it is possible, depending on the domains of f and g, for $f(x)g(y)$ to be defined while $f(y)g(x)$ is not defined).

Remark 5. Let us recall that there is a natural morphism from M to its bidual (see Remark 2). Given $(x_1, \dots, x_p, u_1, \dots, u_q)$ in $M^p \times (M^*)^q$, this allows to consider tensors of $T^p_q(M)$ written in the form $x_1 \otimes \dots \otimes x_p \otimes u_1 \otimes \dots \otimes u_q$, that is the multi-linear form on $(M^*)^p \times M^q$ whose value at $(v_1, \dots, v_p, y_1, \dots, y_q)$ is the scalar $v_1(x_1) \dots v_p(x_p) u_1(y_1) \dots u_q(y_q)$.

2.2.2 Tensors over a finitely generated free module

Let T be a tensor of type $\binom{p}{q}$ over a finitely generated free module M. Suppose that $(a_1, \dots, a_n)$ is a basis of M and $(a^1, \dots, a^n)$ the associated dual basis of M^*.

Following Theorem 1, for any set of linear forms $u_1, \dots, u_p$ and vectors $x_1, \dots, x_q$, we have that

$$T(u_1, \dots, u_p, x_1, \dots, x_q) = \sum_{i_1, \dots, i_p \,;\, j_1, \dots, j_q} u_{1,i_1} \dots u_{p,i_p} x_{1,j_1} \dots x_{q,j_q} T(a^{i_1}, \dots, a^{i_p}, a_{j_1}, \dots, a_{j_q}),$$

where each u_k is decomposed as $\sum_{i=1}^n u_{k,i} a^i$, and each x_k as $\sum_{j=1}^n x_{k,j} a_j$.

We can therefore see that any such tensor T is uniquely determined by the values of $T(a^{i_1}, \dots, a^{i_p}, a_{j_1}, \dots, a_{j_q})$, that we note

$$T^{i_1, \dots, i_p}_{j_1, \dots, j_q}$$

and refer to them as **structure constants**, or **components**, or **coefficients**, or **coordinates**, of T with regards to the chosen basis, as was done in a more general case in section 1.2.2.

Recall that because M is finitely generated, it is isomorphic to its bi-dual M^{**}. Indeed, as was seen in section 1.4, the dual basis of the basis $(a_i)_{1 \le i \le n}$ is formed by the linear forms $(a^j)_{1 \le j \le n}$ such that $a^j(a_i) = \delta_{i,j}$. As such, the family of linear forms on M^*

$$\begin{aligned} f_{a_i} : M^* &\to K \\ u &\mapsto u(a_i) \end{aligned}$$

forms the dual basis of the basis $(a^j)_{1 \le j \le p}$ of M^*, since $f_{a_i}(a^j) = \delta_{i,j}$. We can identify each f_{a_i} in M^{**} with the corresponding a_i in M, and as such M^{**} is isomorphic to M.

The following theorem is an immediate consequence of Theorem 2 and the previous identification.

Theorem 4. *Let M be a module of finite type, $(a_1, \dots, a_n)$ a basis of M and $(a^1, \dots, a^n)$ the associated dual basis of M^* ; then the elements of the form*

$$a_{i_1} \otimes \dots \otimes a_{i_p} \otimes a^{j_1} \otimes \dots \otimes a^{j_q}$$

constitute a basis of $T^p_q(M)$.

2.3 Tensor product of K-modules

The two previous sections introduced tensor products of multi-linear forms and tensors and mutli-linear forms, using the fact that there is a multiplication defined on the scalar ring K. We will now consider the more general notion of the tensor product of modules, which can be used to combine elements of any K-modules, and in particular multi-linear maps between K-modules that are not necessarily multi-linear forms, ie that do not necessarily map to the scalar ring K. Because it can be used to combine elements of arbitrary vector spaces, the tensor product of modules when applied to individual elements is sometimes called the outer product.

We shall also study the links between this definition and the previous notions when finitely generated free modules are involved.

2.3.1 Construction of the tensor product

Let M and N be two K-modules. One way of approaching the tensor product of two modules is to view it as a way to code a bilinear map on $M \times N$ as a linear map on a new module $M \otimes N$. So our goal will be, given a pair of K-modules M and N, to construct a new K-module $M \otimes N$ and a function π from $M \times N$ to $M \otimes N$ such that for every K-module R and every bilinear map $\phi : M \times N \longrightarrow R$, there exists a unique $\bar{\phi}$ from $M \otimes N$ to R such that $\phi = \bar{\phi} \circ \pi$.

We will prove the existence of such an object by constructing it, and then deduce its uniqueness up to isomorphism.

To construct $M \otimes N$, we start by taking the free module over $M \times N$, that is

$$F_{M\times N} = \{ \sum_{(m,n)\in M\times N} \lambda_{m,n}(m,n) \ ; \ \lambda \in A^{M\times N} \text{ and } \#\{(m,n) \in M \times N \ ; \ \lambda_{(m,n)} \neq 0\} < +\infty\} \ .$$

We consider the sub-module G of $F_{M\times N}$ that consists of all the finite linear combinations of elements of $F_{M\times N}$ that can be written

$$\begin{gathered} (m+m',n) - (m,n) - (m',n) \\ a(m,n) - (am,n) \\ a(m,n) - (m,an) \\ (m,n+n') - (m,n) - (m,n'), \end{gathered}$$

for some $a \in K$, m, $m' \in M$ and n, $n' \in N$. Next, we will simply define $M \otimes N = F_{M\times N}/G$ and check that it works.

Here, we note $\overline{(m,n)}$ for the class of (m,n) in $F_{M\times N}/G$. We have

$$\begin{aligned} &\overline{(m+m',n)} \\ &= \overline{(m+m',n) - (m,n) - (m',n) + (m,n) + (m',n)} \\ &= \overline{0} + \overline{(m,n) + (m',n)} = \overline{(m,n)} + \overline{(m',n)}, \end{aligned}$$

and by a similar process, we find

$$\overline{(m,n+n')} = \overline{(m,n)} + \overline{(m,n')} \ .$$

For scalar multiplication, we have

$$\begin{aligned} \overline{(am,n)} &= \overline{(am,n) + 0} \\ &= \overline{(am,n) - a(m,n) + a(m,n)} \\ &= \overline{0 + a(m,n)} = \overline{a(m,n)} = a\overline{(m,n)}, \end{aligned}$$

and similarly

$$\overline{(m,an)} = \overline{a(m,n)} = a\overline{(m,n)} \ .$$

Let $i : M \times N \to F_{M\times N}$ be the canonical injection and let $\pi_0 : F_{M\times N} \to F_{M\times N}/G$ be the canonical quotient projection. The previous paragraph shows that the mapping

$$\pi = \pi_0 \circ i : M \times N \longrightarrow F_{M\times N}/G$$

is bilinear.

Now if we take a bilinear map $\phi : M \times N \longrightarrow R$, where R is any K-module, we can completely define a corresponding map $\tilde{\phi}$ extending ϕ to $F_{M\times N}$

$$\begin{aligned}\tilde{\phi} : F_{M\times N} &\to R\\ \sum_{(m,n)\in F_{M\times N}} \lambda_{m,n}(m,n) &\mapsto \sum_{(m,n)\in F_{M\times N}} \lambda_{m,n}\phi(m,n),\end{aligned}$$

where both sums are over a finite number of non-zero terms, that is $\lambda_{m,n} = 0$ for all but a finite number of (m,n).

This map $\tilde{\phi}$ is linear, as for all scalars μ_1 and μ_2 and for all elements $\sum_{(m,n)\in M\times N} \lambda_{m,n}(m,n)$ and $\sum_{(m,n)\in M\times N} \lambda'_{m,n}(m,n)$ of $F_{M\times N}$,

$$\tilde{\phi}(\mu_1 \sum_{(m,n)\in M\times N} \lambda_{m,n}(m,n) + \mu_2 \sum_{(m,n)\in M\times N} \lambda'_{m,n}(m,n)) = \tilde{\phi}(\sum_{(m,n)\in M\times N} (\mu_1\lambda_{m,n} + \mu_2\lambda'_{m,n})(m,n))$$

$$= \sum_{(m,n)\in M\times N} (\mu_1\lambda_{m,n} + \mu_2\lambda'_{m,n})\phi(m,n) = \sum_{(m,n)\in M\times N} \mu_1\lambda_{m,n}\phi(m,n) + \sum_{(m,n)\in M\times N} \mu_2\lambda'_{m,n}\phi(m,n)$$

$$= \mu_1 \sum_{(m,n)\in M\times N} \lambda_{m,n}\phi(m,n) + \mu_2 \sum_{(m,n)\in M\times N} \lambda'_{m,n}\phi(m,n) = \mu_1\tilde{\phi}(\sum_{(m,n)\in M\times N} \lambda_{m,n}(m,n)) + \mu_2\tilde{\phi}(\sum_{(m,n)\in M\times N} \lambda'_{m,n}(m,n)),$$

where as before only a finite number of $\lambda_{m,n}$ and $\lambda'_{m,n}$ are non nul.

By bilinearity of ϕ, we have $ker\tilde{\phi} \supseteq G$. Thus $\tilde{\phi}$ is factorisable via the quotient: there exists a linear map

$$\begin{aligned}\bar{\phi} : F_{M\times N}/G &\to R\\ \overline{\sum_{i\in I} \lambda_i(m_i,n_i)} &\mapsto \tilde{\phi}(\sum_{i\in I} \lambda_i(m_i,n_i))\end{aligned}$$

such that for all (m,n) in $M \times N$

$$\phi(m,n) = \tilde{\phi}(m,n) = \overline{\phi}(\overline{(m,n)}) = \overline{\phi} \circ \pi(m,n)\ ,$$

which is exactly the property we wanted.

We have successfully shown the existence of a module $M \otimes N$ such that every bilinear form on $M \times N$ can be uniquely factored through the tensor module via a linear map; now we will show that it is unique up to isomorphism.

Suppose that there exists another such module, which we will call P, that has the same property. Let p be the bilinear map from $M \times N$ onto P such that any bilinear map ϕ on $M \times N$ can be factored via a unique linear map $\hat{\phi}$ such that $\phi = \hat{\phi} \circ p$.
Then there is a unique linear map $\hat{\pi} : P \longrightarrow M \otimes N$ such that $\pi = \hat{\pi} \circ p$, and there is also a unique linear map $\bar{p} : M \otimes N \longrightarrow P$ such that $p = \bar{p} \circ \pi$. Therefore we have

$$\begin{aligned}\pi &= \hat{\pi} \circ p = (\hat{\pi} \circ \bar{p}) \circ \pi\ ,\\ p &= \bar{p} \circ \pi = (\bar{p} \circ \hat{\pi}) \circ p\ .\end{aligned}$$

But we also know, by definition of $M \otimes N$ and of P, that there is a unique function from P to P that factors p, because every bilinear function from $M\times N$ to another module can be uniquely factored through P.

Similarly, we have a unique function from $M \otimes N$ into itself that factors π. Now these must both obviously be identities, so we have that $\hat{\pi}$ and $\bar{p}$ are one-to-one and inverses of one another. In conclusion, $\hat{\pi}$ and $\bar{p}$ realize an isomorphism between $M \otimes N$ and P.

This concludes the uniqueness of $M \otimes N$ up to an isomorphism.

We have just shown the following result:

Theorem 5. *Let M and N be K-modules. Up to an isomorphism, there exists a unique K-module $M \otimes N$, and a unique bilinear map $\pi : M \times N \to M \otimes N$, such that for any K-module R, and any bilinear map $\phi : M \times N \to R$, there is a unique linear map $\overline{\phi} : M \otimes N \to R$ such that $\overline{\phi} \circ \pi = \phi$.*

Remark 6. This is known as the universal property of the tensor product of K-modules.

Remark 7. The converse of the result, that if ϕ is a function on $M \times N$ factored though a linear map on the tensor product module (that is of the form $\psi \circ \pi$ with ψ linear) then it is bilinear, is also true. Indeed, by bilinearity of π and linearity of ψ, we have

$$\begin{aligned}\phi(\lambda m + \mu m', n) &= \psi((\lambda m + \mu m') \otimes n) = \psi(\lambda m \otimes n + \mu m' \otimes n) \\ &= \lambda \psi(m \otimes n) + \mu \psi(m' \otimes n) = \lambda \phi(m, n) + \mu \phi(m', n)\ ,\end{aligned}$$

and similarly on the right hand side.

Definition 31. Given M and N two K-modules, their **tensor product** $M \otimes N$ is defined as the K-module quotient of the free module on $M \times N$ by the previously defined submodule G, that is $M \otimes N = F_{M \times N}/G$.

In the future, we will write $m \otimes n$ for the class of (m, n) in $M \otimes N$. As mentioned before, in some sources, the tensor product of two elements of different modules is sometimes also called the outer product. We give a definition using this terminology here.

Definition 32. If M and N are K-modules we define the **outer product** of $m \in M$ and $n \in N$ to be $m \otimes n \in M \otimes N$.

Figure 2.1 illustrates the construction of this space with a commutative diagram, where $\pi = \pi_0 \circ i$.

$$\begin{tikzcd}
M \times N \arrow[d, "i"'] \arrow[drr, "\phi"] & & \\
F_{M \times N} \arrow[d, "\pi_0"'] \arrow[rr, "\tilde{\phi}"] & & R \\
M \otimes N = F_{M \times N}/G \arrow[urr, "\bar{\phi}"'] & &
\end{tikzcd}$$

Figure 2.1: Commutatif diagram

The following proposition generalizes the theorem and remark to the case of p K-modules.

Proposition 14. *Let $M_1, \dots, M_p$ be K-modules. We define by induction*

$$M_1 \otimes \cdots \otimes M_p = M_1 \otimes (M_2 \otimes \cdots \otimes M_p).$$

Let f be a function from $M_1 \times \cdots \times M_p$ into some K-module R. Then f is p-linear if and only if there exists a linear map $\bar{f} : M_1 \otimes \cdots \otimes M_p \to R$ such that $f(x_1, \dots, x_p) = \bar{f}(x_1 \otimes \cdots \otimes x_p)$ for all $(x_i)_i$ in $(M_i)_i$.

Proof. We proceed by induction. The initial case, for $p = 2$ is already proven. Suppose the result true for $p - 1 \geq 2$, and let f be a function from $M_1 \times \cdots \times M_p$ into R. Then if for each $x_p \in M_p$ we call f_{x_p} the function from $M_1 \times \cdots \times M_{p-1}$ to R which associates to each $(x_1, \dots, x_{p-1})$ the element $f(x_1, \dots, x_{p-1}, x_p)$ of R, then we can factor this function through $M_1 \otimes \cdots \otimes M_{p-1}$ into $\overline{f_{x_p}}$. And then we can simply define $\overline{f}(x_1 \otimes \cdots \otimes x_{p-1} \otimes x_p) = \bar{f_{x_p}}(x_1 \otimes \cdots \otimes x_{p-1})$. □

Proposition 15. *The tensor product of K-modules is associative up to isomorphism: we have*

$$L \otimes (M \otimes N) \simeq (L \otimes M) \otimes N \ .$$

Proof. For any n in N we define the mapping

$$\begin{aligned} \phi_n : L \times M &\to L \otimes (M \otimes N) \\ (l, m) &\mapsto l \otimes (m \otimes n). \end{aligned}$$

This mapping is bilinear, so we can factor it via

$$\begin{aligned} \overline{\phi}_n : L \otimes M &\to L \otimes (M \otimes N) \\ l \otimes m &\mapsto l \otimes (m \otimes n). \end{aligned}$$

Since the mapping $n \mapsto \overline{\phi}_n$ is linear, the mapping

$$\begin{aligned} \Phi : (L \otimes M) \otimes N &\to L \otimes (M \otimes N) \\ (l \otimes m) \otimes n &\mapsto \overline{\phi}_n(l \otimes m) \otimes n = l \otimes (m \otimes n) \end{aligned}$$

is linear.

Similarly, we can fix an l in L and construct

$$\begin{aligned} \psi_l : M \times N &\to (L \otimes M) \otimes N \\ (m, n) &\mapsto (l \otimes m) \otimes n \end{aligned}$$

which we then factor via

$$\begin{aligned} \overline{\psi}_l : M \otimes N &\to (L \otimes M) \otimes N \\ m \otimes n &\mapsto (l \otimes m) \otimes n. \end{aligned}$$

Once again, the mapping $l \mapsto \overline{\psi}_l$ is linear so this makes the mapping

$$\begin{aligned} \Psi : L \otimes (M \otimes N) &\to (L \otimes M) \otimes N \\ l \otimes (m \otimes n) &\mapsto \overline{\psi}_l(m \otimes n) = (l \otimes m) \otimes n \end{aligned}$$

linear.

We have that Φ and Ψ are inverses of each other, so they are bijections, hence the result. □

The following proposition gives a commutative property of the tensor product of K-modules.

Proposition 16. *There is a unique isomorphism from $L \otimes M$ to $M \otimes L$ that sends $l \otimes m$ to $m \otimes l$ for each l in L and each m in M.*

Proof. Let $\phi : L \times M \to M \otimes L$ be the bilinear function that to each (l, m) associates $m \otimes l$ in $M \otimes L$. By the universal property proved in Theorem 5, there is a unique $\bar{\phi} : L \otimes M \to M \otimes L$ such that $\bar{\phi}(l \otimes m) = \phi(l, m) = m \otimes l$. Similarly, take $\psi : M \times L \to L \otimes M$ to be the bilinear function that to each (m, l) associates $l \otimes m$. Once again, by the universal property proved in Theorem 5, there is a unique linear map $\bar{\psi} : M \otimes L \to L \otimes M$ such that $\bar{\psi}(m \otimes l) = \psi(m, l) = l \otimes m$.

We have $\bar{\psi} \circ \bar{\phi} = id_{L \otimes M}$ and $\bar{\phi} \circ \bar{\psi} = id_{M \otimes L}$, hence $L \otimes M$ and $M \otimes L$ are isomorphic. ☐

Remark 8. It is possible for the tensor product of two non-trivial K-modules to be trivial. For example, if $M \otimes N = \mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/3\mathbb{Z}$ is isopmorphic to $\{0\}$ as for any m and n we have $m.1_M \otimes n.1_N = mn(3-2)(1_M \otimes 1_N) = mn[(3.1_M \otimes 1_N) - (1_M \otimes 2.1_N)] = mn[(0.1_M \otimes 1_N) - (1_M \otimes 0.1_N)] = mn[0(1_M \otimes 1_N) - 0(1_M \otimes 1_N) = 0_{M \otimes N}$.

2.3.2 Properties

We will now list general properties of the tensor product of K-modules.

From the definition of $M \otimes N$ we get immediately the following proposition.

Proposition 17. *The tensor product $M \otimes N$ is generated by the set of the "**pure tensors**" $m \otimes n$ for m in M and n in N.*

We have shown that π is bilinear, so we get the following equalities:

Proposition 18. *For any modules M and N, and any elements $m \in M$ and $n \in N$,*

$$a(m \otimes n) = (am) \otimes n = m \otimes (an) \ ,$$

$$(m + m') \otimes n = m \otimes n + m' \otimes n \ ,$$

and similarly for the right side.

Moreover, Theorem 5 has the following important corollary.

Proposition 19. *For any K-module R, the module $\mathcal{L}(M, N; R)$ of bilinear maps from $M \times N$ to R is isomorphic to the module $\mathcal{L}(M \otimes N; R)$ of linear maps from $M \otimes N$ to R.*

Proof. In the notation of Theorem 5, consider the map

$$L : \mathcal{L}(M, N; R) \longrightarrow \mathcal{L}(M \otimes N; R) \ ,$$

$$\phi \longmapsto \bar{\phi}$$

clearly well defined on all elements of $\mathcal{L}(M, N; R)$. The map L is linear, as for all $\lambda \in K$ and for all $\phi, \phi_1, \phi_2 \in \mathcal{L}(M, N; R)$

$$\begin{aligned} L(\phi_1 + \phi_2) &= \overline{\phi_1 + \phi_2} = \bar{\phi}_1 + \bar{\phi}_2 = L(\phi_1) + L(\phi_2) \\ L(\lambda\phi) &= \lambda\bar{\phi} = \lambda L(\phi) \end{aligned}$$

by the universal property of $M \otimes N$. Indeed, $\phi_1 + \phi_2$ can be factored via $\overline{\phi_1 + \phi_2}$ since it is a bilinear map on $M \times N$. As ϕ_1 can be factored via $\overline{\phi_1}$ and ϕ_2 can be factored via $\overline{\phi_2}$, $\phi_1 + \phi_2$ can also be factored via $\overline{\phi_1} + \overline{\phi_2}$. But then $\overline{\phi_1 + \phi_2} = \overline{\phi_1} + \overline{\phi_2}$ as the factorisation is unique.

Similarly, we can factor $\overline{\lambda\phi}$ through $\overline{\lambda\phi}$ and through $\lambda\bar{\phi}$, hence the latter two mappings are the same.

Finally, to show that L is a bijection, take any $f \in \mathcal{L}(M \otimes N; R)$, and consider the function

$$\begin{aligned} \phi : M \times N &\to R \\ (m, n) &\mapsto f(m \otimes n). \end{aligned}$$

This ϕ is bilinear, as for all $\lambda \in K$, $m \in M$ and $n \in N$,

$$\phi(m + m', n) = f((m + m') \otimes n) = f(m \otimes n + m' \otimes n) = f(m \otimes n) + f(m' \otimes n) = \phi(m, n) + \phi(m', n)$$
$$\phi(\lambda m, n) = f(\lambda m \otimes n) = \lambda f(m \otimes n) = \lambda \phi(m, n),$$

and similarly on the right hand side. To use the language of Theorem 5 we have $\phi = f \circ \pi$.

This function is also unique, as any other function with the same property would have the same image for all elements of $M \times N$, and as such L is an isomorphism, since every element of $\mathcal{L}(M \otimes N; R)$ has a unique antecedent by L.

The universal property of $M \otimes N$ implies that every bilinear ϕ on $M \times N$ can be associated with a unique linear f on $M \otimes N$. Since we have just shown that every linear function on $M \otimes N$ has a unique antecedent, hence the result.

□

Remark 9. By the previous proposition, the module $\mathcal{L}(M \otimes N; R)$ is isomorphic to $\mathcal{L}(M, N; R)$. Thus, by proposition 6, the modules $\mathcal{L}(M; \mathcal{L}(N; R))$ and $\mathcal{L}(M \otimes N; R)$ are also isomorphic.

Following Proposition 14, this can be generalized to p-linear maps over p modules.

2.4 Links between diverse notions of tensor product and tensors

2.4.1 A mapping from $M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^*$ to $T^p_q(M)$

Proposition 20. *Suppose K is a commutative ring. Let M be a K-module and p and q positive integers. Consider the module $M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^*$, with p occurrences of M and q occurrences of M^*. There exists a unique morphism*

$$j : M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^* \to T^p_q(M)$$

that sends the element $x_1 \otimes \cdots \otimes x_p \otimes u_1 \otimes \cdots \otimes u_q$ of $M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^*$ to the element $x_1 \otimes \cdots \otimes x_p \otimes u_1 \otimes \cdots \otimes u_q$ of $T^p_q(M)$ (following Remark 5, we recall that this latter expression is the multi-linear form on $(M^*)^p \times M^q$ whose value at $(v_1, \ldots, v_p, y_1, \ldots, y_q)$ is the scalar $v_1(x_1) \ldots v_p(x_p) u_1(y_1) \ldots u_q(y_q)$).

Proof. Since the map $M^p \times (M^*)^q$ that sends the element $(x_1, \ldots, x_p, u_1, \ldots, u_q)$ of $M^p \times (M^*)^q$ to the element $x_1 \otimes \cdots \otimes x_p \otimes u_1 \otimes \cdots \otimes u_q$ of $T^p_q(M)$ is multi-linear, the proposition is an immediate corollary of the universal property of the tensor product of modules (see Proposition 14). In other words, the morphism j is completely defined by its values on the pure tensors of $M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^*$.

□

Remark 10. If $M = \mathbb{Z}/p\mathbb{Z}$, and K is $\mathbb{Z}$ (so that M is not a free K-module), then $T^2_0(M) = \{0\}$ is not isomorph to $M \otimes M$. As such, in this case the morphism j introduced above is not an isomorphism. Indeed, consider the mapping from $M \times M$ to M associating to (x, y) the product xy (that is the multiplication of integers modulo p). This is a bilinear non trivial map, and by the universal property of $M \otimes M$, it can be factorized through a linear map on $M \otimes M$, which shows that $M \otimes M$ is not reduced to its zero element. For the case of $T^2(M)$ we have that any element can be uniquely identified by its image of basis element, but $M^* = \{0\}$, and so $T^2(M) = 0$.

Remark 11. We have just shown that while the notions of tensors defined as multi-linear forms and tensors defined as elements of the tensor product of K-modules do coincide when the K-modules are finitely generated free modules, they are not necessarily identical. We have also seen that the tensor product of two non-trivial K-modules can be trivial.

2.4.2 On tensor product of linear maps

The results exhibited here above hold not only for the case of tensor products of linear forms, but also for tensor products of linear map with values in K-modules.

Proposition 21. *Let L, L', M and M' be four K-modules. Let*

$$u : L \to L'$$

and

$$v : M \to M'$$

be two linear maps. Then there exists a unique linear map f from $L \otimes M$ to $L' \otimes M'$ such that $f(x \otimes y) = u(x) \otimes v(y)$, for all x in L and y in M.

Proof. Let $\phi : L \times M \to L' \otimes M'$ be the map that sends (x, y) to $u(x) \otimes v(y)$. We verify that ϕ is indeed bilinear:

$$\begin{aligned}\phi(\lambda x + \lambda' x', y) &= (\lambda x + \lambda' x') \otimes y \\ =&\lambda(x \otimes y) + \lambda'(x' \otimes y) = \lambda\phi(x, y) + \lambda'\phi(x', y)\end{aligned}$$

by bi-linearity of the tensor product, and similarly on the right side. As ϕ is bilinear there is a unique linear map $\bar{\phi} : L \otimes M \to L' \otimes M'$ such that $\bar{\phi}(x \otimes y) = \phi(x, y)$. We chose $f = \bar{\phi}$. □

This unique linear map will now be formally defined.

Definition 33. Let L, L', M, M' be K-modules and $u : L \to L'$ and $v : M \to M'$ be linear maps. The **tensor product of linear maps** u and v is the unique linear map, noted $u\tilde{\otimes}v$ that associates to each $l \otimes m$ in $L \otimes M$ the element $u(l) \otimes v(m)$ in $L' \otimes M'$.

Notation The notation $u\tilde{\otimes}v$ for the tensor product of linear maps u and v is to distinguish $u\tilde{\otimes}v$ from the element $u \otimes v$ of $\mathcal{L}(L, L') \otimes \mathcal{L}(M, M')$. However, let us remark that the mapping

$$\begin{gathered}\mathcal{L}(L; L') \times \mathcal{L}(M; M') \to \mathcal{L}(L \otimes M; L' \otimes M') \\ (u, v) \mapsto u\tilde{\otimes}v\end{gathered}$$

is bilinear, as this is easily deduced from the fact that, for each pure tensors $x \otimes y \in L \otimes M$,

$$\begin{aligned}&[(\lambda u + \mu u')\tilde{\otimes}v](x \otimes y) = (\lambda u + \mu u')(x) \otimes v(y) \\ &= (\lambda u(x) + \mu u'(x)) \otimes v(y) = \lambda u(x) \otimes v(y) + \mu u'(x) \otimes v(y) \\ &= [\lambda u\tilde{\otimes}v](x \otimes y) + [\mu u'\tilde{\otimes}v](x \otimes y),\end{aligned}$$

and similarly on the right hand side. The universal property of Theorem 5 implies that it can be uniquely factored via a linear map

$$\begin{gathered}\mathcal{L}(L; L') \otimes \mathcal{L}(M; M') \to \mathcal{L}(L \otimes M; L' \otimes M') \\ u \otimes v \mapsto u\tilde{\otimes}v\end{gathered}.$$

This morphism is called the **Kronecker morphism**. In this general case, this morphism is neither injective nor surjective, although it can be shown to be injective in the case of finite dimensional vector spaces.

Proposition 22. *Let L, L', L'', M, M' and M'' be modules. Let*

$$\begin{array}{ll} u' : L \to L' & u'' : L' \to L'' \\ v' : M \to M' & v'' : M' \to M'' \end{array}$$

be linear maps. Then

$$(u'' \circ u')\tilde{\otimes}(v'' \circ v') = (u''\tilde{\otimes}v'') \circ (u'\tilde{\otimes}v').$$

Proof. Let x be an element of L and y be an element of M. We have that

$$\begin{aligned} &[(u'' \circ u')\tilde{\otimes}(v'' \circ v')](x \otimes y) = (u'' \circ u')(x) \otimes (v'' \circ v')(y) \\ &= u''(u'(x)) \otimes v''(v'(y)) = u''\tilde{\otimes}v''(u'(x) \otimes v'(y)) \\ &= (u''\tilde{\otimes}v'') \circ (u'\tilde{\otimes}v')(x \otimes y), \end{aligned}$$

for all x in L and y in M, hence the result. □

We now have different types of tensor products. First, there is the tensor product of linear or multi-linear forms (tensors as commonly used in physics being a particular case of this), as described in [3]. Second, we have the tensor product of K-modules, and the definition of pure tensors as elements of this tensor product of K-modules. Finally, we have just seen the tensor product of linear maps, which we noted $\tilde{\otimes}$, so as to distinguish $u \otimes v$ from its image by the Kronecker morphism $u\tilde{\otimes}v$.

2.4.3 The special case of tensor product of finitely generated free modules

In this section K is a commutative ring and L and M are K-modules.

A base for the tensor product of modules

We will begin with a technical result which will enable us to determine a base for the tensor product of two finitely generated free modules.

Proposition 23. *for any families of K-modules $(L_i)_i$ and $(M_j)_j$,*

$$(\bigoplus_i L_i) \otimes (\bigoplus_j M_j) \simeq \bigoplus_{i,j} (L_i \otimes M_j).$$

Proof. Consider the function

$$\begin{aligned} \Pi : (\prod_i L_i) \times (\prod_j M_j) &\to \prod_{i,j} (L_i \otimes M_j) \\ ((l_i)_i), (m_j)_j) &\mapsto (l_i \otimes m_j)_{i,j}. \end{aligned}$$

It is bilinear as

$$\begin{aligned} \Pi(\lambda(l_i)_i + \mu(l'_i)_i, (m_j)_j) &= ((\lambda l_i + \mu l'_i) \otimes m_j)_{i,j} \\ &= (\lambda l_i \otimes m_j + \mu l'_i \otimes m_j)_{i,j} = \lambda(l_i \otimes m_j)_{i,j} + \mu(l'_i \otimes m_j)_{i,j} \\ &= \lambda\Pi((l_i)_i, (m_j)_j) + \mu\Pi((l'_i)_i, (m_j)_j)\ , \end{aligned}$$

and similarly on the right hand side. By the universal property of the tensor product of K-modules, there exists therefore a unique linear map, determined by the values on the pure tensors

$$\overline{\Pi} : (\prod_i L_i) \otimes (\prod_j M_j) \to \prod_{i,j}(L_i \otimes M_j)$$
$$(l_i)_i \otimes (m_j)_j \mapsto (l_i \otimes m_j)_{i,j}.$$

When we compose this linear map with the canonical injection

$$i : (\bigoplus_i L_i) \otimes (\bigoplus_j M_j) \to (\prod_i L_i) \otimes (\prod_j M_j)$$

we find that the mapping

$$\Phi : (\bigoplus_i L_i) \otimes (\bigoplus_j M_j) \to \bigoplus_{i,j}(L_i \otimes M_j)$$
$$(l_i)_i \otimes (m_j)_j \mapsto (l_i \otimes m_j)_{i,j}$$

is linear, as it is a composition of two linear maps, and well defined, as the image of any family with a finite number of non nul elements will be a family with a finite number of non nul elements.

For each pair of indices i, j we define

$$\phi_{i,j} : L_i \times M_j \to (\bigoplus_i L_i) \otimes (\bigoplus_j M_j)$$
$$(l_i, m_j) \mapsto (l_i) \otimes (m_j).$$

These functions are bilinear, as can be simply verified, so there exists for each i, j a unique linear map

$$\overline{\phi_{i,j}} : L_i \otimes M_j \to (\bigoplus_i L_i) \otimes (\bigoplus_j M_j)$$
$$(l_i \otimes m_j) \mapsto (l_i) \otimes (m_j).$$

These linear maps can be combined to create a linear map

$$\Psi : \bigoplus_{i,j}(L_i \otimes M_j) \to (\bigoplus_i L_i) \otimes (\bigoplus_j M_j)$$
$$(l_i \otimes m_j)_{i,j} \mapsto (l_i)_i \otimes (m_j)_j$$

which is the inverse of Φ, hence the isomorphism. □

Proposition 24. *Let L and M be finitely generated free modules. Let $(a_i)_{1\leq i\leq p}$ be a basis of L and $(b_j)_{1\leq j\leq q}$ be a basis of M. The products*

$$(a_i \otimes b_j)_{1\leq i\leq p, 1\leq j\leq q}$$

form a basis of $L \otimes M$.
The rank of $L \otimes M$ is the product of the rank of L and the rank of M. In particular, if K is a commutative field, then

$$dim(L \otimes M) = dim(L)dim(M).$$

Proof. By Proposition 23 we have

$$L \otimes M = (\bigoplus_i Ka_i)_i \otimes (\bigoplus_j Kb_j)_j \simeq \bigoplus_{i,j} K(a_i \otimes b_j).$$

□

Isomorphism between $M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^*$ and $T^p_q(M)$

We come back to the mapping j from $M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^*$ to $T^p_q(M)$ of Proposition 20.

Proposition 25. *Suppose M be a free K-module of finite type, p and q positive integers and consider the module $M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^*$, with p occurrences of M and q occurrences of M^*.*
The map

$$j : M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^* \to T^p_q(M)$$

that sends the element $x_1 \otimes \cdots \otimes x_p \otimes u_1 \otimes \cdots \otimes u_q$ of $M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^*$ to the element $x_1 \otimes \cdots \otimes x_p \otimes u_1 \otimes \cdots \otimes u_q$ of $T^p_q(M)$ is an isomorphism.

Proof. Suppose that M is finitely generated, with basis $(a_i)_{1\leq i\leq n}$. Then we have that the elements

$$(a_{i_1} \otimes \cdots \otimes a_{i_p} \otimes a^{j_1} \otimes \cdots \otimes a^{j_q})_{1\leq i_1,\ldots,i_p,j_1\ldots,j_p\leq n}$$

form a basis of $M \otimes \cdots \otimes M \otimes M^* \otimes \cdots \otimes M^*$, and their images by j form a basis of $T^p_q(M)$: j is an isomorphism.

□

Tensor product of linear maps and Kronecker product of matrices

Let L, L', M and M' be free K-modules of finite type. Let

$$u : L \to L' \text{ and } v : M \to M'$$

be linear maps. We are going to show how to write the matrix of $u\tilde{\otimes}v$ as a function of the matrices of u and v.

Let us consider $\mathcal{A} = (a_i)_{1\leq i\leq p}$, $\mathcal{B} = (b_j)_{1\leq j\leq q}$, $\mathcal{C} = (c_k)_{1\leq k\leq r}$ and $\mathcal{D} = (d_l)_{1\leq l\leq s}$ bases of L, L', M and M' respectively. Let $A = (\alpha_{ij})_{(i,j)\in\{1,\ldots,p\}\times\{1,\ldots,q\}}$ and $B = (\beta_{kl})_{(k,l)\in\{1,\ldots,r\}\times\{1,\ldots,s\}}$ be the matrices of u and v in these bases:

$$A = \mathcal{M}_{\mathcal{A},\mathcal{B}}(u) = \begin{bmatrix} \alpha_{11} & \ldots & \alpha_{1p} \\ \vdots & \ddots & \vdots \\ \alpha_{q1} & \ldots & \alpha_{qp} \end{bmatrix},$$

$$B = \mathcal{M}_{\mathcal{C},\mathcal{D}}(v) = \begin{bmatrix} \beta_{11} & \ldots & \beta_{1r} \\ \vdots & \ddots & \vdots \\ \beta_{s1} & \ldots & \beta_{sr} \end{bmatrix}.$$

For the sake of convenience, we identify in the following elements of the module L, L', M and M' with their column matrices of coordinates in the chosen bases. If x and y are elements of L and M, then $(u\tilde{\otimes}v)(x,y)$ will be the element $x' \otimes y'$ with

$$x' = \begin{pmatrix} x'_1 = \sum_i \alpha_{1i}x_i \\ \vdots \\ x'_q = \sum_i \alpha_{qi}x_i \end{pmatrix}$$

and

$$y' = \begin{pmatrix} y'_1 = \sum_i \beta_{1i} y_i \\ \vdots \\ y'_s = \sum_i \beta_{si} y_i \end{pmatrix}.$$

In order to write the tensor products $x \otimes y$ and $x' \otimes y$ as column matrices, we will need to chose a way of ordering the elements of the bases $(a_i \otimes c_k)_{1 \leq i \leq p, 1 \leq k \leq r}$ or $L \otimes M$ and $(b_j \otimes d_l)_{1 \leq j \leq q, 1 \leq l \leq s}$ of $L' \otimes M'$. We will chose to increase the indices of the c_k before the a_i and the d_l before the b_j.

Then, we can write the tensor product of vectors in the form

$$x \otimes y = \begin{pmatrix} x_1 y_1 \\ \vdots \\ x_1 y_r \\ x_2 y_1 \\ \vdots \\ x_2 y_r \\ \vdots \\ x_p y_1 \\ \vdots \\ x_p y_r \end{pmatrix} \text{ and } x' \otimes y' = \begin{pmatrix} x'_1 y'_1 \\ \vdots \\ x'_1 y'_r \\ x'_2 y'_1 \\ \vdots \\ x'_2 y'_s \\ \vdots \\ x'_q y'_1 \\ \vdots \\ x'_q y'_s \end{pmatrix}.$$

We note that this convention of having the y_k change before the x_i and the y'_l change before the x'_j is completely arbitrary. It is possible to use a different symbolic formalism, for instance alternating the x_i before the y_i, and arrive at the result so long the same formalism is used during the entirety of the calculation.

When the bases are ordered according to the previous convention, the matrix of $u \tilde{\otimes} v$ is equal to

$$\begin{bmatrix}
\alpha_{11}\beta_{11} & \dots & \alpha_{11}\beta_{1r} & \alpha_{12}\beta_{11} & \dots & \alpha_{12}\beta_{1r} & \dots & \alpha_{1p}\beta_{11} & \dots & \alpha_{1p}\beta_{1r} \\
\alpha_{11}\beta_{21} & \dots & \alpha_{11}\beta_{2r} & \alpha_{12}\beta_{21} & \dots & \alpha_{21}\beta_{r2} & \dots & \alpha_{p1}\beta_{21} & \dots & \alpha_{p1}\beta_{2r} \\
\vdots & & & & & & & & \vdots & \\
\alpha_{11}\beta_{s1} & \dots & \alpha_{11}\beta_{sr} & \alpha_{12}\beta_{s1} & \dots & \alpha_{12}\beta_{sr} & \dots & \alpha_{1p}\beta_{s1} & \dots & \alpha_{1p}\beta_{sr} \\
\alpha_{21}\beta_{11} & \dots & \alpha_{21}\beta_{1r} & \alpha_{22}\beta_{11} & \dots & \alpha_{22}\beta_{1r} & \dots & \alpha_{2p}\beta_{11} & \dots & \alpha_{2p}\beta_{1r} \\
\vdots & & & & & & & & \vdots & \\
\alpha_{21}\beta_{s1} & \dots & \alpha_{21}\beta_{sr} & \alpha_{22}\beta_{s1} & \dots & \alpha_{22}\beta_{sr} & \dots & \alpha_{2p}\beta_{s1} & \dots & \alpha_{2p}\beta_{sr} \\
\vdots & & & & & & & & \vdots & \\
\vdots & & & & & & & & \vdots & \\
\alpha_{q1}\beta_{11} & \dots & \alpha_{q1}\beta_{1r} & \alpha_{q2}\beta_{11} & \dots & \alpha_{q2}\beta_{r1} & \dots & \alpha_{qp}\beta_{11} & \dots & \alpha_{qp}\beta_{1r} \\
\vdots & & & & & & & & \vdots & \\
\alpha_{q1}\beta_{s1} & \dots & \alpha_{q1}\beta_{sr} & \alpha_{q2}\beta_{s1} & \dots & \alpha_{q2}\beta_{sr} & \dots & \alpha_{qp}\beta_{s1} & \dots & \alpha_{qp}\beta_{sr}
\end{bmatrix}.$$

We can see that this matrix is obtained from the matrix A of u replacing each coordinate α_{ij} by the matrix block $\alpha_{ij} B$.

This is sometimes called the **Kronecker Product** of two matrices and noted $A \otimes B$.

.

Remark 12. Were we to adopt the convention of having the x_i change before the the y_k and the x'_j change before the y'_l, the matrix of $A \otimes B$ would be the matrix composed of the blocks $b_{ij} A$, rather than the blocks $aijB$ as is with the current convention.

Chapter 3

Hypermatrices and tensors

In this section, F will denote a commutative field.

Hypermatrices are generalizations of matrices to families of scalars with more than two indexes. We will also see that they can be seen as representing the coefficients in a given basis, or set of bases, of a tensor.

Notation We will sometimes use the notation $\langle n \rangle$ to denote the set $\{1, \ldots, n\}$.

3.1 Basic definitions

Definition 34. A function $f : \langle n_1 \rangle \times \cdots \times \langle n_d \rangle \to F$ will be referred to as a **hypermatrix** of order d.

Notation The set of hypermatrices on $n_1, \ldots n_d$ with coefficients in the field F will be denoted $F^{n_1 \times \cdots \times n_d}$.

A hypermatrix can also be represented in the form $A = [a_{i_1 \ldots i_d}]$ with $i_k \in \langle n_k \rangle$ for $k \in \langle d \rangle$. A hypermatrix of order 2 is a standard matrix, and as such the set of $m \times n$ matrices over F can be referred to as either $\mathcal{M}_{m \times n}(F)$ or $F^{m \times n}$.

Much as with matrices, we define termwise addition in the following manner: if $A = [a_{i_1 \ldots i_d}]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle}$ and $B = [b_{i_1 \ldots i_d}]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \langle n_d \rangle}$ are two hypermatrices, of $F^{n_1 \times \cdots \times n_d}$, then

$$A + B = [a_{i_1 \ldots i_d} + b_{i_1 \ldots i_d}]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle}$$

will be their sum. For scalar multiplication, if λ is a scalar of F, and $A = [a_{i_1 \ldots i_d}]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle}$ is a hypermatrix, then

$$\lambda A = [\lambda a_{i_1 \ldots i_d}]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle}$$

will be the their product.

We define the standard basis of $F^{n_1 \times \cdots \times n_d}$ to be $\mathcal{E} = \{E_{i_1 \ldots i_d} : i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle\}$ where $E_{i_1 \ldots i_d}$ denotes the hypermatrix with a $(i_1, \ldots, i_d)$ coordinate of 1 and zeros everywhere else.

The standard matrix multiplication can be generalized to hypermatrices.

Definition 35. Let $X_1 = (x^1_{ij}) \in F^{m_1 \times n_1}, \ldots, X_d = (x^d_{ij}) \in F^{m_d \times n_d}$ are matrices and $A = [a_{i_1 \ldots i_d}] \in F^{n_1 \times \cdots \times n_d}$ is a hyper matrix, their mutli-linear matrix product is

$$A' = (X_1, \ldots, X_d) \cdot A = [a'_{i_1 \ldots i_d}] \in F^{m_1 \times \cdots \times m_d}$$

defined by

$$a'_{i_1 \ldots i_d} = \sum_{k_1, \ldots, k_d} x^1_{i_1 k_1} \ldots x^d_{i_d k_d} a_{k_1 \ldots k_d}.$$

We have the following properties:

Proposition 26. *If $A \in F^{n_1 \times \cdots \times n_d}$ is a hypermatrix, and for $i \in \langle d \rangle$ let $X_i \in F^{l_i \times m_i}$ and $Y_i \in F^{m_i \times n_i}$ are matrices, then*

$$(X_1, \ldots, X_d) \cdot ((Y_1, \ldots, Y_d) \cdot A) = (X_1 Y_1, \ldots, X_d Y_d) \cdot A.$$

Proof. By calculating we have

$$\begin{aligned}
&(X_1, \ldots, X_d) \cdot ((Y_1, \ldots, Y_d) \cdot A) \\
&= (X_1, \ldots, X_d) \cdot \left[\sum_{k_1, \ldots, k_d}^{n_1, \ldots, n_d} y^1_{j_1 k_1} \ldots y^d_{j_d k_d} \cdot a_{k_1 \ldots k_d} \right]_{j_1 \in \langle m_1 \rangle, \ldots, j_d \in \langle m_d \rangle} \\
&= \left[\sum_{j_1, \ldots, j_d}^{n_1, \ldots n_d} x^1_{i_1 j_1} \ldots x^d_{i_d j_d} \sum_{k_1, \ldots, k_d}^{n_1, \ldots, n_d} y^1_{j_1 k_1} \ldots y^d_{j_d k_d} \cdot a_{k_1 \ldots k_d} \right]_{i_1 \in \langle l_1 \rangle, \ldots, i_d \in \langle l_d \rangle} \\
&= \left[\sum_{j_1, \ldots, j_d} \sum_{k_1, \ldots, k_d} x^1_{i_1 j_1} y^1_{j_1 k_1} \ldots x^d_{i_d j_d} y^d_{j_d k_d} a_{k_1 \ldots k_d} \right]_{i_1 \in \langle l_1 \rangle, \ldots, i_d \in \langle l_d \rangle} \\
&= \left[\sum_{k_1, \ldots, k_d} (\sum_{j_1} x^1_{i_1 j_1} y^1_{j_1 k_1}) \ldots (\sum_{j_d} x^d_{i_d j_d} y^d_{j_d k_d}) a_{k_1 \ldots k_d} \right]_{i_1 \in \langle l_1 \rangle, \ldots, i_d \in \langle l_d \rangle} \\
&= (X_1 Y_1, \ldots, X_d Y_d) \cdot A.
\end{aligned}$$

□

Proposition 27. *If A and B are hypermatrices in $F^{n_1 \times \cdots \times n_d}$, α and β are scalars and for $k \in \langle d \rangle$ $X_k \in F^{m_k \times n_k}$ are matrices, then*

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

Proof. Once again by calculating we find

$$\begin{aligned}
&(X_1, \ldots, X_d) \cdot [\alpha a_{i_1 \ldots i_d} + \beta b_{i_1 \ldots i_d}]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle} = \\
&\left[\sum_{k_1, \ldots, k_d} x^1_{i_1 k_1} \ldots x^d_{i_d k_d} (\alpha a_{i_1 \ldots i_d} + \beta b_{i_1 \ldots i_d} \right]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle} \\
&= \alpha \left[\sum_{k_1, \ldots, k_d} x^1_{i_1 k_1} \ldots x^d_{i_d k_d} a_{k_1 \ldots k_d} \right]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle} + \beta \left[\sum_{k_1, \ldots, k_d} x^1_{i_1 k_1} \ldots x^d_{i_d k_d} b_{k_1 \ldots k_d} \right]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle} \\
&= \alpha (X_1, \ldots, X_d) \cdot A + \beta (X_1, \ldots, X_d) \cdot B.
\end{aligned}$$

□

Definition 36. If $\pi \in \mathfrak{S}_d$ is a permutation, and $A = [a_{i_1 \ldots i_d}]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle nd \rangle}$ is a hypermatrix of order d, then we define the **π-transpose of** A to be

$$A^\pi = [a_{\pi(i_1) \ldots \pi(i_d)}]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle}.$$

The space of $F^{n_1 \times n_p}$ of order d hypermatrices on a field F is a vector space for these operations. Indeed, it is an abelian group with a scalar multiplication that is distributive with regards to the addition operation.

Hypermatrices and tensors Let $V_1, \ldots, V_d$ are finitely generated free vector spaces of dimension $n_1, \ldots, n_d$ over the field $F = \mathbb{R}$ or $F = \mathbb{C}$, with respective bases $\mathcal{B}^1, \ldots, \mathcal{B}^d$, where each $\mathcal{B}^i = \{b^i_1, \ldots, b^i_{n_i}\}$. If the tensor $T \in V_1 \otimes \cdots \otimes V_d$ has the structure constants or coordinates in the basis $(b^1_{i_1} \otimes \cdots \otimes b^d_{i_d})_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \rangle n_d \rangle}$ of $(a_{i_1 \ldots i_d})_{i_1, \ldots, i_d}$ then we can represent it in the basis as the hypermatrix

$$A = [a_{i_1 \ldots i_d}] \in F^{n_1 \times \cdots \times n_d}.$$

Furthermore, if $\mathcal{B}'^1, \ldots, \mathcal{B}'^d$ is a second set of bases for $V_1, \ldots, V_d$, and $X_1, \ldots, X_d$ are the respective change of basis matrices from $\mathcal{B}^1, \ldots, \mathcal{B}^d$ to $\mathcal{B}'^1, \ldots, \mathcal{B}'^d$, then hypermatrix of T in the new set of bases is

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

Definition 37. Let F be a field and let $u^1 \in F^{n_1}, u^2 \in F^{n_2}, \ldots, u^d \in F^{n_d}$ be vectors. Their **Segre outer product**, noted $\otimes b \otimes c$ is defined as

$$[u^1_{i_1} u^2_{i_2} \ldots u^d_{i_d}]^{n_1 n_2 \ldots n_d}_{i_1 = 1 i_2 = 1 \ldots i_d = 1}.$$

The Segre outer product can help us to define an isomorphism between $F^{n_1} \otimes \cdots \otimes F^{n_d}$ and $F^{n_1 \times \cdots \times n_d}$ when F is $\mathbb{R}$ or $\mathbb{C}$ (and possibly some other fields, but it is best to avoid overgeneralizing and thereby including pathological cases). Indeed, consider the Segre map

$$\begin{aligned} \phi : F^{n_1} \times \cdots \times F^{n_d} &\to F^{n_1 \times \cdots \times n_d} \\ (u^1, \ldots, u^d) &\mapsto u^1 \otimes \cdots \otimes u^d. \end{aligned}$$

It is bilinear, and as such by the universal property of the tensor product of F-modules there is a unique $\theta : F^{n_1} \otimes \cdots \otimes F^{n_d}$ such that $\theta(u^1 \otimes \cdots \otimes u^d) = u^1 \otimes \cdots \otimes u^d$. This mapping is evidently injective, and since $F^{n_1 \times \cdots \times n_d}$ and $F^{n_1} \otimes \cdots \otimes F^{n_d}$ have the same dimension, it is an isomorphism. As such, we have just proved the following result

Proposition 28. *Let F be a the field $\mathbb{R}$ or $\mathbb{C}$ and $n_1, \ldots, n_d$ be positive integers. The vector spaces $F^{n_1 \times \cdots \times n_d}$ and $F^{n_1} \otimes \cdots \otimes F^{n_d}$ are isomorphic.*

Definition 38. Let $A = [a_{i_1 \ldots i_p}] \in F^{n_1 \times \cdots \times n_p}$ and $B = [b_{j_1 \ldots j_q}] \in F^{m_1 \times \cdots \times m_q}$ be hypermatrices. Their **outer product of hypermatrices** noted $A \otimes B$ is the hypermatrix

$$A \otimes B = [a_{i_1 \ldots i_p} b_{j_1 \ldots j_q}] \in F^{n_1 \times \cdots \times n_p \times m_1 \times \cdots \times m_q}.$$

3.1.1 Some useful concepts

The terminology *mode*, when used to refer to a hypermatrix, will designate one of its dimensions. For example, if $A \in F^{2 \times 3 \times 4}$ is a hypermatrix, its first mode, or mode 1 will be two, second mode 3, and third mode 4.

Definition 39. Let $A = [a_{i_1 \ldots i_d}]_{i_1 \in \langle n_1 \rangle, \ldots, i_d \in \langle n_d \rangle}$ be a hypermatrix. The **mode q fibers** of A are the vectors formed by fixing all the indices but the q mode index.

The mode q fibers are noted $a_{i_1 \ldots i_{q-1} : i_{q+1} \ldots i_d}$.

To illustrate this, if A is as before an element of $F^{n_1 \times \cdots \times n_d}$, then the mode q fibers of A will be vectors of the form

$$a_{i_1 \ldots i_{q-1} : i_{q+1} \ldots i_d} = \begin{pmatrix} a_{i_1 \ldots i_{q-1} 1 i_{q+1} \ldots i_d} \\ \vdots \\ a_{i_1 \ldots i_{q-1} n_q i_{q+1} \ldots i_d} \end{pmatrix}$$

Fibers are the hypermatrix analogue of rows and columns of a matrix. In fact, the rows of a matrix are its mode 1 fiber, and its columns are its mode 2 fibers.

Definition 40. Given a hypermatrix of order d $\mathcal{X} \in F^{n_1 \times \cdots \times n_d}$, the **mode k flattening** of $\mathcal{X}$ is the matrix noted $X^{(k)}$ whose columns are the mode i fibers of $\mathcal{X}$, arranged in reverse lexicographical order.

Reverse lexicographical order simply means that the leftmost index will vary first, then the second leftmost, and so on, with the rightmost index will vary last. It is the lexicographical order applied to the reverse of each string of indices. For example, given a hypermatrix of order 4 where $n_1 = n_2 = n_3 = n_4 = 2$, if we chose to take its mode 2 representation, we will have a 2×8 matrix with columns

$$X^{(2)} = \begin{bmatrix} x_{1:11} & x_{2:11} & x_{1:21} & x_{2:21} & x_{1:12} & x_{2:12} & x_{1:22} & x_{2:22} \end{bmatrix}.$$

More generally, the mode k flattening is described by mapping each $x_{i_1 \ldots i_k \ldots i_d}$ in the hypermatrix to $x_{i_k j}$ in the mode k flattening, where

$$j = 1 + \sum_{l=1, l \neq k}^{d} (i_l - 1)(\prod_{m=1, m \neq k}^{l-1} n_m).$$

3.1.2 Alternative matrix products

In addition to the Kronecker product, there are other possible matrix products. We will describe two of them, the Khatri-Rao product and the Hadamard product, although other types, such as the Semi-Tensor product and the Tracy-Singh product, also exist. Furthermore, we note that the formalism we have used to describe the Khatri-Rao product is not unique, and that other ways of defining this product exist.

Definition 41. Let $A \in F^{l \times n}$ and $B \in F^{m \times n}$ be two matrices with the same number of columns. The **Khatri-Rao product** of A and B, noted $A \odot B$, is the $lm \times n$ matrix whoses columns are the Kronecker products of the corresponding columns of A and B.

To illustrate this, suppose

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{l1} & \ldots & a_{ln} \end{bmatrix}$$

and

$$B = \begin{bmatrix} b_{11} & \ldots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \ldots & b_{mn} \end{bmatrix}.$$

Then $A \odot B$ will take the form

$$A \odot B = \begin{bmatrix} a_{11}b_{11} & \ldots & a_{1n}b_{1n} \\ \vdots & \ddots & \vdots \\ a_{11}b_{m1} & \ldots & a_{1n}b_{mn} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ a_{l1}b_{11} & \ldots & a_{ln}b_{1n} \\ \vdots & \ddots & \vdots \\ a_{l1}b_{m1} & \ldots & a_{ln}b_{mn} \end{bmatrix}$$

using the previous formalism of varying the indices of the second matrix before those of the first.

Definition 42. Let $A \in F^{m\times n}$ and $B \in F^{m\times n}$ be two matrices with the same dimensions. Their **Hadamard product**, noted $A * B$ is the $m \times n$ matrix whose elements are the products of the corresponding elements of A and B.

To illustrate this, suppose

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

and

$$B = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix}.$$

The Hadamard product will be

$$A * B = \begin{bmatrix} a_{11}b_{11} & \dots & a_{1n}b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & \dots & a_{mn}b_{mn} \end{bmatrix}.$$

3.2 Graphs, hypergraphs, and networks

This section will introduce the basic terminology used to describe graphs, networks and hypergraphs.

Definition 43. A **graph** $G = (V_G, E_G)$ is a set of nodes (or vertices) V_G and a set of edges between those nodes $E_G \subseteq V_G \times V_G$.

Graphs are also sometimes referred to as **networks** and we will use both terminologies. For the sake of simplicity, we will assume that for a finite graph, the nodes are labeled $1, \dots, n$ where n is the number of nodes. We will also consider exclusively finite graphs.

A graph is called **simple** if it has no loops and at most one edge between each pair of nodes. A **multigraph** is a graph with multiple edges between nodes.

The concept of a graph can be generalized to a hypergraph, by weakening the condition that edges must be pairs.

Definition 44. A **hypergraph** $H = (V_H, E_H)$ is a collection of vertices or nodes along with a set of hyperedges $E_H \subseteq \mathcal{P}(V_H)$ linking those nodes.

In both hypergraphs and graphs, edges can be undirected or directed, and weighted or unweighted.

Definition 45. A **directed graph** or **digraph** is one where the edges (a, b) and (b, a) are distinct. In other words, edges in a directed graph have a direction from one node to another.

3.2.1 Multiplex networks

Graphs and networks are frequently used to represent data and their interactions. Beyond the visualization of the interaction network (often impossible due to the size of the network), this representation is useful for data analysis and to infer, from a topological analysis of the graph, underlying properties in the data set. In molecular biology, new technologies provide a huge quantity of heterogeneous data, that can be interpreted as interactions of the biological components at different scales: protein-protein interactions, gene regulation through transcription factors, gene regulation through non-codant RNA, signalling pathways, correlation of expression levels... [8]. While it is important to consider these data as a whole, there is also evidence that it

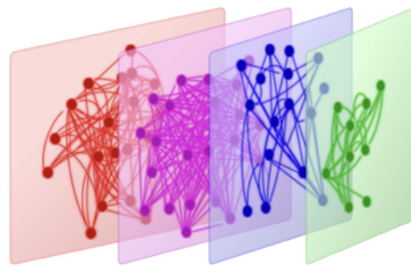

Figure 3.1: A multiplex network

is better to study each of these networks separately, rather than lumping them together into a single network [2]. Multiplex networks are composed of several layers of simple (monoplex) networks. Each layer shares the same set of nodes, but their edges belong to different categories (they have a different meaning). Hence, multiplex allow to encode all the different type of interactions between the components, while keeping them separate. An illustration of this is given Figure 3.1.

We need efficient methods to mine and analyse multiplex networks despite their huge size. For example, in [11] is proposed a random walk with restart method to predict key components around a gene of interest. Tensors can be used to encode multiplexes (cf the following subsections), and then could shed new light on the data and their structure.

3.2.2 Some uses of hypermatrices

Hypermatrices can be used to encode hypergraphs and multiplexes. For hypergraphs, if the hypergraph contains n nodes, it can be coded as an order n hypercubical hypermatrix will all indices in $\langle n+1 \rangle$. To the hypergraph $H = (V, E)$ we can associate the hypermatrix $A_H = [a_{i_1 \dots i_n} \in \mathbb{R}^{n+1 \times \dots \times n+1}$ such that each $a_{i_1 \dots i_n}$ is the weight of the hyperedge between nodes $i_1, \dots, i_n$. Of course, this will only work for hyperedges of size n, which is why we add a "0" node to the node set. A hyperedge linking nodes i, j and k will be encoded by the index $a_{0 \dots ijk \dots 0}$. However, this poses problems of redundancy - the same hyperedge of size r will be in $\binom{n+1}{r}$ different places in the hypermatrix!

Coding multiplexes is a somewhat simpler task. Any multiplex $M = (V, E_1, \dots, E_p)$ can be encoded in an order 3 hypergraph

$$A_M = [a_{ijk}] \in \mathbb{R}^{n \times n \times p}$$

where n is the size of V. Each a_{ijk} is the value of the edge (i, j) in the layer k.

3.3 Tensor rank decomposition

Definition 46. The **rank** of a tensor $T \in F^{n_1} \otimes \dots \otimes F^{n_d}$ is the number of simple or pure tensors needed to write the tensor as a linear combination of simple tensors. In other words, it is the smallest r such that there exists $(a_i^{(1)} \otimes \dots \otimes a_i^{(d)})_{i \in \langle r \rangle} \in F^{n_1} \otimes \dots \otimes F^{n_d}$ and $(\lambda_i)_{i \in \langle r \rangle}$ such that

$$T = \sum_i^r \lambda_i a_i^{(1)} \otimes \dots \otimes a_i^{(k)}.$$

Notation The following terms will be used interchangeably - simple tensor, pure tensor, and rank-1 tensor.

The **tensor rank decomposition** of a hypermatrix, also sometimes called CANDECOMP, PARAFAC, or CP-decomposition, is the process of taking a hypermatrix representation of a given tensor and finding the hypermatrix representations and associated scalars of the simple tensors that compose it.

A proof of the following can be found in [5].

Proposition 29. *Computing the rank of a tensor over any field that contains $\mathbb{Q}$ is NP-hard.*

In practice, most tensor rank decompositions are done as approximations up to a pre-specified rank. The decompositions are not nested - the best rank $r-1$ approximation may not be part of the best rank r approximation. Approximate tensor rank decompositions can be done using the tool Tensorly [7]. We will also give a description of an algorithm, called the alternating least squares algorithm, which can be used to calculate approximate tensor rank decompositions, which can be found, along with an in depth discussion of other methods, in [6].

This algorithm will use the Moore-Penrose generalized matrix inverse, a brief discussion of which can be found in the annex.

We will also quickly define a tensor norm, which is simply a generalization of the standard euclidian norm on $\mathbb{R}^n$.

Definition 47. The **tensor norm** of a tensor $\mathbb{X} = [x_{i_1 \dots i_d} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ is

$$||\mathcal{X} = \sqrt{\sum_{i_1,\dots,i_d} x^2_{i_1\dots i_d}}||.$$

Given an order d hypermatrix $\mathcal{X} \in \mathbb{R}^{n_1\times\dots\times n_d}$ and a pre-specified rank R, our goal will be to find scalars $\lambda_1,\dots,\lambda_R$ and vectors $a_1^{(1)},\dots a_R^{(1)} \in \mathbb{R}^{n_1}, a_1^{(2)},\dots a_R^{(2)} \in \mathbb{R}^{n_2}\dots, a_1^{(d)},\dots a_R^{(d)} \in \mathbb{R}^{n_d}$ so as to minimize$||\mathcal{X} - \sum_R \lambda_r a_r^{(1)} \otimes \dots \otimes a_r^{(d)}||$.

We will organise the vectors $a_1^{(i)},\dots,a_R^{(i)}$ into the columns of a matrix $A^{(i)} = \begin{bmatrix} a_1^{(i)} & a_2^{(i)} & \dots & a_R^{(i)} \end{bmatrix}$. We will next state the following lemma, which can be proved by a simple calculation.

Lemma 5.1. *If $\mathcal{X} \in \mathbb{R}^{n_1\times\dots\times n_d}$ is a hypermatrix with mode i flattening $X^{(i)}$ and with tensor rank decomposition*

$$\mathcal{X} = \sum_{r=1}^{R} \lambda_r a_r^{(1)} \otimes \dots \otimes a_r^{(d)}$$

then

$$X^{(i)} = A^{(i)}\Lambda(A^{(d)} \odot \dots \odot A^{(i+1)} \odot A^{(i-1)} \odot \dots \odot A^{(1)})^T$$

where Λ is the diagonal matrix with $\lambda_1,\dots,\lambda_R$ as its diagonal.

The algorithm starts initializing each of the $A^{(i)}$ matrices. The simplest way to do this is randomly, although other ways are possible, such as for example taking the R left singular vectors of the SVD of $X^{(i)}$. Once all matrices are initialized, we fix all but the first one, and solve for that algebraically, and normalize it by storing the norms of each column as λ_r. In short, we preform the following operations

$$\begin{aligned}
&X^{(i)} = A^{(i)}\Lambda(A^{(d)} \odot \dots \odot A^{(i+1} \odot A^{(i-1)} \odot \dots \odot A^{(1)})^T \\
\therefore &X^{(i)}((A^{(d)} \odot \dots \odot A^{(i+1} \odot A^{(i-1)} \odot \dots \odot A^{(1)})^T)^+ = A^{(i)}\Lambda \\
&\lambda_r = ||\lambda_r a_r^{(i)}||
\end{aligned}$$

where as in the annex, A^+ denotes the Moore-Penrose pseudo-inverse of A.

We then repeat the process for each one of the other matrices. We keep doing this loop until a stopping criteria is met - either the norm of $\mathcal{X} - \sum_r \lambda_r a_r^{(1)} \otimes \dots \otimes a_r^{(d)}$ stops decreasing, or a pre-specified maximal number of iterations is reached.

In algorithmic form, this gives 1.

Data: $\mathcal{X}, R$,maxIter
Result: $\lambda_1, \dots, \lambda_R, A^{(1)}, \dots, A^{(d)}$
initialize $A^{(1)}, \dots, A^{(d)}$;
iter = 0 ;
newFit = $||\mathcal{X} - \sum_r \lambda_r a^{(1)} \otimes \cdots \otimes a^{(d)}||$;
oldFit = newFit + 1 ;
while *oldFit* $\geq$ *newFit and iter* $<$ *maxIter* **do**
    oldFit = newFit ;
    for $i \in \langle d \rangle$ **do**
        $\hat{A}^{(i)} = X^{(i)}((A^{(d)} \odot \cdots \odot A^{(i+1)} \odot A^{(i-1)} \odot \cdots \odot A^{(1)})^T)^+$;
    newFit = $||\mathcal{X} - \sum_r \lambda_r a^{(1)} \otimes \cdots \otimes a^{(d)}||$;
    iter = iter + 1 ;

Complexity A tensor rank approximation of rank R of a hypermatrix $F^{n_1 \times \cdots \times n_d}$ will require NR^d space, where $N = \prod_i n_i$. For time complexity, in [4] there is an algorithm that calculates the SVD of an $m \times n$ matrix in $O(max(m^2 n, n^3))$ time. A matrix multiplication of an $l \times m$ matrix with an $m \times n$ matrix requires $O(lmn)$ operations. A Khatri-Rao product of an $l \times n$ matrix with an $m \times n$ matrix also requires $O(lmn)$ operations. Each iteration of the for loop involves calculation $d-2$ Khatri-Rao products, which requires $O(N/n_i)$ operations in total, followed by a calculation of the Moore-Penrose inverse (which is more or less the same as a calculation the SVD of an $N/n_i \times R$ matrix, which will be be $O((N/n_i)^3)$ if $N/n_i > R$, and $O(R^2 N)$ if not, and finally there is a multiplication of the mode i flattening of the hypermatrix, of dimension $n_i \times N/n_i$, with the $N/n_i \times R$ matrix, which takes $O(NR)$ time. This means that depending on R, n_i and N, one iteration of the for loop takes $O((N/n_i)^3)$ time or $O(NR^2)$ time. We then have at least d iterations of the for loop, and an indefinite number of while loop iterations. As such, we can say that if $N \geq R$, the whole process is cubic in N.

Applications In practice, this can be used to break up a multiplex network into a sum of smaller component networks, which can be more easily studied.

Bibliography

[1] Daizhan Cheng. *Analysis and Control of Boolean Networks: A Semi-Tensor Approach.*

[2] Gilles Didier, Christine Brun, and Anais Baudot. In: *PeerJ* (2015). URL: https://doi.org/10.7717/peerj.1525.

[3] Roger Godement. *Cours d'algèbre (French) [Algebra].*

[4] Gene H. Golub and Charles F. Van Loan. *Matrix Computations 4th edition.*

[5] Christopher J. Hillar and Lek-Heng Lim. *Most Tensor Problems are NP-Hard.*

[6] Tamara G. Kolda and Brett W. Bader. *Tensor Decompositions and Applications.*

[7] Jean Kossaifi et al. *TensorLy: Tensor Learning in Python.*

[8] Nicolas Le Novère. "Quantitative and logic modelling of molecular and gene networks". In: *Nature Publishing Group* February (2015). ISSN: 1471-0056. DOI: 10.1038/nrg3885. URL: http://dx.doi.org/10.1038/nrg3885.

[9] Robert Rolland. *Produit Tensoriel D'Espaces Vectoriels.*

[10] Marc Sage. *Algèbre Multilinéaire(French) [Multi-linear Algebra].*

[11] Alberto Valdeolivas et al. "Random Walk with Restart on Multiplex and Heterogeneous Biological Networks." In: *Bioinformatics (Oxford, England)* (July 2018). Ed. by Alfonso Valencia. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/bty637. URL: https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/bty637/5055408%20http://www.ncbi.nlm.nih.gov/pubmed/30020411.