B

Functional Analysis

In this appendix we provide a very loose review of linear functional analysis. Due to the extension of the topic and the scope of the book, this presentation relies on intuition more than mathematical rigor. In order to support intuition we present the subject as a generalization to the infinite-dimensional world of calculus of the familiar formalism and concepts of linear algebra. For this reason we parallel as closely as possible the discussion in Appendix A. For a more rigorous discussion the reader is referred to references such as Smirnov (1964), Reed and Simon (1980), Rudin (1991), and Whittaker and Watson (1996).

B.1 Vector space

The natural environment of linear functional analysis are infinite-dimensional spaces of functions that are a direct extension of the finite-dimensional Euclidean space discussed in Appendix A.1.

The main difference (and analogy) between the Euclidean space and a vector space of functions is that the discrete integer index n of the Euclidean vectors becomes a continuous index $\mathbf{x}$. Furthermore, we also let the value of the vector be complex. Therefore we define an element of the yet to be defined vector space by extending $(A.2)$ as follows:

$$v: \mathbf{x} \in \mathbb{R}^N \to v(\mathbf{x}) \in \mathbb{C}. \tag{B.1}$$

Notice that we denote as v the function, to be compared with the boldface notation $\mathbf{v}$ in $(A.2)$, which denotes a vector; on the other hand we denote as $v(\mathbf{x})$ the specific value of that function in $\mathbf{x}$, to be compared with the entry of the vector v_n in $(A.2)$.

We represent the analogy between $(B.1)$ and $(A.2)$ graphically in Figure B.1, which parallels Figure A.1.

The set of functions $(B.1)$ is a vector space, since the following operations are properly defined on its elements.

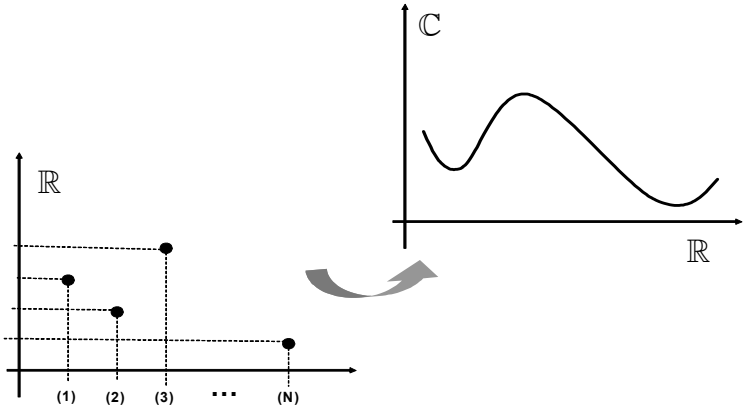

Fig. B.1. From linear algebra to functional analysis

The *sum* of two functions is defined point-wise as follows:

$$[u+v](\mathbf{x}) \equiv u(\mathbf{x}) + v(\mathbf{x}). \tag{B.2}$$

This is the infinite-dimensional version of the "parallelogram rule" of a Euclidean space, compare with $(A.3)$.

The *multiplication by a scalar* is defined point-wise as follows:

$$[\alpha v](\mathbf{x}) \equiv \alpha v(\mathbf{x}), \tag{B.3}$$

compare with $(A.4)$.

Combining sums and multiplications by a scalar we obtain *linear combinations* of functions.

We have seen the striking resemblance of the definitions introduced above with those introduced in Appendix A.1. With the further remark that finite sums become integrals in this infinite-dimensional world, we can obtain almost all the results we need by simply changing the notation in the results for linear algebra. We summarize the main notational analogies between linear algebra and linear functional analysis in the following table:

	linear algebra	**functional analysis**
index/dimension	$n \in \{1,\ldots,N\}$	$\mathbf{x} \in \mathbb{R}^N$
element	$\mathbf{v}:\ n \to v_n \in \mathbb{R}$	$v:\ \mathbf{x} \to v(\mathbf{x}) \in \mathbb{C}$
sum	$\sum_{n=1}^{N}[\cdot]$	$\int_{\mathbb{R}^N}[\cdot]\,d\mathbf{x}$

(B.4)

A set of functions are *linearly independent* if the parallelotope they generate is non-degenerate, i.e. if no function can be expressed as a linear combination of the others.

Using the analogies of Table B.4 we can generalize the definition of inner product given in $(A.5)$ and endow our space of functions with the following *inner product*:

$$\langle u, v\rangle \equiv \int_{\mathbb{R}^N} u(\mathbf{x})\,\overline{v(\mathbf{x})}d\mathbf{x}, \tag{B.5}$$

where $\bar{\cdot}$ denotes the *conjugate transpose*:

$$\overline{a+ib} \equiv a-ib, \qquad a,b\in\mathbb{R}, \quad i\equiv\sqrt{-1}. \tag{B.6}$$

By means of an inner product we can define orthogonality. Similarly to $(A.11)$ a pair of functions (u, v) are *orthogonal* if:

$$\langle u, v\rangle = 0. \tag{B.7}$$

Orthogonal functions are in particular linearly independent, since the parallelotope they generate is not skewed, and thus non-degenerate. For a geometrical interpretation refer to the Euclidean case in Figure A.2.

As in the finite-dimensional setting of linear algebra, the inner product allows us to define the *norm* of a function, i.e. its "length", by means of the *Pythagorean theorem*. Using the analogies of Table B.4 the definition of norm given in $(A.6)$ becomes:

$$\|v\| \equiv \sqrt{\langle v, v\rangle} = \sqrt{\int_{\mathbb{R}^N} |v(\mathbf{x})|^2\, d\mathbf{x}}. \tag{B.8}$$

When defined, this is a norm, since it satisfies the properties $(A.7)$.

Nevertheless, unlike in the finite-dimensional setting of linear algebra, for most functions the integrals $(B.5)$ and $(B.8)$ are not defined. Therefore, at this stage we restrict our space to the set of vectors with finite length:

$$L_2\left(\mathbb{R}^N\right) \equiv \left\{ v \text{ such that } \int_{\mathbb{R}^N} |v(\mathbf{x})|^2\, d\mathbf{x} < \infty \right\}. \tag{B.9}$$

This set is clearly a restriction of the original set of functions $(B.1)$. Furthermore, we extend this set to include in a natural way a set of *generalized functions*, namely elements that behave like functions inside an integral, but are not functions as $(B.1)$ in the common sense of the word. This way the space $(B.9)$ becomes a *complete vector space*.[1]

[1] This extension can be understood intuitively as follows. Consider the set of numbers:

$$\mathcal{S} \in \left\{ \frac{1}{x}, x\in\mathbb{R} \right\}. \tag{B.10}$$

This set is the real axis deprived of the zero. Adding the zero element to this set makes it a *complete* set, which is a much richer object.

A complete vector space where the norm is defined in terms of an inner product as in $(B.8)$ is called a Hilbert space. Therefore the space of functions $L_2\left(\mathbb{R}^N\right)$ is a Hilbert space: this is the closest infinite-dimensional generalization of a finite-dimensional Euclidean vector space. In Table B.11 we summarize how the properties of the Hilbert space $L_2\left(\mathbb{R}^N\right)$ compare to the properties of the Euclidean space $\mathbb{R}^N$.

	linear algebra	**functional analysis**
space	Euclid $\mathbb{R}^N$	Hilbert $L_2\left(\mathbb{R}^N\right)$
inner product	$\langle \mathbf{u}, \mathbf{v} \rangle \equiv \sum_{n=1}^{N} u_n v_n$	$\langle u, v \rangle \equiv \int_{\mathbb{R}^N} u(\mathbf{x}) \overline{v(\mathbf{x})} d\mathbf{x}$
norm (length)	$\|\mathbf{v}\| \equiv \sqrt{\sum_{n=1}^{N} v_n^2}$	$\|v\| \equiv \sqrt{\int_{\mathbb{R}^N} \|v(\mathbf{x})\|^2 d\mathbf{x}}$

(B.11)

We conclude this section mentioning a more general vector space of functions. Indeed, instead of $(B.8)$ we can define a norm as follows:

$$\|v\|_p \equiv \left(\int_{\mathbb{R}^N} |v(\mathbf{x})|^p d\mathbf{x} \right)^{\frac{1}{p}}, \tag{B.12}$$

where $1 \leq p < \infty$. Notice that $(B.8)$ corresponds to the particular case $p \equiv 2$ in $(B.12)$. It can be proved that $(B.12)$ is also a norm, as it satisfies the properties $(A.7)$. This norm is defined on the following space of functions:

$$L_p\left(\mathbb{R}^N\right) \equiv \left\{ v \text{ such that } \int_{\mathbb{R}^N} |v(\mathbf{x})|^p d\mathbf{x} < \infty \right\}. \tag{B.13}$$

Unlike $(B.8)$, in the general case $p \neq 2$ this norm is not induced by an inner product. A complete normed space without inner product is called a *Banach space*. Therefore the spaces $L_p\left(\mathbb{R}^N\right)$ are Banach spaces.

B.2 Basis

A basis is a set of linearly independent elements of a vector space that can generate any vector in that space by linear combinations. According to Table B.4, the discrete integer index n of the Euclidean vectors becomes a continuous index $\mathbf{y} \in \mathbb{R}^N$. Therefore the definition $(A.13)$ of a basis for a Euclidean space is generalized to the Hilbert space $L_2\left(\mathbb{R}^N\right)$ as follows. A basis for $L_2\left(\mathbb{R}^N\right)$ is a set of linearly independent functions indexed by $\mathbf{y}$:

$$e^{(\mathbf{y})}, \quad \mathbf{y} \in \mathbb{R}^N, \tag{B.14}$$

such that any function v of $L_2\left(\mathbb{R}^N\right)$ can be expressed as a linear combination:

$$v = \int_{\mathbb{R}^N} \alpha(\mathbf{y})\, e^{(\mathbf{y})} d\mathbf{y}. \tag{B.15}$$

In analogy with $(A.16)$, the *canonical basis* of $L_2\left(\mathbb{R}^N\right)$ is the set of functions $\delta^{(\mathbf{y})}$ indexed by $\mathbf{y}$ such that the inner product of a generic element of this basis $\delta^{(\mathbf{y})}$ with a generic function v in $L_2\left(\mathbb{R}^N\right)$ yields the "$\mathbf{y}$-th entry", i.e. the value of the function at that point:

$$\left\langle v, \delta^{(\mathbf{y})} \right\rangle \equiv v(\mathbf{y}). \tag{B.16}$$

The generic element $\delta^{(\mathbf{x})}$ of this basis is called the *Dirac delta centered in* $\mathbf{x}$.

We notice that the Dirac delta is not a standard function: it is a generalized function, since it only makes sense within an integral. Indeed, no regular function can possibly satisfy $(B.16)$, since from $(B.5)$ for all functions v this hypothetical regular function should satisfy the following equality:

$$\int_{\mathbb{R}^N} v(\mathbf{x})\, \overline{\delta^{(\mathbf{y})}(\mathbf{x})} d\mathbf{x} = v(\mathbf{y}). \tag{B.17}$$

In order for $(B.17)$ to be true, $\delta^{(\mathbf{y})}(\mathbf{x})$ should be zero for all values of $\mathbf{x}$, except for $\mathbf{x} \equiv \mathbf{y}$. In this case the above integral would be zero, no matter the value of $\delta^{(\mathbf{y})}$ in $\mathbf{x} \equiv \mathbf{y}$.

Therefore, the Dirac delta is not a standard function: instead, it is a limit case of standard functions. Define an approximation of the Dirac delta in terms of the *Gauss exponential function* as follows:

$$\delta_\epsilon^{(\mathbf{y})}(\mathbf{x}) \equiv \frac{1}{(2\pi)^{\frac{N}{2}}\, \epsilon^N} e^{-\frac{1}{2\epsilon^2}(\mathbf{x}-\mathbf{y})'(\mathbf{x}-\mathbf{y})}. \tag{B.18}$$

This is a bell-shaped, smooth function that reaches its peak in $\mathbf{x} \equiv \mathbf{y}$ and whose width is of the order of ϵ. We plot in Figure B.2 this function for different values of ϵ. As the width ϵ approaches zero the bell becomes taller and thinner around the peak $\mathbf{y}$. Intuitively, as $\epsilon \to 0$ the function $\delta_\epsilon^{(\mathbf{y})}$ becomes zero everywhere, except at the point $\mathbf{y}$ where its value becomes infinite. This profile generalizes the finite-dimensional canonical basis $(A.15)$.

Furthermore, the integral of the approximate Dirac delta function $(B.18)$ over the whole real axis is one, since it is a specific case of the multivariate normal probability density function (2.156). Thus the inner product of $\delta_\epsilon^{(\mathbf{y})}$ with another function v is a weighted average of the values of v, where the most weight is given to the points in a neighborhood of $\mathbf{y}$ of radius ϵ. Therefore:

$$\left\langle v, \delta_\epsilon^{(\mathbf{y})} \right\rangle \approx v(\mathbf{y}). \tag{B.19}$$

In the limit $\epsilon \to 0$ this approximation becomes the equality $(B.16)$.

We might be puzzled that the elements of the basis of the space $L_2\left(\mathbb{R}^N\right)$, which is a set of functions, is not a function. In reality, this is not a problem:

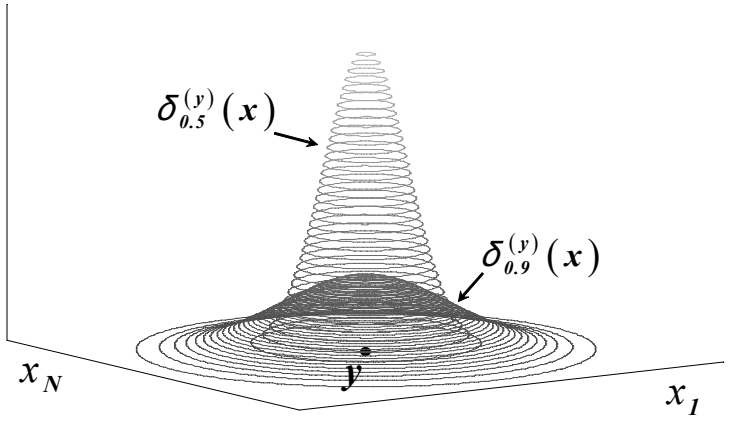

Fig. B.2. Approximation of the Dirac delta with Gaussian exponentials

we recall that in its definition we extended the space $L_2\left(\mathbb{R}^N\right)$ to include all the natural limit operations, in such a way to make it complete, see $(B.10)$.

We summarize in the table below the analogies between the basis in the finite-dimensional Euclidean vector space $\mathbb{R}^N$ and in the infinite-dimensional Hilbert space $L_2\left(\mathbb{R}^N\right)$ respectively.

	linear algebra	**functional analysis**
basis	$\left\{\mathbf{e}^{(n)}\right\}_{n\in\{1,\ldots,N\}}$	$\left\{e^{(\mathbf{y})}\right\}_{\mathbf{y}\in\mathbb{R}^N}$
canonical basis	$\left\langle \mathbf{v}, \boldsymbol{\delta}^{(n)} \right\rangle = v_n$	$\left\langle v, \delta^{(\mathbf{y})} \right\rangle = v(\mathbf{y})$

(B.20)

As an application, consider a random a variable $\mathbf{X}$ that takes on specific values $\mathbf{x}_1, \mathbf{x}_2, \ldots$ with finite probabilities $p_{\mathbf{x}_1}, p_{\mathbf{x}_2}, \ldots$ respectively. The variable $\mathbf{X}$ has a *discrete distribution*. In this situation no regular probability density function $f_{\mathbf{X}}$ can satisfy (2.4). Indeed, if this were the case, the following equality would hold:

$$p_{\mathbf{x}_i} = \mathbb{P}\{\mathbf{X} = x_i\} = \int_{\{\mathbf{x}_i\}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}. \tag{B.21}$$

Nevertheless, the integral of any regular function on the singleton $\{\mathbf{x}_i\}$ is null. On the other hand, if we express the probability density function $f_{\mathbf{X}}$ as a generalized function this problem does not exist. Indeed, if we express $f_{\mathbf{X}}$ in terms of the Dirac delta $(B.16)$ as follows:

$$f_{\mathbf{X}} = \sum_i p_{\mathbf{x}_i} \delta^{(\mathbf{x}_i)}, \tag{B.22}$$

then this generalized function satisfies $(B.21)$.

In particular, consider the case of a discrete random variable that can only take on one specific value $\widetilde{\mathbf{x}}$ with associated probability $p_{\widetilde{\mathbf{x}}} \equiv 1$: this is not a random variable, as the outcome of its measurement is known with certainty. Instead, it is a constant vector $\widetilde{\mathbf{x}}$. The formalism of generalized functions allows us to treat constants as special cases of a random variable. The visualization of the probability density function of this "not-too-random" variable in terms of the regularized Dirac delta is a bell-shaped function centered around $\widetilde{\mathbf{x}}$ that spikes to infinity as the approximation becomes exact.

B.3 Linear operators

Consider in analogy with $(A.17)$ a transformation A that maps functions v of the Hilbert space $L_2\left(\mathbb{R}^N\right)$ into functions that might belong to the same Hilbert space, or to some other space of functions F:

$$A: \; v \in L_2\left(\mathbb{R}^N\right) \mapsto u \equiv A[v] \in F. \tag{B.23}$$

In the context of functional analysis, such transformations are called *functionals* or *operators* and generalize the finite-dimensional concept of function.

In analogy with $(A.18)$, a functional A is called a *linear operator* if it preserves the sum and the multiplication by a scalar:

$$\begin{aligned} A[u+v] &= A[u] + A[v] \\ A[\alpha v] &= \alpha A[v]. \end{aligned} \tag{B.24}$$

Geometrically, this means that infinite-dimensional parallelotopes are mapped into infinite-dimensional parallelotopes, as represented in Figure A.3.

For example, consider the *differentiation operator*:

$$\mathcal{D}_n[v](\mathbf{x}) \equiv \frac{\partial v(\mathbf{x})}{\partial x_n}. \tag{B.25}$$

This operator is defined on a subset of smooth functions in $L_2\left(\mathbb{R}^N\right)$. It is easy to check that the differentiation operator is linear.

The inverse A^{-1} of a linear operator A is the functional that applied either before or after the linear operator A cancels the effect of the operator A. In other words, in analogy with $(A.20)$, for all functions v the inverse functional A^{-1} satisfies:

$$\left(A^{-1} \circ A\right)[v] = v = \left(A \circ A^{-1}\right)[v]. \tag{B.26}$$

As in the finite-dimensional case, in general the inverse transformation is not defined. If it is defined, it is linear.

For example, consider the *integration operator*, defined as follows:

$$\mathcal{I}_n[v](\mathbf{x}) \equiv \int_{-\infty}^{x_n} v(x_1, \ldots, z_n, \ldots x_N)\, dz_n \tag{B.27}$$

The *fundamental theorem of calculus* states that the integration operator is the inverse of the differentiation operator $(B.25)$. It is easy to check that the integration operator is linear.

B.3.1 Kernel representations

In Appendix A.3.1 we saw that in the finite-dimensional case every linear transformation A admits a matrix representation A_{mn}. Therefore we expect that every linear operator on $L_2\left(\mathbb{R}^N\right)$ be expressible in terms of a continuous version of a matrix. By means of the notational analogies of Table B.4 and Table B.11 this "continuous" matrix must be an integral *kernel*, i.e. a function $A(\mathbf{y}, \mathbf{x})$ such that

$$A[v](\mathbf{y}) \equiv \int_{\mathbb{R}^N} A(\mathbf{y}, \mathbf{x})\, v(\mathbf{x})\, d\mathbf{x}, \tag{B.28}$$

which parallels $(A.25)$. Such a representation does not exist in general. The operators that admit a kernel representation are called *Hilbert-Schmidt operators*. Nevertheless, we can always find and use an approximate kernel, which becomes exact in a limit sense.

For example the kernel of the differentiation operator $(B.25)$ is not defined (we consider the one-dimensional case for simplicity). Nevertheless, the following kernel is well defined:

$$D^\epsilon(y, x) \equiv \frac{1}{\epsilon}\left(\delta_\epsilon^{(y+\epsilon)}(x) - \delta_\epsilon^{(y)}(x)\right), \tag{B.29}$$

where $\delta_\epsilon^{(y)}$ is the one-dimensional approximate Dirac delta $(B.18)$. In the limit $\epsilon \to 0$ we obtain:

$$\lim_{\epsilon \to 0} \int_{\mathbb{R}} D^\epsilon(y, x)\, v(x)\, dx = [\mathcal{D}v](y). \tag{B.30}$$

Therefore $(B.29)$ is the kernel representation of the differentiation operator in a limit sense.

B.3.2 Unitary operators

Unitary operators are the generalization of rotations to the infinite-dimensional world of functional analysis. Therefore, in analogy with $(A.30)$, an operator U

is unitary if it does not alter the length, i.e. the norm $(B.8)$, of any function in $L_2\left(\mathbb{R}^N\right)$:

$$\|U[v]\| = \|v\|. \tag{B.31}$$

For example, it is immediate to check that the *reflection operator* defined below is unitary:

$$\text{Refl}[v](\mathbf{x}) \equiv v(-\mathbf{x}). \tag{B.32}$$

Similarly the *shift operator* defined below is unitary:

$$\text{Shift}_{\mathbf{a}}[v](\mathbf{x}) \equiv v(\mathbf{x}-\mathbf{a}). \tag{B.33}$$

The most notable application of unitary operators is the *Fourier transform*. This transformation is defined in terms of its kernel representation as follows:

$$\mathcal{F}[v](\mathbf{y}) \equiv \int_{\mathbb{R}^N} e^{i\mathbf{y}'\mathbf{x}} v(\mathbf{x})\, d\mathbf{x}. \tag{B.34}$$

We prove below that for all functions in $L_2\left(\mathbb{R}^N\right)$ the following result holds:

$$\|\mathcal{F}[v]\| = (2\pi)^{\frac{N}{2}} \|v\|. \tag{B.35}$$

Therefore the Fourier transform is a (rescaled) unitary operator.

For example, consider the normal probability density function (2.156):

$$f^{\text{N}}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) \equiv \frac{1}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{\mu}-\mathbf{x})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\mathbf{x})}. \tag{B.36}$$

From (2.14), the Fourier transform of the normal pdf is the characteristic function of the normal distribution. From (2.157) it reads:

$$\mathcal{F}\left[f^{\text{N}}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}\right](\mathbf{y}) = e^{i\boldsymbol{\mu}'\mathbf{y}-\frac{1}{2}\mathbf{y}'\boldsymbol{\Sigma}\mathbf{y}}. \tag{B.37}$$

In particular from $(B.37)$ and $(B.18)$, i.e. the fact that in the limit $\boldsymbol{\Sigma} \to \mathbf{0}$ the normal density $f^{\text{N}}_{\mathbf{x},\boldsymbol{\Sigma}}$ becomes the Dirac delta $\delta^{(\mathbf{x})}$, we obtain the following notable result:

$$\mathcal{F}\left[\delta^{(\mathbf{x})}\right] = \exp(i\mathbf{x}'\cdot). \tag{B.38}$$

In a Euclidean space rotations are invertible and the inverse is a rotation. Similarly, a unitary operator is always invertible and the inverse is a unitary operator. Furthermore, in a Euclidean space the representation of the inverse rotation is the transpose matrix, see $(A.31)$. Similarly, it is possible to prove that the kernel representation of the inverse of a unitary transformation is the complex conjugate of the kernel representation of the unitary operator. In formulas:

$$U^{-1}[v](\mathbf{x}) = \int_{\mathbb{R}^N} \overline{U(\mathbf{y},\mathbf{x})} v(\mathbf{y})\, d\mathbf{y}. \tag{B.39}$$

By this argument, the *inverse Fourier transform* is defined in terms of its kernel representation as follows:

$$\mathcal{F}^{-1}[v](\mathbf{x}) \equiv (2\pi)^N \int_{\mathbb{R}^N} e^{-i\mathbf{x}'\mathbf{y}} v(\mathbf{y})\, d\mathbf{y}, \tag{B.40}$$

where the factor $(2\pi)^N$ appears because the Fourier transform is a rescaled unitary transformation.

In particular, inverting $(B.38)$ and substituting $v(\mathbf{y}) \equiv e^{i\mathbf{z}'\mathbf{y}}$ in $(B.40)$ we obtain the following useful identity:

$$\delta^{(\mathbf{z})}(\mathbf{x}) = \mathcal{F}^{-1}[\exp(i\mathbf{z}'\cdot)](\mathbf{x}) = (2\pi)^N \int_{\mathbb{R}^N} e^{i(\mathbf{z}-\mathbf{x})'\mathbf{y}} d\mathbf{y}. \tag{B.41}$$

Using this identity we can show that the Fourier transform is a rescaled unitary transformation. Indeed:

$$\begin{aligned} \|\mathcal{F}[v]\|^2 &\equiv \int_{\mathbb{R}^N} \left[\int_{\mathbb{R}^N} e^{i\mathbf{y}'\mathbf{x}} v(\mathbf{x})\, d\mathbf{x}\right] \overline{\left[\int_{\mathbb{R}^N} e^{i\mathbf{y}'\mathbf{z}} v(\mathbf{z})\, d\mathbf{z}\right]} d\mathbf{y} \\ &= \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \left(\int_{\mathbb{R}^N} e^{\left(i\mathbf{y}'(\mathbf{x}-\mathbf{z})\right)} d\mathbf{y}\right) v(\mathbf{x})\, \overline{v(\mathbf{z})} d\mathbf{x} d\mathbf{z} \\ &= (2\pi)^N \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \delta^{(\mathbf{z})}(\mathbf{x})\, v(\mathbf{x})\, \overline{v(\mathbf{z})} d\mathbf{x} d\mathbf{z} = (2\pi)^N \|v\|^2. \end{aligned} \tag{B.42}$$

B.4 Regularization

In the Hilbert space of functions $L_2\left(\mathbb{R}^N\right)$ it is possible to define another operation that turns out very useful in applications.

The *convolution* of two functions u and v in this space is defined as follows:

$$[u * v](\mathbf{x}) \equiv \int_{\mathbb{R}^N} u(\mathbf{y})\, v(\mathbf{x}-\mathbf{y})\, d\mathbf{y}. \tag{B.43}$$

The convolution shares many of the features of the multiplication between numbers. Indeed it is commutative, associative and distributive:

$$\begin{aligned} u * v &= v * u \\ (u * v) * z &= u * (v * z) \\ (u + v) * z &= u * z + v * z. \end{aligned} \tag{B.44}$$

Furthermore, the Fourier transform $(B.34)$ of the convolution of two functions is the product of the Fourier transforms of the two functions:

$$\mathcal{F}[u * v] = \mathcal{F}[u]\,\mathcal{F}[v]. \tag{B.45}$$

This follows from the series of identities:

$$\begin{aligned}\mathcal{F}[u*v](\mathbf{y}) &\equiv \int_{\mathbb{R}^N} e^{i\mathbf{y}'\mathbf{x}}[u*v](\mathbf{x})\,d\mathbf{x}\\ &= \int_{\mathbb{R}^N} u(\mathbf{z})\left[\int_{\mathbb{R}^N} e^{i\mathbf{y}'\mathbf{x}} v(\mathbf{x}-\mathbf{z})\,d\mathbf{x}\right]d\mathbf{z} \qquad \text{(B.46)}\\ &= \int_{\mathbb{R}^N} u(\mathbf{z})\left[e^{i\mathbf{y}'\mathbf{z}}\mathcal{F}[v](\mathbf{y})\right]d\mathbf{z}\\ &= \mathcal{F}[v](\mathbf{y})\,\mathcal{F}[u](\mathbf{y}).\end{aligned}$$

An important application of the convolution stems from the immediate result that the Dirac delta $(B.16)$ centered in zero is the neutral element of the convolution:

$$\left[\delta^{(\mathbf{0})} * v\right] = v. \qquad \text{(B.47)}$$

By approximating the Dirac delta with the smooth function $\delta_\epsilon^{(\mathbf{0})}$ defined in $(B.18)$, we obtain from $(B.47)$ an approximate expression for a generic function v, which we call the *regularization* of v with *bandwidth* ϵ:

$$v_\epsilon \equiv \left[\delta_\epsilon^{(\mathbf{0})} * v\right] \approx v. \qquad \text{(B.48)}$$

From $(B.43)$, the regularization of v reads explicitly as follows:

$$v_\epsilon(\mathbf{x}) \equiv \frac{1}{(2\pi)^{\frac{N}{2}}\,\epsilon^N}\int_{\mathbb{R}^N} e^{-\frac{(\mathbf{y}-\mathbf{x})'(\mathbf{y}-\mathbf{x})}{2\epsilon^2}} v(\mathbf{y})\,d\mathbf{y}. \qquad \text{(B.49)}$$

Due to the bell-shaped profile of the Gaussian exponential in the above integral, the regularized function $v_\epsilon(\mathbf{x})$ is a *moving average* of $v(\mathbf{x})$ with its surrounding values: the effect of the surrounding values fades away as their distance from $\mathbf{x}$ increases. The size of the "important" points that determine the moving average is determined by the bandwidth ϵ of the bell-shaped Gaussian exponential.

The regularization $(B.49)$ becomes exact as the bandwidth ϵ tends to zero: indeed, in the limit where ϵ tends to zero, the Gaussian exponential tends to the Dirac delta, and we recover $(B.47)$. Furthermore, the regularized function v_ϵ is smooth: indeed, since the Gaussian exponential is smooth, the right hand side of $(B.49)$ can be derived infinite times with respect to $\mathbf{x}$.

To become more acquainted with the regularization technique we use it to compute the derivative of the Heaviside function $H^{(\mathbf{y})}$ defined in $(B.73)$. The partial derivative of the Heaviside function along any coordinate is zero everywhere, except in $\mathbf{y}$, where the limit that defines the partial derivative diverges to infinity. This behavior resembles that of the Dirac delta $\delta^{(\mathbf{y})}$, so we are led to conjecture that the combined partial derivatives of the Heaviside function are the Dirac delta:

$$(\mathcal{D}_1 \circ \cdots \circ \mathcal{D}_N)\left[H^{(\mathbf{y})}\right] = \delta^{(\mathbf{y})}. \qquad \text{(B.50)}$$

We can verify this conjecture with the newly introduced operations. First we notice that in general the convolution of a function with the Heaviside function is the combined integral $(B.27)$ of that function:

$$\begin{aligned}\left(v * H^{(\mathbf{y})}\right)(\mathbf{x}) &\equiv \int_{\mathbb{R}^N} v(\mathbf{z})\, H^{(\mathbf{y})}(\mathbf{x}-\mathbf{z})\, d\mathbf{z} \\ &= \int_{-\infty}^{x_1-y_1} \cdots \int_{-\infty}^{x_N-y_N} v(\mathbf{z})\, d\mathbf{z} = (\mathcal{I}_1 \circ \cdots \circ \mathcal{I}_N)[v](\mathbf{x}-\mathbf{y}).\end{aligned} \tag{B.51}$$

Applying this result to the regularization $(B.48)$ of the Heaviside function and recalling from $(B.27)$ that the integration is the inverse of the differentiation, we obtain:

$$\begin{aligned}(\mathcal{D}_1 \circ \cdots \circ \mathcal{D}_N)\left[H_\epsilon^{(\mathbf{y})}\right] &\equiv (\mathcal{D}_1 \circ \cdots \circ \mathcal{D}_N)\left[H^{(\mathbf{y})} * \delta_\epsilon^{(\mathbf{0})}\right] \\ &= (\mathcal{D}_1 \circ \cdots \circ \mathcal{D}_N)\left[(\mathcal{I}_1 \circ \cdots \circ \mathcal{I}_N)\left[\delta_\epsilon^{(\mathbf{y})}\right]\right] = \delta_\epsilon^{(\mathbf{y})}.\end{aligned} \tag{B.52}$$

Taking the limit $\epsilon \to 0$ we obtain the proof of the conjecture $(B.50)$.

Using $(B.50)$ we can compute the cumulative distribution function of a discrete distribution, whose probability density function is $(B.22)$. Indeed, from the definition of cumulative distribution function (2.10) we obtain:

$$F_{\mathbf{X}} \equiv (\mathcal{I}_1 \circ \cdots \circ \mathcal{I}_N)\left[\sum_{\mathbf{x}_i} p_{\mathbf{x}_i} \delta^{(\mathbf{x}_i)}\right] = \sum_{\mathbf{x}_i} p_{\mathbf{x}_i} H^{(\mathbf{x}_i)}, \tag{B.53}$$

where we used the fact that the integration operator is linear.

An important application of the regularization technique concerns the probability density function $f_{\mathbf{X}}$ of a generic random variable $\mathbf{X}$. Indeed, consider the regularization $(B.49)$ of $f_{\mathbf{X}}$, which can be a very irregular function, or even a generalized function:

$$f_{\mathbf{X};\epsilon}(\mathbf{x}) \equiv \frac{1}{(2\pi)^{\frac{N}{2}}\, \epsilon^N} \int_{\mathbb{R}^N} e^{-\frac{(\mathbf{y}-\mathbf{x})'(\mathbf{y}-\mathbf{x})}{2\epsilon^2}} f(\mathbf{y})\, d\mathbf{y}. \tag{B.54}$$

It is immediate to check that $f_{\mathbf{X};\epsilon}$ is strictly positive everywhere and integrates to one over the entire domain: therefore it is a probability density function.

Furthermore, we notice that in general the probability density function $f_{\mathbf{X}}$ of a random variable $\mathbf{X}$ is not univocally defined. Indeed, from its very definition (2.4) the probability density function only makes sense within an integral. For instance, if we change its value at one specific point, the ensuing altered probability density is completely equivalent to the original one. More precisely, a probability density function is an equivalence class of functions that are identical *almost everywhere*, i.e. they are equal to each other except possibly on a set of zero probability (such as one point). The regularization

technique ($B.54$) provides a univocally defined, smooth and positive probability density function.

Therefore, whenever needed, we can replace the original probability density function $f_{\mathbf{X}}$ with its regularized version $f_{\mathbf{X};\epsilon}$, which is smooth and approximates the original probability density function $f_{\mathbf{X}}$ to any degree of accuracy. If necessary, we can eventually consider the limit $\epsilon \to 0$ in the final solution to our problem, in order to recover an exact answer that does not depend on the bandwidth ϵ. Nevertheless, from a more "philosophical" point of view, in most cases we do not need to consider the limit $\epsilon \to 0$ in the final solution. Indeed, in most applications it is impossible to distinguish between a statistical model based on the original probability density function $f_{\mathbf{X}}$ and one based on the regularized probability density function $f_{\mathbf{X};\epsilon}$, provided that the bandwidth ϵ is small enough. Therefore it becomes questionable which of the two probability density functions is the "real" and which is the "approximate" model.

B.5 Expectation operator

Consider a random variable $\mathbf{X}$, whose probability density function is $f_{\mathbf{X}}$. Consider a new random variable Y defined in terms of a generic function g of the original variable $\mathbf{X}$:

$$Y \equiv g(\mathbf{X}). \tag{B.55}$$

We recall that the set of functions of the random variable $\mathbf{X}$ is a vector space, since sum and multiplication by a scalar are defined in a natural way as in ($B.2$) and ($B.3$) respectively.

To get a rough idea of the possible outcomes of the random variable Y defined in ($B.55$) it is intuitive to weigh each possible outcome by its respective probability. This way we are led to the definition of the *expectation operator* associated with the distribution $f_{\mathbf{X}}$. This operator associates with any function of a random variable the probability-weighted average of all its possible outcomes:

$$\mathrm{E}\{g(\mathbf{X})\} \equiv \mathrm{E}_{\mathbf{X}}\{g\} \equiv \int_{\mathbb{R}^N} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \tag{B.56}$$

To simplify the notation we might at times drop the symbol $\mathbf{X}$. From this definition it is immediate to check that the expectation operator is linear, i.e. it satisfies ($B.24$).

The expectation operator endows the functions of $\mathbf{X}$ with a norm, and thus with the structure of Banach space. Indeed, consider an arbitrary positive number p. We define the p-norm of g as follows:

$$\|g\|_{\mathbf{X};p} \equiv (\mathrm{E}_{\mathbf{X}}\{|g|^p\})^{\frac{1}{p}}. \tag{B.57}$$

When defined, it is possible to check that this is indeed a norm, as it satisfies the properties ($A.7$). In order to guarantee that the norm is defined, we restrict the generic space of functions of $\mathbf{X}$ to the following subspace:

$$L^p_{\mathbf{X}} \equiv \{g \text{ such that } \mathrm{E}_{\mathbf{X}}\{|g|^p\} < \infty\}. \tag{B.58}$$

Given the norm, we can define the distance $\|g-h\|_{\mathbf{X};p}$ between two generic functions g and h in $L^p_{\mathbf{X}}$.

The space $L^2_{\mathbf{X}}$ is somewhat special, as it can also be endowed with the following inner product:

$$\langle g, h\rangle_{\mathbf{X}} \equiv \mathrm{E}\left\{g\overline{h}\right\}. \tag{B.59}$$

It is easy to check that in $L^2_{\mathbf{X}}$ the norm is induced by the inner product:

$$\|\cdot\|_{\mathbf{X};2} = \sqrt{\langle\cdot,\cdot\rangle_{\mathbf{X}}}. \tag{B.60}$$

Therefore, $L^2_{\mathbf{X}}$ is a Hilbert space and in addition to the properties $(A.7)$ also the Cauchy-Schwartz inequality $(A.8)$ is satisfied:

$$|\langle g, h\rangle_{\mathbf{X}}| \leq \|g\|_{\mathbf{X};2}\,\|h\|_{\mathbf{X};2}. \tag{B.61}$$

As in $(A.9)$-$(A.10)$ the equality in this expression holds if and only if $g \equiv \alpha h$ almost everywhere for some scalar α. If the scalar α is positive:

$$\langle g, h\rangle_{\mathbf{X}} = \|g\|_{\mathbf{X};2}\,\|h\|_{\mathbf{X};2}; \tag{B.62}$$

if the scalar α is negative:

$$\langle g, h\rangle_{\mathbf{X}} = -\|g\|_{\mathbf{X};2}\,\|h\|_{\mathbf{X};2}. \tag{B.63}$$

It is easy to check that the operator $\langle\cdot,\cdot\rangle_{\mathbf{X}}$ is, like all inner products, *symmetric* and *bilinear*. Explicitly, this means that for all functions g and h in $L^2_{\mathbf{X}}$:

$$\langle g, h\rangle_{\mathbf{X}} = \langle h, g\rangle_{\mathbf{X}}; \tag{B.64}$$

and that for any function g in $L^2_{\mathbf{X}}$ the application $\langle g,\cdot\rangle_{\mathbf{X}}$ is linear. This implies in particular that the inner product of a linear combination of functions with itself can be expressed as follows:

$$\left\langle \sum_{m=1}^{M} \alpha_m g_m, \sum_{m=1}^{M} \alpha_m g_m \right\rangle_{\mathbf{X}} = \boldsymbol{\alpha}'\mathbf{S}\boldsymbol{\alpha}, \tag{B.65}$$

where $\mathbf{S}$ is an $M \times M$ matrix:

$$S_{mn} \equiv \langle g_m, g_n\rangle_{\mathbf{X}}. \tag{B.66}$$

It is easy to check that this matrix is symmetric, i.e. it satisfies $(A.51)$, and positive, i.e. it satisfies $(A.52)$.

In particular, if we consider the functions

$$g_m(\mathbf{X}) \equiv X_m - \mathrm{E}\{X_m\} \tag{B.67}$$

the matrix $(B.66)$ becomes the covariance matrix (2.67):

$$\begin{aligned} S_{mn} &\equiv \langle X_m - \mathrm{E}\{X_m\}, X_n - \mathrm{E}\{X_n\}\rangle_{\mathbf{X}} \\ &= \mathrm{Cov}\{X_m, X_n\}. \end{aligned} \tag{B.68}$$

The Cauchy-Schwartz inequality $(B.61)$ in this context reads:

$$|\mathrm{Cov}\{X_m, X_n\}| \leq \mathrm{Sd}\{X_m\}\,\mathrm{Sd}\{X_n\}. \tag{B.69}$$

In particular, from $(B.62)$ and the affine equivariance of the expected value (2.56) we obtain:

$$\mathrm{Cov}\{X_m, X_n\} = \mathrm{Sd}\{X_m\}\,\mathrm{Sd}\{X_n\} \Leftrightarrow X_m = a + bX_n, \tag{B.70}$$

where a is a scalar and b is a positive scalar. Similarly, from $(B.63)$ and the affine equivariance of the expected value (2.56) we obtain:

$$\mathrm{Cov}\{X_m, X_n\} = -\mathrm{Sd}\{X_m\}\,\mathrm{Sd}\{X_n\} \Leftrightarrow X_m = a - bX_n, \tag{B.71}$$

where a is a scalar and b is a positive scalar. These properties allow us to define the correlation matrix.

B.6 Some special functions

We conclude with a list of special functions that recur throughout the text. See Abramowitz and Stegun (1974) and `mathworld.com` for more information.

The *indicator function* of a set $\mathcal{S} \in \mathbb{R}^N$ is defined as follows:

$$\mathbb{I}_{\mathcal{S}}(\mathbf{x}) \equiv \begin{cases} 1 \text{ if } \mathbf{x} \in \mathcal{S} \\ 0 \text{ if } \mathbf{x} \notin \mathcal{S}. \end{cases} \tag{B.72}$$

The *Heaviside function* $H^{(\mathbf{y})}$ is a step function:

$$H^{(\mathbf{y})}(\mathbf{x}) \equiv \begin{cases} 1 \text{ where } x_1 \geq y_1, \ldots, x_N \geq y_N \\ 0 \text{ otherwise.} \end{cases} \tag{B.73}$$

We can define equivalently the Heaviside function in terms of the indicator function $(B.72)$ as follows:

$$H^{(\mathbf{y})} \equiv \mathbb{I}_{[y_1,+\infty)\times\cdots[y_N,+\infty)}. \tag{B.74}$$

The *error function* is defined as the integral of the Gaussian exponential:

$$\mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du. \tag{B.75}$$

The error function is odd:

$$\text{erf}(-x) = -\text{erf}(x). \tag{B.76}$$

Furthermore, the error function is normalized in such a way that:

$$\text{erf}(\infty) = 1. \tag{B.77}$$

This implies the following relation for the *complementary error function*:

$$\text{erfc}(x) \equiv \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-u^2} du = 1 - \text{erf}(x). \tag{B.78}$$

The *factorial* is a function defined only on integer values:

$$n! \equiv 1 \times 2 \times 3 \times \cdots \times (n-1) \times n. \tag{B.79}$$

The *gamma function* is defined by the following integral:

$$\Gamma(a) \equiv \int_0^{+\infty} u^{a-1} \exp(-u)\, du. \tag{B.80}$$

The gamma function is an extension to the complex and real numbers of the factorial. Indeed it is easy to check from the definition $(B.80)$ that the following identity holds:

$$\Gamma(n) = (n-1)!. \tag{B.81}$$

For half-integer arguments, it can be proved that the following identity holds:

$$\Gamma\left(\frac{n}{2}\right) = \frac{(n-2)(n-4)\cdots n_0\sqrt{\pi}}{2^{\frac{n-1}{2}}}, \tag{B.82}$$

where $n_0 \equiv 1$ if n is odd and $n_0 \equiv 2$ if n is even.

The *lower incomplete gamma function* is defined as follows:

$$\gamma(x;a) \equiv \int_0^x u^{a-1} e^{-u} du. \tag{B.83}$$

The *upper incomplete gamma function* is defined as follows:

$$\Gamma(x;a) \equiv \int_x^{+\infty} u^{a-1} e^{-u} du. \tag{B.84}$$

The *lower regularized gamma function* is defined as follows:

$$P(x;a) \equiv \frac{\gamma(x;a)}{\Gamma(a)}. \tag{B.85}$$

The *upper regularized gamma function* is defined as follows:

$$Q(x;a) \equiv \frac{\Gamma(x;a)}{\Gamma(a)}. \tag{B.86}$$

The regularized gamma functions satisfy:

$$P(x;a)+Q(x;a)=1. \tag{B.87}$$

The *beta function* is defined by the following integral:

$$B(a,b)\equiv\int_0^1 u^{a-1}(1-u)^{b-1}\,du. \tag{B.88}$$

The beta function is related to the gamma function through this identity:

$$B(a,b)=\frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}. \tag{B.89}$$

The *incomplete beta function* is defined by the following integral:

$$B(x;a,b)\equiv\int_0^x u^{a-1}(1-u)^{b-1}\,du. \tag{B.90}$$

The *regularized beta function* is a normalized version of the incomplete beta function:

$$I(x;a,b)\equiv\frac{B(x;a,b)}{B(a,b)}. \tag{B.91}$$

Therefore the regularized beta function satisfies

$$I(0;a,b)=0, \qquad I(1;a,b)=1. \tag{B.92}$$

The *Bessel functions of first, second, and third kind* are solutions to the following differential equation:

$$x^2\frac{d^2w}{dx^2}+x\frac{dw}{dx}+\left(x^2-\nu^2\right)w=0.$$

In particular, the Bessel function of the second kind admits the following integral representation, see Abramowitz and Stegun (1974) p. 360:

$$\begin{aligned} Y_\nu(x)=&\frac{1}{\pi}\int_0^\pi \sin(x\sin(\theta)-\nu\theta)\,d\theta \\ &-\frac{1}{\pi}\int_0^{+\infty}\left(e^{\nu u}+e^{-\nu u}\cos(\nu\pi)\right)e^{x\sinh(u)}du. \end{aligned} \tag{B.93}$$