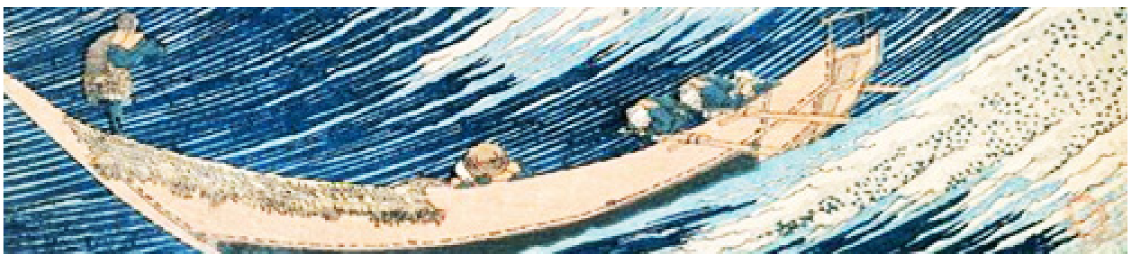
Mathematical Foundations of Data Sciences



Gabriel Peyré
CNRS & DMA
École Normale Supérieure
gabriel.peyre@ens.fr
www.gpeyre.com
www.numerical-tours.com

November 12, 2017

Chapter 2

Fourier Transforms

The main references for this chapter is [29]. The Fourier transforms offers a perfect blend of analysis (solution of PDEs, approximation of functions), algebra (characters of groups, representation theory) and computer science (the FFT). This chapter offers a glimpse of all these different facets.

2.1 Hilbert spaces and Fourier Transforms

2.1.1 Hilbertian bases.

An Hilbert space $(\mathcal{H}, \langle \cdot,\, \cdot \rangle)$ is complete. If it is separable, it can be equipped with an Hilbertian orthogonal basis $(\varphi_n)_{n\in\mathbb{N}}$, which means that one can expand any $f \in \mathcal{H}$ as

$$f = \sum_n \langle f,\, \varphi_n \rangle \varphi_n$$

where the convergence is in the sense of $\|f\|^2 \stackrel{\text{def.}}{=} \langle f,\, f \rangle$, i.e. $\|f - \sum_{n=0}^N \langle f,\, \varphi_n \rangle \varphi_n\| \to 0$ as $N \to +\infty$. One also have the conservation of energy

$$\|f\|^2 = \sum_n \langle f,\, \varphi_n \rangle^2.$$

A way to construct such an ortho-basis is using Gram-Schmidt orthogonalization procedure. On $L^2([0,1])$ equipped with the usual inner product, orthogonalization of monomials defines the Legendre polynomials. On $L^2(\mathbb{R})$ equipped with a Gaussian measure, this leads to functions of the form $P_n(x)e^{-x^2}$ where P_n are Laguerre polynomials. Intuitively, orthogonality forces φ_n to have n "oscillations", e.g. orthogonal polynomials have exactly n zeros.

Figure 2.1, left, shows examples of the real part of Fourier atoms.

2.1.2 Fourier basis on $\mathbb{R}/2\pi\mathbb{Z}$.

On $L^2(\mathbb{T})$ where $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$, equipped with $\langle f,\, g \rangle \stackrel{\text{def.}}{=} \frac{1}{2\pi}\int_{\mathbb{T}} f(x)\bar{g}(x)\mathrm{d}x$, one can use the Fourier basis

$$\varphi_n(x) \stackrel{\text{def.}}{=} e^{\mathrm{i}nx} \quad \text{for} \quad n \in \mathbb{Z}. \tag{2.1}$$

One thus has

$$f = \sum_n \hat{f}_n e^{\mathrm{i}n\cdot} \quad \text{where} \quad \hat{f}_n \stackrel{\text{def.}}{=} \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-\mathrm{i}nx}\mathrm{d}x, \tag{2.2}$$

in $L^2(\mathbb{T})$ sense. Pointwise convergence is delicate, see Section 1.2.

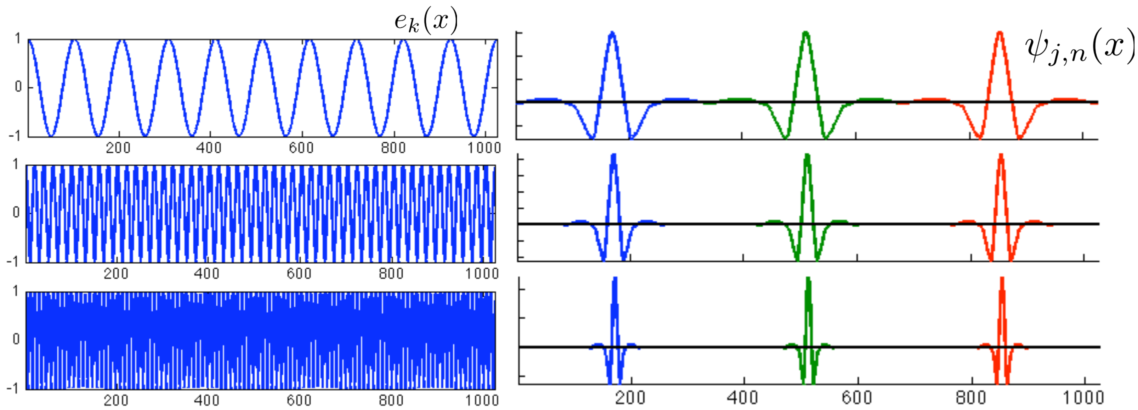


Figure 2.1: Left: 1D Fourier (real part), right: wavelet bases.

We recall that for $f \in L^1(\mathbb{R})$, its Fourier transform is defined as

$$\forall\, \omega \in \mathbb{R}, \quad \hat{f}(\omega) \overset{\text{def.}}{=} \int_{\mathbb{R}} f(x) e^{-\mathrm{i}x\omega} \mathrm{d}x.$$

and this is extended to $L^2(\mathbb{R})$ by density.

The connexion between the Fourier transform on $\mathbb{R}$ and the Fourier coefficients on $\mathbb{T}$ is given by the following diagram

$$\begin{array}{cccccc} & f(x) & \xrightarrow{\mathcal{F}} & \hat{f}(\omega) & \\ \text{sampling} & \downarrow & & \downarrow & \text{periodization} \\ & (f(n))_n & \xrightarrow{\text{Fourier serie}} & \sum_n f(n) e^{-\mathrm{i}\omega n} & \end{array}.$$

Its commutativity sates

$$\sum_n f(n) e^{-\mathrm{i}\omega n} = \sum_n \hat{f}(\omega - 2\pi n) \tag{2.3}$$

and this is in fact the celebrated Poisson formula (Proposition 1).

2.2 Convolution on $\mathbb{R}$ and $\mathbb{T}$

2.2.1 Convolution

On $\mathbb{X} = \mathbb{R}$ or $\mathbb{T}$, one defines

$$f \star g(x) = \int_{\mathbb{X}} f(t) g(x-t) \mathrm{d}t. \tag{2.4}$$

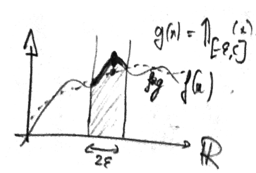


Figure 2.3: Convolution on $\mathbb{R}$.

Young's inequality shows that this quantity is well defined if $(f, g) \in L^p(\mathbb{X}) \times L^q(\mathbb{X})$

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r} \quad \Longrightarrow \quad f \star g \in L^r(\mathbb{X}) \quad \text{and} \quad \|f \star g\|_{L^r(\mathbb{X})} \leqslant \|f\|_{L^p(\mathbb{X})} \|g\|_{L^q(\mathbb{X})}. \tag{2.5}$$

This shows that if $f \in L^1(\mathbb{X})$, then one has the map $g \in L^p(\mathbb{X}) \mapsto f \star g \in L^p(\mathbb{X})$ is a continuous map on $L^p(\mathbb{X})$. Furthermore, when $r = \infty$, $f \star g \in \mathcal{C}^0(\mathbb{X})$ is a

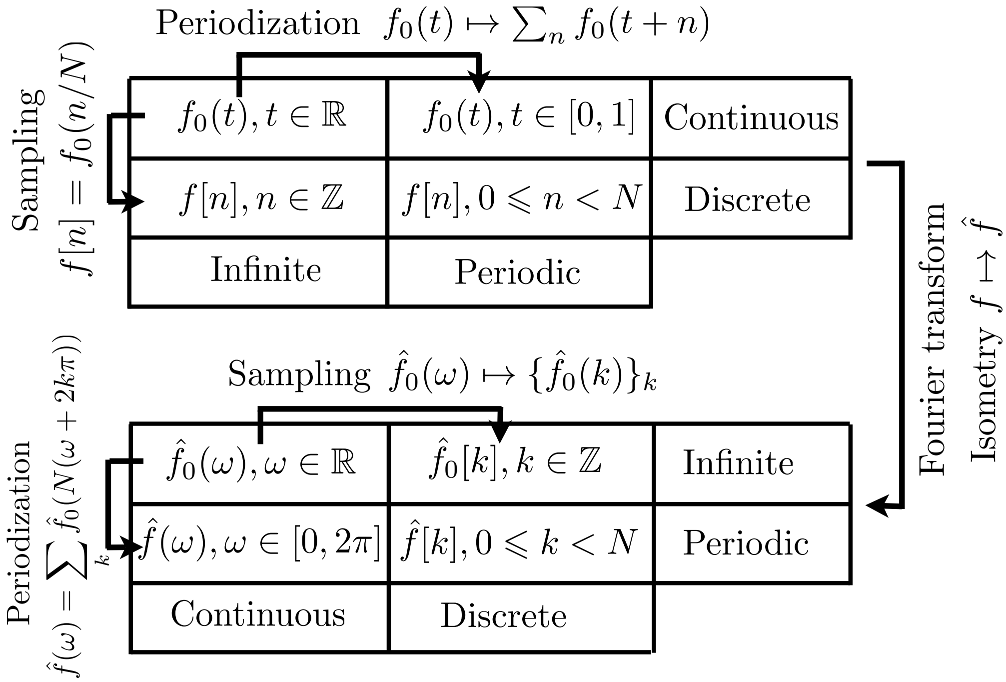


Figure 2.2: The four different settings for Fourier analysis, and the sampling-periodization relationship.

continuous function (which shows some regularizing effect). Note that for $\mathbb{X} = \mathbb{T}$, $p < q \implies L^q(\mathbb{T}) \subset L^p(\mathbb{T})$, so that $L^\infty(\mathbb{X})$ is the smallest space.

Convolution is mostly used in order to regularize functions. For instance, if $f \in L^1(\mathbb{X})$ and $g \in \mathcal{C}^1(\mathbb{X})$ is bounded, then $f \star g$ is differentiable and $(f \star g)' = f \star g'$. This is used to produce smooth approximate identity $(\rho_\varepsilon = \frac{1}{\varepsilon}\rho(\cdot/\varepsilon))_\varepsilon$ with convergence $f \star \rho_\varepsilon \to f$ in $L^p(\mathbb{X})$ for $1 \leqslant p < +\infty$ of smooth approximations (and convergence in $L^\infty(\mathbb{X})$ if f is uniformly continuous). This is also used for denoising applications in signal and image processing.

For $(f, g) \in L^1(\mathbb{X})^2$ (so on $\mathbb{X} = \mathbb{T}$, also in any $L^p(\mathbb{X})$), one has

$$\mathcal{F}(f \star g) = \hat{f} \odot \hat{g} \tag{2.6}$$

which means that $\mathcal{F}$ is a morphism of algebra. For instance if $\mathbb{X} = \mathbb{R}$, its range is included in the algebra of continuous functions with vanishing limits in $\pm\infty$.

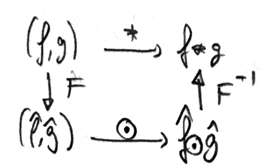


Figure 2.6: Commutative diagram of convolution-Fourier.

2.2.2 Translation Invariant Operators

Translation invariant operators (which commutes with translation) are fundamental because in most situations, input (signal, image, etc) should be processed without spatial preference. The following propositions shows that any translation invariant[1] (i.e. which commutes with translations) operator is actually a "convolution" against a distribution with bounded Fourier transform. The proof and conclusion (regularity of the convolution kernel) vary depending on the topology on the input and output spaces. We first study the case of convolution mapping to continuous functions.

Proposition 2. *We define $T_\tau f = f(\cdot - \tau)$. A bounded linear operator $H : (L^2(\mathbb{X}), \|\cdot\|_2) \to (C^0(\mathbb{X}), \|\cdot\|_\infty)$*

[1]One should rather actually say "translation equivariant".

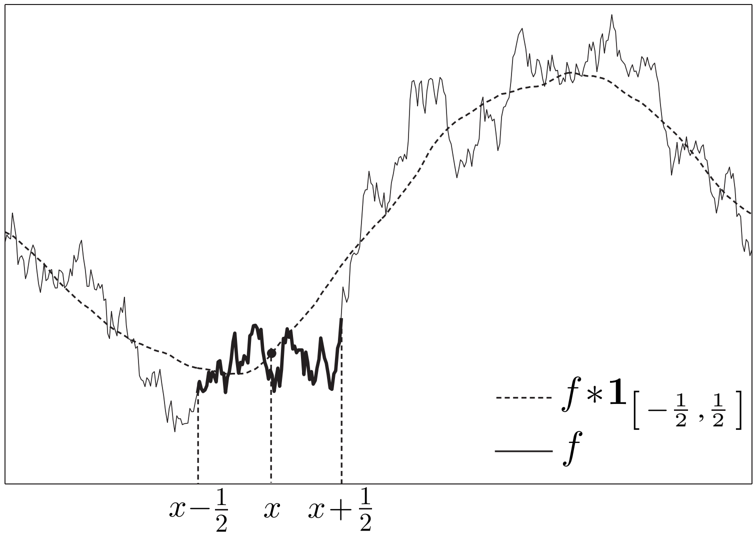


Figure 2.4: Signal filtering with a box filter (running average).

is such that for all τ, $H \circ T_\tau = T_\tau \circ H$ *if and only if*

$$\forall\, f \in L^2(\mathbb{T}), \quad H(f) = f \star g$$

with $g \in L^2(\mathbb{X})$.

Proof. Thanks to (2.5) (and the remark in the case $r = \infty$), $T : f \mapsto f \star g$ with $g \in L^2(\mathbb{X})$ is indeed a continuous operator from $L^2(\mathbb{X})$ to $C^0(\mathbb{X})$. Furthermore

$$(H \circ T_\tau)(f) = \int_{\mathbb{X}} f((\cdot - \tau) - y) g(y) \mathrm{d}\tau = (f \star g)(\cdot - \tau) = T_\tau(Hf),$$

so that such an H is translation-invariant.

Conversely, we define $\ell : f \mapsto H(f)(0) \in \mathbb{R}$, which is legit since $H(f) \in \mathcal{C}^0(\mathbb{X})$. Since H is continuous, there exists C such that $\|Hf\|_\infty \leqslant C\|f\|_2$, and hence $|\ell(f)| \leqslant C\|f\|_2$, so that f is a continuous linear form on the Hilbert space $L^2(\mathbb{X})$. Hence, according to Fréchet-Riesz theorem, there exists $h \in L^2(\mathbb{X})$ such that $\ell(f) = \langle f,\, h\rangle$. Hence, $\forall\, x \in \mathbb{X}$,

$$H(f)(x) = T_{-x}(Hf)(0) = H(T_{-x}f)(0) = \ell(T_{-x}f) = \langle T_{-x}f,\, h\rangle = \int_{\mathbb{X}} f(y+x)h(y)\mathrm{d}y = f \star \bar{h}(x).$$

where $g \overset{\text{def.}}{=} \bar{h} = h(-\cdot) \in L^2(\mathbb{X})$. $\square$

We now study, on $\mathbb{T}$, the case of convolution which can output non-continuous functions. In this case, the kernel can be a "distribution", so the convolution is defined over the Fourier domain.

$$\begin{tikzcd} f \arrow[r, "H"] \arrow[d, "T_\tau"'] & H(f) \\ f(\cdot-\tau) \arrow[r, "H"] & H(f(\cdot-\tau)) \arrow[u, "T_{-\tau}"'] \end{tikzcd}$$

Figure 2.7: Commutative diagram for translation invariance.

Proposition 3. *A bounded linear operator* $H : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ *is such that for all* τ, $H \circ T_\tau = T_\tau \circ H$ *if and only if*

$$\forall\, f \in L^2(\mathbb{T}), \quad \mathcal{F}(H(f)) = \hat{f} \odot c$$

where $c \in \ell^\infty(\mathbb{Z})$ *(a bounded sequence).*

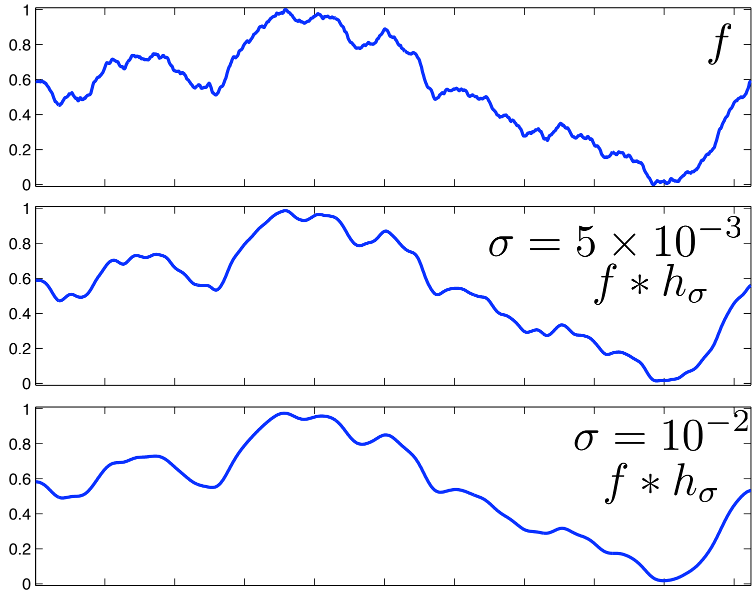


Figure 2.5: Filtering an irregular signal with a Gaussian filter of increasing filter size σ.

Proof. We denote $\varphi_n \overset{\text{def.}}{=} e^{\mathrm{i}n\cdot}$. One has

$$H(\varphi_n) = e^{\mathrm{i}n\tau} H(T_\tau(\varphi_n)) = e^{\mathrm{i}n\tau} T_\tau(H(\varphi_n)).$$

Thus, for all n,

$$\langle H(\varphi_n), \varphi_m \rangle = e^{\mathrm{i}n\tau} \langle T_\tau \circ H(\varphi_n), \varphi_m \rangle = e^{\mathrm{i}n\tau} \langle H(\varphi_n), T_{-\tau}(\varphi_m) \rangle = e^{\mathrm{i}(n-m)\tau} \langle H(\varphi_n), \varphi_m \rangle$$

So for $n \neq m$, $\langle H(\varphi_n), \varphi_m \rangle = 0$, and we define $c_n \overset{\text{def.}}{=} \langle H(\varphi_n), \varphi_n \rangle$. Since H is continuous, $\|Hf\|_{L^2(\mathbb{T})} \leqslant C\|f\|_{L^2(\mathbb{T})}$ for some constant C, and thus by Cauchy-Schwartz

$$|c_n| = |\langle H(\varphi_n), \varphi_n \rangle| \leqslant \|H(\varphi_n)\| \|\varphi_n\| \leqslant C$$

because $\|\varphi_n\| = 1$, so that $c \in \ell^\infty(\mathbb{Z})$. By continuity, recalling that by definition $\hat{f}_n \overset{\text{def.}}{=} \langle f, \varphi_n \rangle$,

$$H(f) = \lim_N H\Big(\sum_{n=-N}^{N} \hat{f}_n \varphi_n\Big) = \lim_N \sum_{n=-N}^{N} \hat{f}_n H(\varphi_n) = \lim_N \sum_{n=-N}^{N} c_n \hat{f}_n \varphi_n = \sum_{n \in \mathbb{Z}} c_n \hat{f}_n \varphi_n$$

so that in particular one has the desired result. $\square$

This theorem thus states that translation invariant operators are those which are "diagonal" in the Fourier ortho-basis.

2.2.3 Revisiting Poisson formula using distributions.

Informally, the Fourier series

$$\sum_n f(n) e^{-\mathrm{i}\omega n}$$

can be thought as the Fourier transform $\mathcal{F}(\Pi_1 \odot f)$ of the discrete distribution

$$\Pi_1 \odot f = \sum_n f(n)\delta_n \quad \text{where} \quad \Pi_s = \sum_n \delta_{sn}$$

for $s \in \mathbb{R}$, where δ_a is the Dirac mass at location $a \in \mathbb{R}$, i.e. the distribution such that $\int f \mathrm{d}(\mathrm{d}_a) = f(a)$ for any continuous f. Indeed, one can multiply a distribution by a continuous function, and the definition of the Fourier transform of a distribution μ is a distributions $\mathcal{F}(\mu)$ such that that

$$\forall\, g \in \mathcal{S}(\mathbb{R}), \quad \int_{\mathbb{R}} g(x)\mathrm{d}\mathcal{F}(\mu) = \int_{\mathbb{R}} \mathcal{F}^*(g)\mathrm{d}\mu, \quad \text{where} \quad \mathcal{F}^*(g) \stackrel{\text{def.}}{=} \int_{\mathbb{R}} g(x)e^{\mathrm{i}x\cdot}\mathrm{d}x,$$

where $\mathcal{S}(\mathbb{R})$ are smooth and rapidly decaying (Schwartz class) functions.

The Poisson formula (2.3) can thus be interpreted as

$$\mathcal{F}(\Pi_1 \odot f) = \sum_n \hat{f}(\cdot - 2\pi n) = \int_{\mathbb{R}} \hat{f}(\cdot - \omega)\mathrm{d}\Pi_{2\pi}(\omega) = \hat{f} \star \Pi_{2\pi}$$

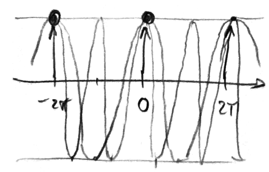


Figure 2.8: Sine wave being summed in the Poisson formula.

Since $\mathcal{F}^{-1} = \frac{1}{2\pi}\mathcal{S} \circ \mathcal{F}$ where $\mathcal{S}(f) = f(-\cdot)$, one has, applying this operator on both sides

$$\Pi_1 \odot f = \frac{1}{2\pi}\,\mathcal{S}\circ\mathcal{F}(\hat{f}\star\Pi_{2\pi}) = \mathcal{S}(\frac{1}{2\pi}\mathcal{F}(\hat{f})\odot\hat{\Pi}_{2\pi}) = \mathcal{S}(\frac{1}{2\pi}\mathcal{F}(\hat{f}))\odot\mathcal{S}(\hat{\Pi}_{2\pi}) = \hat{\Pi}_{2\pi}\odot f.$$

This can be interpreted as the relation

$$\hat{\Pi}_{2\pi} = \Pi_1 \quad \Longrightarrow \quad \hat{\Pi}_1 = 2\pi\Pi_{2\pi}.$$

To intuitively understand this relation, one can compute a finite Fourier series

$$\sum_{n=-N}^{N} e^{-\mathrm{i}n\omega} = \frac{\sin((N+1/2)x)}{\sin(x/2)}$$

which is a smooth function which grows unbounded with $N \to +\infty$ as $N \to +\infty$.

2.3 Finite Fourier Transform and Convolution

2.3.1 Discrete Ortho-bases

Discrete signals are finite dimensional vector $f \in \mathbb{C}^N$ where N is the number of samples and where each f_n is the value of the signal at a 1D or 2D location. For a 2-D images $f \in \mathbb{C}^N \simeq \mathbb{C}^{N_0 \times N_0}$, $N = N_0 \times N_0$, where N_0 is the number of pixels along each direction.

Discrete signals and images are processed using a discrete inner product that mimics the continuous L^2 inner product

$$\langle f,\, g\rangle = \sum_{n=0}^{N-1} f_n \bar{g}_n.$$

One thus defines a distance between discretized vectors as

$$\|f - g\|^2 = \sum_{n=0}^{N-1} |f_n - g_n|^2.$$

Exactly as in the continuous case, a discrete orthogonal basis $\{\psi_m\}_{0\leqslant m<N}$ of $\mathbb{C}^N$, satisfies

$$\langle \psi_m,\, \psi_{m'}\rangle = \delta_{m-m'}. \tag{2.7}$$

The decomposition of a signal in such an ortho-basis is written

$$f = \sum_{m=0}^{N-1} \langle f,\, \psi_m\rangle \psi_m.$$

It satisfies a conservation of energy

$$\|f\|^2 = \sum_{n=0}^{N-1} |f_n|^2 = \sum_{m=0}^{N-1} |\langle f,\, \psi_m\rangle|^2$$

Computing the set of all inner product $\{\langle f,\, \psi_m\rangle\}_{0\leqslant m<N}$ is done in a brute force way in $O(N^2)$ operations. This is not feasible for large datasets where N is of the order of millions. When designing an ortho-basis, one should keep this limitation in mind and enforce some structure in the basis elements so that the decomposition can be computed with fast algorithm. This is the case for the Fourier and wavelet bases, that enjoy respectively $O(N\log(N))$ and $O(N)$ algorithms.

2.3.2 Discrete Fourier transform

We denote $f = (f_n)_{n=0}^{N-1} \in \mathbb{R}^N$, but we insists that such vector should really be understood as being indexed by $n \in \mathbb{Z}/N\mathbb{Z}$, which is a finite commutative group for the addition. This corresponds to using periodic boundary conditions.

The discrete Fourier transform is defined as

$$\forall\, k = 0,\dots,N-1, \quad \hat f_k \overset{\text{def.}}{=} \sum_{n=0}^{N-1} f_n e^{-\frac{2\mathrm{i}\pi}{N}kn} = \langle f,\, \varphi_k\rangle \quad \text{where} \quad \varphi_k \overset{\text{def.}}{=} (e^{\frac{2\mathrm{i}\pi}{N}kn})_{n=0}^{N-1} \in \mathbb{C}^N \tag{2.8}$$

where the canonical inner product on $\mathbb{C}^N$ is $\langle u,\, v\rangle = \sum_{n=1}^N u_n \bar v_n$ for $(u, v) \in (\mathbb{C}^N)^2$. This definition can intuitively be motivated by sampling the Fourier basis $x \mapsto e^{\mathrm{i}kx}$ on $\mathbb{R}/2\pi\mathbb{Z}$ at equi-spaced points $(\frac{2\pi}{N}n)_{n=0}^{N-1}$. The following proposition shows that this corresponds to a decomposition in an ortho-basis.

Proposition 4. *One has*

$$\langle \varphi_k,\, \varphi_\ell\rangle = \begin{cases} N & \text{if} \quad k=\ell, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, this implies

$$\forall\, n = 0,\dots,N-1, \quad f_n = \frac{1}{N}\sum_k \hat f_k e^{\frac{2\mathrm{i}\pi}{N}kn}. \tag{2.9}$$

Proof. One has, if $k \neq \ell$

$$\langle \varphi_k,\, \varphi_\ell\rangle = \sum_n e^{\frac{2\mathrm{i}\pi}{N}(k-\ell)n} = \frac{1-e^{\frac{2\mathrm{i}\pi}{N}(k-\ell)N}}{1-e^{\frac{2\mathrm{i}\pi}{N}(k-\ell)}} = 0$$

which is the sum of a geometric serie (equivalently, sum of equi-spaced points on a circle). The inversion formula is simply $f = \sum_k \langle f,\, \varphi_k\rangle \frac{\varphi_k}{\|\varphi_k\|^2}$. □

2.3.3 Fast Fourier transform

Assuming $N = 2N'$, one has

$$\hat f_{2k} = \sum_{n=0}^{N'-1} (f_n + f_{n+N/2}) e^{-\frac{2\mathrm{i}\pi}{N'}kn}$$
$$\hat f_{2k+1} = \sum_{n=0}^{N'-1} e^{-\frac{2\mathrm{i}\pi}{N'}n} (f_n - f_{n+N/2}) e^{-\frac{2\mathrm{i}\pi}{N'}kn}.$$

For the second line, we used the computation

$$e^{-\frac{2\mathrm{i}\pi}{N}(2k+1)(n+N/2)} = e^{-\frac{2\mathrm{i}\pi}{N}(2kn+kN+n+N/2)} = -e^{-\frac{2\mathrm{i}\pi}{N'}n} e^{-\frac{2\mathrm{i}\pi}{N'}kn}.$$

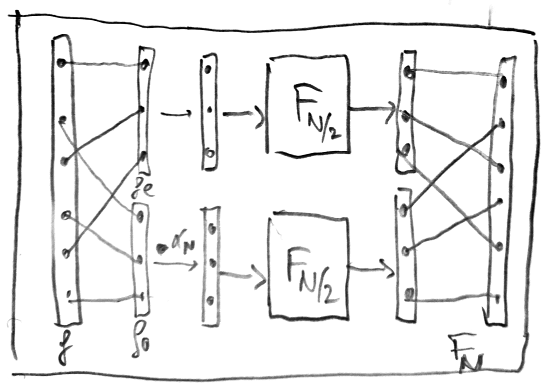


Figure 2.9: Diagram of one step of the FFT.

Denoting $\mathcal{F}_N(f) = \hat{f}$ the discrete Fourier transform on $\mathbb{R}^N$, and introducing the notation $f_e = (f_n + f_{n+N/2})_n \in \mathbb{R}^{N'}$, $f_o = (f_n - f_{n+N/2})_n \in \mathbb{R}^{N'}$ and $\alpha_N = (e^{-\frac{2i\pi}{N'}n})_n \in \mathbb{R}^{N'}$, one has the following recursion formula

$$\mathcal{F}_N(f) = \mathcal{I}_N(\mathcal{F}_{N/2}(f_e), \mathcal{F}_{N/2}(f_o \odot \alpha_N))$$

where $\mathcal{I}_N$ is the "interleaving" operator, defined by $\mathcal{I}_N(a,b) \stackrel{\text{def.}}{=} (a_1, b_1, a_2, b_2, \ldots, a_{N'}, b_{N'})$. There iterations define the so-called Fast Fourier Transform algorithm, which works here when N is a power of 2. These iterations can be extended to arbitrary number N, but a workaround is to simply pad with 0 (or use more complicated extensions) to have vector with size that are power of 2.

Denoting $C(N)$ the numerical complexity (number of elementary operations) associated to the computation of $\hat{f}$, one thus has

$$C(N) = 2C(N/2) + NK \tag{2.10}$$

where K is the complexity of N complex additions and $N/2$ multiplications. Making the change of variable

$$\ell \stackrel{\text{def.}}{=} \log_2(N) \quad \text{and} \quad T(\ell) \stackrel{\text{def.}}{=} \frac{C(N)}{N}$$

i.e. $C(N) = 2^\ell T(\ell)$, the relation (2.10) becomes

$$2^\ell T(\ell) = 2 \times 2^{\ell-1} T(\ell-1) + 2^\ell K \quad \Longrightarrow \quad T(\ell) = T(\ell-1) + K \quad \Longrightarrow \quad T(\ell) = T(0) + K\ell$$

and using the fact that $T(0) = C(1)/1 = 0$, one obtains

$$C(N) = KN\log_2(N).$$

This complexity should be contrasted with the complexity $O(N^2)$ of directly computing the N coefficients (2.8), each involving a sum of size N.

2.3.4 Finite convolution

For $(f,g) \in (\mathbb{R}^N)^2$, one defines $f \star g \in \mathbb{R}^N$ as

$$\forall\, n = 0, \ldots, N-1, \quad (f \star g)_n \stackrel{\text{def.}}{=} \sum_{k=0}^{N-1} f_k g_{n-k} = \sum_{k+\ell=n} f_k g_\ell \tag{2.11}$$

where one should be careful that here $+$ and $-$ should be understood modulo N (vectors should be seen as being defined on the group $\mathbb{Z}/N\mathbb{Z}$, or equivalently, one uses periodic boundary conditions). This defines an commutative algebra structure $(\mathbb{R}^N,+,\star)$, with neutral element the "Dirac" $\delta_0 \overset{\text{def.}}{=} (1,0,\ldots,0)^\top \in \mathbb{R}^N$. The following proposition shows that $\mathcal{F} : f \mapsto \hat{f}$ is an algebra bijective isometry (up to a scaling by $\sqrt{N}$ of the norm) mapping to $(\mathbb{R}^N,+,\odot)$ with neutral element $\mathbb{1}_N = (1,\ldots,1) \in \mathbb{R}^N$.

Proposition 5. *One has $\mathcal{F}(f\star g) = \hat{f}\odot\hat{g}$.*

Proof. We denote $T : g \mapsto f\star g$. One has

$$(T\varphi_\ell)_n = \sum_k f_k e^{\frac{2\mathrm{i}\pi}{N}\ell(n-k)} = e^{\frac{2\mathrm{i}\pi}{N}\ell n}\hat{f}_\ell.$$

This shows that $(\varphi_\ell)_\ell$ are the N eigenvectors of T with associated eigenvalues $\hat{f}_\ell$. So T is diagonalizable in this basis. Denoting $F = (e^{-\frac{2\mathrm{i}\pi}{N}kn})_{k,n}$ the matrix of the Fourier transform, the Fourier inversion formula (2.9) reads $F^{-1} = \frac{1}{N}F^*$ where $F^* = \bar{F}^\top$ is the adjoint matrix (trans-conjugate). The diagonalization of T now reads

$$T = F^{-1}\operatorname{diag}(\hat{f})F = \quad\Longrightarrow\quad \mathcal{F}(Tg) = \operatorname{diag}(\hat{f})Fg \quad\Longrightarrow\quad \mathcal{F}(f\star g) = \operatorname{diag}(\hat{f})\hat{g}.$$

$\square$

This proposition shows that one can compute in $O(N\log(N))$ operation via the formula

$$f\star g = \mathbb{F}^{-1}(\hat{f}\odot\hat{g}).$$

This is very advantageous with respect to the naive implementation of formula (2.11), in the case where f and g have large support. In case where $|\operatorname{Supp}(g)| = P$ is small, then direct implementation is $O(PN)$ which might be advantageous. An example is $g = [1,1,0,\ldots,0,1]/3$, the moving average, where

$$(f\star g)_n = \frac{f_{n-1}+f_n+f_{n+1}}{3}$$

needs $3N$ operations.

An example of application of the FFT is the multiplication of large polynomial, and thus of large integers (viewing the expansion in a certain basis as a polynomial). Indeed

$$(\sum_{i=0}^{A} a_i X_i)(\sum_{j=0}^{B} b_i X^j) = \sum_{k=0}^{A+B}(\sum_{i+j=k} a_i b_j)X^k$$

One can write $\sum_{i+j=k} a_i b_j = (\bar{a}\star\bar{b})_k$ when one defines $\bar{a},\bar{b}\in\mathbb{R}^{A+B}$ by zero padding.

2.4 Discretisation Issues

Beside computing convolutions, another major application of the FFT is to approximate the Fourier transform and its inverse, thus leading to a computationally efficient spectral interpolation method.

2.4.1 Fourier approximation via spatial zero padding.

It is possible to view the discrete finite Fourier transform (2.8) as a first order approximation to compute Fourier coefficients, or rather actually samples from the Fourier transform (1.2). Supposing that f is a smooth enough function is supported on $[0,1]$, we consider the discrete Fourier transform of the vector $f^Q \overset{\text{def.}}{=} (f(n/N))_{n=0^{Q-1}} \in \mathbb{R}^Q$ where $Q \geqslant N$ induced a padding by 0 (since $f(n/N) = 0$ for $n > N$)

$$\forall\, k \in [\![-\frac{Q}{2},\frac{Q}{2}]\!], \quad \frac{1}{N}\hat{f}^Q_k = \frac{1}{N}\sum_{n=0}^{N-1} f\left(\frac{n}{N}\right)e^{-\frac{2\mathrm{i}\pi}{Q}nk} \approx \int_0^1 f(x)e^{-\frac{2k\mathrm{i}\pi}{T}x}\mathrm{d}x = \hat{f}\left(\frac{2k\pi}{T}\right) \quad \text{where} \quad T \overset{\text{def.}}{=} \frac{Q}{N}.$$

The approximation is first order accurate, i.e. $O(1/N)$ for a C^1 function f. Increasing the amount Q of zero padding is a way to compute larger frequencies. Increasing the discretization precision N is on contrary a way to increase the precision of the Fourier sampling (using smaller step size $2\pi/T$).

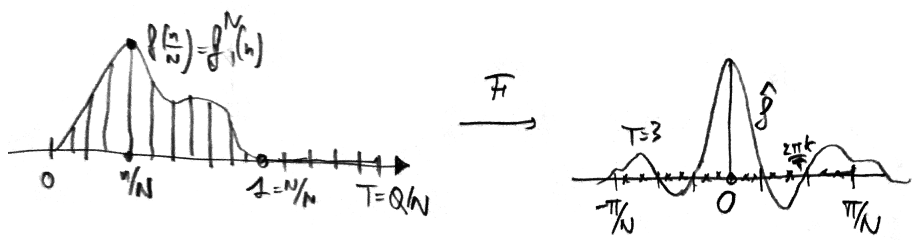


Figure 2.10: Fourier transform approximation by zero-padding in the spatial domain.

2.4.2 Fourier approximation via spatial zero padding.

If one has at its disposal N uniform discrete samples $f^N = (f_n^N)_{n=0}^{N-1}$, one can compute its discrete Fourier transform $\mathcal{F}(f^N) = \hat{f}^N$ (in $O(N\log(N))$ operations with the FFT),

$$\hat{f}_k^N \overset{\text{def.}}{=} \sum_{n=0}^{N-1} f_n^N e^{-\frac{2\mathrm{i}\pi}{N}nk},$$

and then zero-pad it to obtain a vector of length Q. For simplicity, we assume $N = 2N'+1$ is odd, and this computation can be also done (but is more involved) with even size. Indexing the frequencies as $-N' \leqslant k \leqslant N'$ The padding vector is of the form,

$$\tilde{f}^Q \overset{\text{def.}}{=} (0,\ldots,0,\hat{f}^N,0,\ldots,0) \in \mathbb{R}^Q$$

One can then compute the (with a normalization constant Q/N) inverse discrete Fourier transform of size Q (in $O(Q\log(Q))$ operations with the FFT) to obtain

$$\begin{aligned}
\frac{Q}{N}\mathcal{F}^{-1}(\tilde{f}^Q)_\ell &= \frac{Q}{N} \times \frac{1}{N} \sum_{k=-N'}^{N'} \hat{f}_k^N e^{\frac{2\mathrm{i}\pi}{Q}\ell k} = \frac{1}{N} \sum_{k=-N'}^{N'} \sum_{n=0}^{N-1} f_n^N e^{\frac{2\mathrm{i}\pi}{N}nk} e^{\frac{2\mathrm{i}\pi}{Q}\ell k} \\
&= \sum_{n=0}^{N-1} f_n^N \frac{1}{N} \sum_{k=-N'}^{N'} e^{2\mathrm{i}\pi\left(-\frac{n}{N}+\frac{\ell}{Q}\right)k} = \sum_{n=0}^{N-1} f_n^N \frac{\sin\left[\pi N\left(\frac{\ell}{Q}-\frac{n}{N}\right)\right]}{N\sin\left[\pi\left(\frac{\ell}{Q}-\frac{n}{N}\right)\right]} \\
&= \sum_{n=0}^{N-1} f_n^N \operatorname{sinc}_N\left(\frac{\ell}{T}-n\right) \quad \text{where} \quad T \overset{\text{def.}}{=} \frac{Q}{N} \quad \text{and} \quad \operatorname{sinc}_N(u) \overset{\text{def.}}{=} \frac{\sin(\pi u)}{N\sin(\pi u/N)}.
\end{aligned}$$

Here we use the following summation rule for geometric series for $\rho = e^{\mathrm{i}\omega}$, $a = -b$, $\omega = 2\pi\left(-\frac{n}{N}+\frac{\ell}{Q}\right)$,

$$\sum_{i=a}^{b} \rho^i = \frac{\rho^{a-\frac{1}{2}} - \rho^{b+\frac{1}{2}}}{\rho^{-\frac{1}{2}} - \rho^{\frac{1}{2}}} = \frac{\sin((b+\frac{1}{2})\omega)}{\sin(\omega/2)}.$$

This zero-padding method leads to a discrete version of the Shannon interpolation formula (1.9), which allows to comput the interpolation on a grid of size Q are cost $O(Q\log(Q))$. Increasing N increases the accuracy of the formula, since $\operatorname{sinc}_N \to \operatorname{sinc}$ as $N \to +\infty$.

2.5 Fourier in Multiple Dimensions

The Fourier transform is extended from 1-D to arbitrary finite dimension $d > 1$ by tensor product.

2.5.1 On Continuous Domains

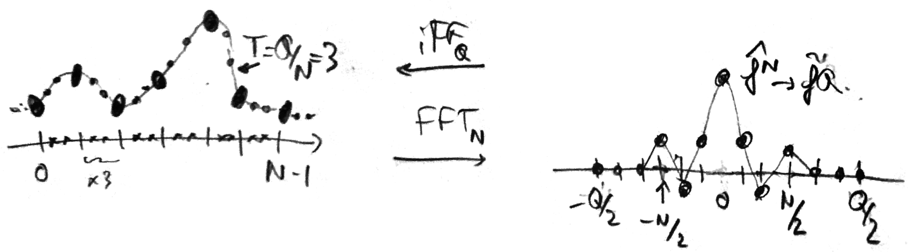


Figure 2.11: Interpolation by zero-padding in the frequency domain.

On $\mathbb{R}^d$. The crux of the power of Fourier transform in arbitrary dimension is that a product of elementary 1-D sine waves is still a sine wave

$$\prod_{\ell=1}^{d} e^{\mathrm{i}x_\ell \omega_\ell} = e^{\mathrm{i}\langle x,\, \omega\rangle}$$

moving orthogonally to the wave vector $\omega = (\omega_\ell)_{\ell=1}^d \in \mathbb{R}^d$. Here $\langle x,\, \omega\rangle = \sum_\ell x_\ell \omega_\ell$ is the canonical inner product on $\mathbb{R}^d$.

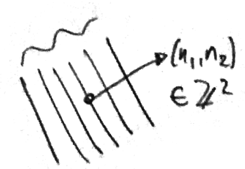


Figure 2.12: 2-D sine wave.

The definition of the Fourier transform and its inverse are

$$\forall\, \omega \in \mathbb{R}^d, \quad \hat{f}(\omega) \overset{\text{def.}}{=} \int_{\mathbb{R}^d} f(x) e^{-\mathrm{i}\langle x,\, \omega\rangle} \mathrm{d}x,$$

$$\forall\, x \in \mathbb{R}^d, \quad f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f(x) e^{\mathrm{i}\langle x,\, \omega\rangle} \mathrm{d}\omega,$$

under hypotheses of integrability matching exactly those in 1-D.

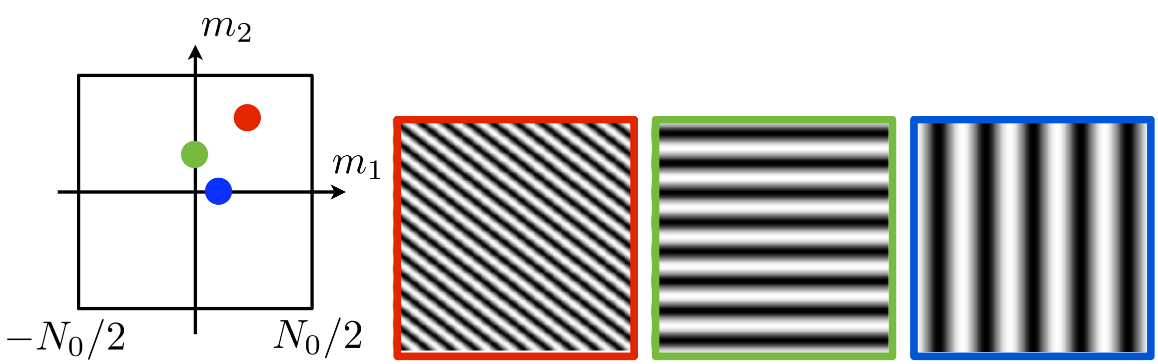


Figure 2.13: 2D Fourier orthogonal bases.

On $(\mathbb{R}/2\pi\mathbb{Z})^d$. Given an Hilbertian basis $(\varphi_{n_1})_{n_1 \in \mathbb{N}}$ of $L^2(\mathbb{X})$, one construct an Hilbertian basis of $L^2(\mathbb{X}^d)$ by tensorization

$$\forall\, n = (n_1, \ldots, n_d) \in \mathbb{N}^d, \quad \forall\, x \in \mathbb{X}^d, \quad \varphi_n(x) = \varphi_{n_1}(x_1) \ldots \varphi_{n_d}(x_d). \tag{2.12}$$

Orthogonality is simple to check, and one can also prove convergence for sum of the form $\sum_{\|n\|_\infty \leqslant N} \langle f,\, \varphi_n\rangle \varphi_n \to f$ in $L^2(\mathbb{X}^d)$.

For the multi-dimensional torus $(\mathbb{R}/2\pi\mathbb{Z})^d$, using the Fourier basis (2.1), this leads to consider the basis

$$\forall\, n \in \mathbb{R}^d, \quad \varphi_n(x) = e^{\mathrm{i}\langle x,\, n\rangle}$$

which is indeed an Hilbertian orthonormal basis for the inner product $\langle f,\, g\rangle \stackrel{\text{def.}}{=} \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} f(x)\bar{g}(x)\mathrm{d}x$. This defines the Fourier transform and the reconstruction formula on $L^2(\mathbb{T}^d)$

$$\hat{f}_n \stackrel{\text{def.}}{=} \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} f(x)e^{-\mathrm{i}\langle x,\, n\rangle}\mathrm{d}x \quad \text{and} \quad f = \sum_{n\in\mathbb{Z}^d} \hat{f}_n e^{\mathrm{i}\langle x,\, n\rangle}.$$

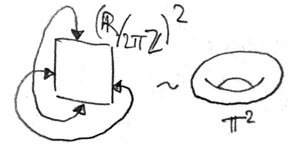


Figure 2.14: The 2-dimensional torus $\mathbb{T}^2 = (\mathbb{R}/2\pi\mathbb{Z})^2$

2.5.2 On Discrete Domains

Discrete Fourier Transform. On d-dimensional discrete domain of the form

$$n = (n_1, \ldots, n_d) \in \mathbb{Y}_d \stackrel{\text{def.}}{=} [\![1, N_1]\!] \times \ldots \times [\![1, N_d]\!]$$

(we denote $[\![a, b]\!] \stackrel{\text{def.}}{=} \{i \in \mathbb{Z}\,;\; a \leqslant i \leqslant b\}$) of $N = N_1 \ldots N_d$ points, with periodic boundary conditions, one defines an orthogonal basis $(\varphi_k)_k$ by the same tensor product formula as (2.12) but using the 1-D discrete Fourier basis (2.8)

$$\forall\, (k, n) \in \mathbb{Y}_d^2, \quad \varphi_k(n) = \varphi_{k_1}(n_1)\ldots\varphi_{k_d}(n_d) = \prod_{\ell=1}^{d} e^{\frac{2\mathrm{i}\pi}{N_\ell}k_\ell n_\ell} = e^{2\mathrm{i}\pi\langle k,\, n\rangle_{\mathbb{Y}_d}} \tag{2.13}$$

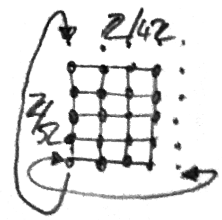


Figure 2.15: Discrete 2-D torus.

where we used the (rescaled) inner product

$$\langle k,\, n\rangle_{\mathbb{Y}_d} \stackrel{\text{def.}}{=} \sum_{\ell=1}^{d} \frac{k_\ell n_\ell}{N_\ell}. \tag{2.14}$$

The basis $(\varphi_k)_k$ is indeed orthonormal for this inner product. The Fourier transform gathers inner products in this basis, and (similarly to the 1-D case) the convention is to not normalize them with $(N_\ell)_\ell$, so that

$$\begin{aligned} \forall\, k \in \mathbb{Y}_d, \quad \hat{f}_k &\stackrel{\text{def.}}{=} \sum_{n\in\mathbb{Y}_d} f_n e^{-\mathrm{i}\langle k,\, n\rangle_{\mathbb{Y}_d}}, \\ \forall\, n \in \mathbb{Y}_d, \quad f_n &= \frac{1}{N}\sum_{k\in\mathbb{Y}_d} \hat{f}_k e^{\mathrm{i}\langle k,\, n\rangle_{\mathbb{Y}_d}}. \end{aligned}$$

Fast Fourier Transform. We detail the algorithm in dimension $d = 2$ for notation simplicity, but this extends similarly in arbitrary dimension. The general idea is that if a fast algorithm is available to compute ortho-decompositions on two 1-D bases $(\varphi^1_{k_1})_{k_1=1}^{N_1}$, $(\varphi^2_{k_2})_{k_2=1}^{N_2}$, is extended to compute decomposition on the tensor product basis $(\varphi^1_{k_1} \otimes \varphi^2_{k_2})_{k_1,k_2}$ by apply succesively the algorithm on the "rows" and then "columns" (the order does not matters) of the matrix $(f_n)_{n=(n_1,n_2)} \in \mathbb{R}^{N_1\times N_2}$. Indeed

$$\forall\, k = (k_1, k_2), \quad \langle f,\, \varphi^1_{k_1} \otimes \varphi^2_{k_2}\rangle = \sum_{n=(n_1,n_2)} f_n \varphi^1_{k_1}(n_1)\varphi^1_{k_1}(n_2) = \sum_{n_1}\left(\sum_{n_2} f_{n_1,n_2}\varphi^1_{k_1}(n_2)\right)\varphi^1_{k_1}(n_1).$$

Denoting $C(N_1)$ the complexity of the 1-D algorithm on $\mathbb{R}^{N_1}$, the complexity of the resulting 2-D decomposition is $N_2C(N_1) + N_1C(N_2)$, and hence for the FFT, it is $O(N_1N_2\log(N_1N_2)) = O(N\log(N))$ for $N = N_1N_2$.

If we represent $f \in \mathbb{R}^{N_1 \times N_2}$ as a matrix, and denote $F_N = (e^{-\frac{2i\pi}{N} kn})_{k,n}$ the Fourier transform matrix (or the matrix where rows are the φ_k^*), then one can compute the 2-D Fourier transform as matrix-matrix products

$$\hat{f} = F_{N_1} \times f \times F_{N_2}^* \in \mathbb{R}^{N_1 \times N_2}.$$

But of course these multiplications are not computed explicitly (one uses the FFT).

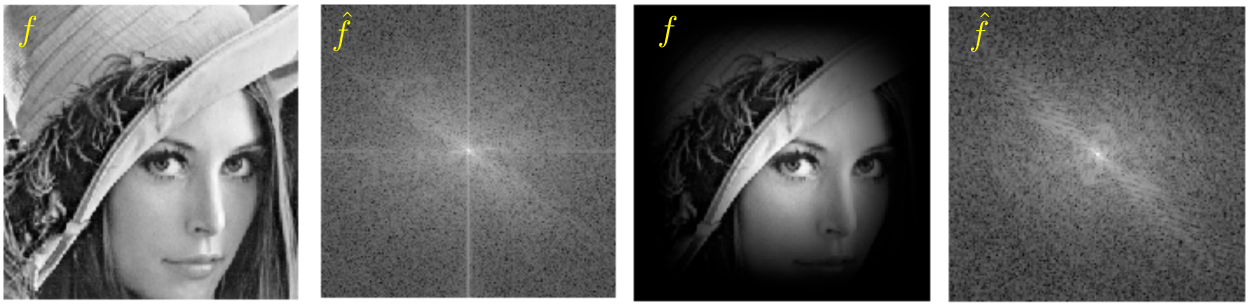


Figure 2.16: 2D Fourier analysis of a image (left), and attenuation of the periodicity artifact using masking (right).

2.5.3 Shannon sampling theorem.

The sampling Theorem 1 extends easily to $\mathbb{R}^d$ by tensorization, assuming that the sampling is on a uniform Cartesian grid. In 2-D for instance, if $\text{supp}(\hat{f}) \subset [-\pi/s_1, \pi/s_1] \times [-\pi/s_2, \pi/s_2]$ and f is decaying fast enough,

$$\forall\, x \in \mathbb{R}^2, \quad f(x) = \sum_{n \in \mathbb{Z}^2} f(n_1 s_1, n_2 s_2)\, \text{sinc}(x_1/s_1 - n_1)\, \text{sinc}(x_2/s_2 - n_2) \quad \text{where} \quad \text{sinc}(u) = \frac{\sin(\pi u)}{\pi u}.$$

2.5.4 Convolution in higher dimension.

Convolution on $\mathbb{X}^d$ with $\mathbb{X} = \mathbb{R}$ or $\mathbb{X} = \mathbb{R}/2\pi\mathbb{Z}$ are defined in the very same way as in 1-D (2.4) as

$$f \star g(x) = \int_{\mathbb{X}^d} f(t) g(x-t) \mathrm{d}t.$$

Similarly, finite discrete convolution of vectors $f \in \mathbb{R}^{N_1 \times N_2}$ extend formula (2.11) as

$$\forall\, n \in [\![0, N_1 - 1]\!] \times [\![0, N_2 - 1]\!], \quad (f \star g)_n \stackrel{\text{def.}}{=} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} f_k g_{n-k}$$

where additions and subtractions of vectors are performed modulo (N_1, N_2).

The Fourier-convolution theorem is still valid in all this cases, namely $\mathcal{F}(f \star g) = \hat{f} \odot \hat{g}$. In the finite case, this offers a fast $O(N \log(N))$ method to compute convolutions even if f and g do not have small support.

2.6 Application to ODEs and PDEs

2.6.1 On Continuous Domains

We here give only the intuition without formal proofs.

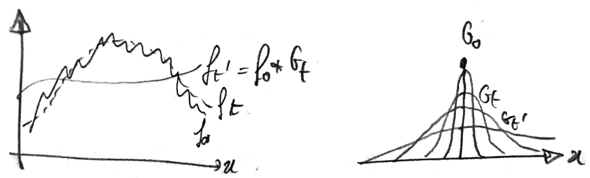


Figure 2.17: Heat diffusion as a convolution.

One $\mathbb{X} = \mathbb{R}$ or $\mathbb{T}$, one has

$$\mathcal{F}(f^{(k)})(\omega) = (\mathrm{i}\omega)^k \hat{f}(\omega).$$

Intuitively, $f^{(k)} = f \star \delta^{(k)}$ where $\delta^{(k)}$ is a distribution with Fourier transform $\hat{\delta}^{(k)}(\omega) = (\mathrm{i}\omega)^k$. Similarly on $\mathbb{X} = \mathbb{R}^d$ (see Section 2.5 for the definition of the Fourier transform in dimension d), one has

$$\mathcal{F}(\Delta f)(\omega) = -\|\omega\|^2 \hat{f}(\omega) \tag{2.15}$$

(and similarly on $\mathbb{T}$ replacing ω by $n \in \mathbb{Z}^d$). The Fourier transform (or Fourier coefficients) are thus powerful to study linear differential equations with constant coefficients, because they are turned into algebraic equations.

As a typical example, we consider the heat equation

$$\frac{\partial f_t}{\partial t} = \Delta f_t \quad \Longrightarrow \quad \forall\, \omega, \quad \frac{\partial \hat{f}_t(\omega)}{\partial t} = -\|\omega\|^2 \hat{f}(\omega).$$

This shows that $\hat{f}_t(\omega) = \hat{f}_0(\omega) e^{-\|\omega\|^2 t}$ and by inverse Fourier transform and the convolution theorem

$$f_t = G_t \star f_0 \quad \text{where} \quad G_t = \frac{1}{(4\pi t)^{d/2}} e^{-\frac{\|x\|^2}{4t}}$$

which is a Gaussian of standard deviation $\sqrt{2t}$.

2.6.2 Finite Domain and Discretization

On $\mathbb{R}/N\mathbb{Z}$ (i.e. discrete domains with periodic boundary conditions), one typically considers forward finite differences (first and second order)

$$D_1 f \stackrel{\text{def.}}{=} N(f_{n+1} - f_n)_n = f \star d_1 \quad \text{where} \quad d_1 = [-1, 0, \ldots, 0, 1]^\top \in \mathbb{R}^N, \tag{2.16}$$

$$D_2 f = D_1^\top D_1 f \stackrel{\text{def.}}{=} N^2(f_{n+1} + f_{n-1} - 2f_n)_n = f \star d_2 \quad \text{where} \quad d_2 = d_1 \star \bar{d}_1 = [2, -1, 0, \ldots, 0, -1]^\top \in \mathbb{R}^N. \tag{2.17}$$

Thanks to Proposition 5, one can alternatively computes

$$\mathcal{F}(D_2 f) = \hat{d}_2 \odot \hat{f} \quad \text{where} \quad (\hat{d}_2)_k = N^2(e^{\frac{2\mathrm{i}\pi}{N}} + e^{-\frac{2\mathrm{i}\pi}{N}} - 2) = -4N^2 \sin\left(\frac{\pi k}{N}\right)^2. \tag{2.18}$$

For $N \gg k$, one thus has $(\hat{d}_2)_k \sim -(2\pi k)^2$ which matches the scaling of (2.15).

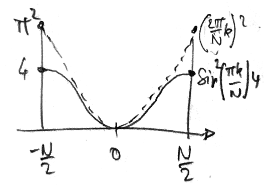


Figure 2.18: Comparison of the spectrum of Δ and D_2.

2.7 A Bit of Group Theory

The reference for this section is [32].

2.7.1 Characters

For $(G,+)$ a commutative group, a character is a group morphism χ : $(G,+) \to (\mathbb{C}^*,\cdot)$, i.e. is satisfies

$$\forall\, (n,m) \in G, \quad \chi(n+m) = \chi(n)\chi(m).$$

The set of characters is the so-called dual $(\hat{G},\odot)$ and is a group for the pointwise multiplication $(\chi_1 \odot \chi_2)(n) \overset{\text{def.}}{=} \chi_1(n)\chi_2(n)$. Indeed, the inverse of a character χ is $\chi^{-1}(n) = \chi(-n)$.

Note that for a finite group G with $|G| = N$, then since $N \times n = 0$ for any $n \in G$, then $\chi(n)^N = \chi(Nn) = \chi(0) = 1$, so that characters assume values in the unit circle, and more precisely

$$\chi(n) \in \left\{ e^{\frac{2i\pi}{N}k} \;;\; 0 \leqslant k \leqslant N-1 \right\}. \tag{2.19}$$

So in particular $\hat{G}$ is a finite group (since there is a finite number of application between two finite sets) and $\chi^{-1} = \bar{\chi}$. In the case of a cyclic group, the dual is actually simple to describe.

Proposition 6. *For $G = \mathbb{Z}/N\mathbb{Z}$, then $\hat{G} = (\varphi_k)_{k=0}^{N-1}$ where $\varphi_k = (e^{\frac{2i\pi}{N}nk})_n$ and $k \mapsto \varphi_k$ defines a (non-canonical) isomorphism $G \sim \hat{G}$.*

Proof. The φ_k are indeed characters.

Conversely, for any $\chi \in \hat{G}$, according to (2.19), $\chi(1) = e^{\frac{2i\pi}{N}k}$ for some k. Then

$$\chi(n) = \chi(1)^n =^{\frac{2i\pi}{N}kn} = \varphi_k(n).$$

Note that all these applications are different (because $\varphi_k(1)$ are all distincts) which shows that $|G| = |\hat{G}|$ so that they are isomorphic. ☐

This proposition thus shows that characters of cyclic groups are exactly the discrete Fourier orthonormal basis defined in (2.8).

Commutative groups. For more general commutative groups with a finite number of generators, according to the celebrated structure theorem, one can "decompose" them as a product of cyclic group (which are in some sense the basic building blocks), i.e. there is the following isomorphism of groups

$$G \sim (\mathbb{Z}/N_1\mathbb{Z}) \times \ldots \times (\mathbb{Z}/N_d\mathbb{Z}) \times \mathbb{Z}^Q. \tag{2.20}$$

If G is finite, then $Q = 0$ and $N = N_1 \times N_d$. In this case, G is simply a discrete d-dimensional "rectangle" with periodic boundary conditions.

For two finite groups (G_1, G_2) one has

$$\widehat{G_1 \times G_2} = \hat{G}_1 \otimes \hat{G}_2 = \left\{ \chi_1 \otimes \chi_2 \;;\; (\chi_1,\chi_2) \in \hat{G}_1 \times \hat{G}_2 \right\}. \tag{2.21}$$

Here $\otimes$ is the tensor product of two functions

$$\forall\, (n_1,n_2) \in G_1 \times G_2, \quad (\chi_1 \otimes \chi_2)(n_1,n_2) \overset{\text{def.}}{=} \chi_1(n_1)\chi_2(n_2).$$

Indeed, one verifies that $\chi_1 \otimes \chi_2$ is a morphism, and in fact one has the factorization $\chi = \chi(\cdot,0) \otimes \chi(0,\cdot)$ because one decomposes $(n_1,n_2) = (n_1,0) + (0,n_2)$.

This construction, thanks to the structure theorem, leads to a constructive proof of the isomorphism theorem.

Proposition 7. *If G is commutative and finite then $\hat{G} \sim G$.*

Proof. The structure theorem (2.20) for $Q=0$ and the dual of a product (2.21) shows that

$$\hat{G} \sim \hat{G}_1 \otimes \ldots \otimes \hat{G}_d$$

where we denoted $G_\ell \overset{\text{def.}}{=} \mathbb{Z}/N_\ell\mathbb{Z}$. One then remark that $\hat{G}_1 \otimes \hat{G}_2 \sim G_1 \times \hat{G}_2$. One conclude thanks to Proposition 6, since one has $\hat{G}_k \sim G_k$. $\square$

Note that the isomorphism $\hat{G} \sim G$ is not "canonical" since it depends on the indexing of the roots of unity on the circle. Similarly to the case of duality of vector space, the isomorphism $\hat{\hat{G}} \sim G$ can be made canonical by considering the evaluation map

$$g \in G \longmapsto e_g \in \hat{\hat{G}} \quad \text{where} \quad \left(e_g : \chi \in \hat{G} \mapsto \chi(g) \in \mathbb{C}^*.\right)$$

Discrete Fourier transform from character's point of view. One can be even more constructive by remarking that characters in $\hat{G}_\ell$ are the discrete Fourier atoms (2.8), i.e. are of the form

$$(e^{\frac{2\mathrm{i}\pi}{N_\ell}k_\ell n_\ell})_{n_\ell=0}^{N_\ell-1} \quad \text{for some} \quad 0 \leqslant k_\ell < N_\ell.$$

Identifying G and $G_1 \times \ldots \times G_d$, by tensorizing these functions together, one thus obtains that the characters composing $\hat{G}$ are exactly the orthogonal multi-dimensional discrete Fourier basis (2.13).

2.7.2 More General cases

Infinite groups. For an infinite group with a finite number of generator, one has $Q>0$, and the definition of $\hat{G}$ should impose the continuity of the characters (and also use an invariant measure on G to define inner products). In the case $G=\mathbb{Z}$, the dual are indexed by a continuous parameter,

$$\hat{\mathbb{Z}} = \left\{\varphi_\omega : n \mapsto e^{\mathrm{i}n\omega} \in L^2(\mathbb{R}/2\pi\mathbb{Z}) \; ; \; \omega \in \mathbb{R}/2\pi\mathbb{Z}\right\}$$

so that $\hat{\mathbb{Z}} \sim \mathbb{R}/2\pi\mathbb{Z}$. The case $G=\mathbb{Z}^Q$ follows by tensorization. The $(\varphi_\omega)_\omega$ are "orthogonal" in the sense that $\langle \varphi_\omega, \varphi_{\omega'} \rangle_{\mathbb{Z}} = \delta(\omega-\omega')$ can be understood as a Dirac kernel (this is similar to the Poisson formula), where $\langle u, v \rangle_{\mathbb{Z}} \overset{\text{def.}}{=} \sum_n u_n \bar{v}_n$. The "decomposition" of a sequence $(c_n)_{n\in\mathbb{Z}}$ on the set of characters is equivalent to forming a Fourier series $\sum_n c_n e^{-\mathrm{i}n\omega}$.

Similarly, for $G=\mathbb{R}/2\pi\mathbb{Z}$, one has $\hat{G}=\mathbb{Z}$, with orthonormal characters $\varphi_n = e^{\mathrm{i}\cdot n}$, so that the decomposition of functions in $L^2(G)$ is the computation of Fourier coefficients.

Non-commutative groups. For non-commutative group, one also observe that G is not isometric to $\hat{G}$. A typical example is the symmetric group Σ_N of N elements, where one can show that $\hat{G} = \{\mathrm{Id}, \varepsilon\}$ where $\varepsilon(\sigma) = (-1)^q$ is the signature, where q is the number of permutations involved in a decomposition of $\sigma \in \Sigma_N$.

In order to study non-commutative groups, one has to replace morphisms $\chi : G \to \mathbb{C}^*$ by morphisms $\rho : G \to \mathrm{GL}(\mathbb{C}^d)$ for some d, which are called "representations" of the group G. For $(g,g') \in G$ (denoting now multiplicatively the operation on G), one should thus have $\rho(gg') = \rho(g) \circ \rho(g')$. When $d=1$, identifying $\mathrm{GL}(\mathbb{C}) \sim \mathbb{C}^*$, one retrieve the definition of characters. Note that if ρ is a representation, then $\chi(g) \overset{\text{def.}}{=} \mathrm{tr}(\rho(g))$, where tr is the trace, defines a character. In order to limit the set of such representations, one is only interested in "elementary" ones, which does not have invariant sub-spaces, and are called "irreducible" (otherwise one could create arbitrary large representation by stacking others in a block diagonal way). One can show that the dimensions of these irreducible representations sum to N and that the entries of the matrices involved in these representation define an orthogonal basis of the space of functions $f : G \to \mathbb{C}$, thus leading to a theory extending the one on commutative groups. In particular, this Fourier transform in some sense also diagonalizes convolutions over G

$$f \star h(a) \overset{\text{def.}}{=} \sum_{bc=a} f(b)h(c).$$

For the symmetric group, there is an explicit description of the set of irreducible representations.

2.8 A Bit of Spectral Theory

In order to define Fourier methods on general domains $\mathbb{X}$, one can use the aforementioned group-theoretic approach if $\mathbb{X} = G$ is a group, or also if a group acts transitively on $\mathbb{X}$. An alternative way is to describe the equivalent of Fourier basis functions as diagonalizing a specific differential operator (as we have seen in Section 2.6 that it is in some sense a way to characterise the Fourier basis). Of particular interest is the Laplacian, since it is the lowest order rotation-invariant differential operator, and that there exists natural generalization on domains such as surfaces or graphs.

2.8.1 On a Surface or a Manifold

The presentation here is very informal. One can define the Laplacian of a smooth function $f : \mathbb{X} \to \mathbb{C}$ defined on a "surface" $\mathbb{X}$ as

$$\forall\, x \in \mathbb{X}, \quad (\Delta f)(x) \overset{\text{def.}}{=} \lim_{\varepsilon \to 0} \frac{1}{\mathrm{Vol}(B_\varepsilon(x))} \int_{B_\varepsilon(x)} f(x) \mathrm{d}\mu(x) - f(x).$$

Here $\mu(x)$ is the area measure on $\mathbb{X}$, $\mathrm{Vol}(B) \overset{\text{def.}}{=} \int_B \mathrm{d}\mu(x)$, and $B_\varepsilon(x) = \{y \,;\, d_{\mathbb{X}}(x, y) \leqslant \varepsilon\}$ is the geodesic ball of radius ε at x, where $d_{\mathbb{X}}$ is the geodesic distance on $\mathbb{X}$ (length of the shortest path).

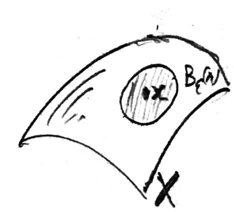


Figure 2.19: Computing Laplacian on a surface

If the surface $\mathbb{X}$ is smooth, compact and connected, then it is possible to show that Δ is itself a compact operator with a negative spectrum $0 > \lambda_1 > \lambda_2 > \ldots$ and an orthogonal set of eigenvectors $(\varphi_n)_{n \geqslant 0}$ where $\varphi_1 = 1$. Here the inner product is $\langle f, g \rangle_{\mathbb{X}} \overset{\text{def.}}{=} \int_{\mathbb{X}} f(x) g(x) \mathrm{d}\mu(x)$ on $L^2(\mathbb{X})$. In the case of a flat torus $\mathbb{X} = (\mathbb{R}/\mathbb{Z})^d$, then writing $x = (x_1, \ldots, x_d)$,

$$\Delta f = \sum_{s=1}^{d} \frac{\partial^2 f}{\partial^2 x_s}.$$

Similarly to (2.15) (which was for an unbounded domain), then one can chose for this eigen-functions φ_n the Fourier basis (2.1) and $\lambda_n = -\|n\|^2$

2.8.2 Spherical Harmonics

Of particular interest is the special case of the previous construction on the $(d-1)$-dimensional sphere $\mathbb{S}^{d-1} = \left\{x \in \mathbb{R}^d \,;\, \|x\|_{\mathbb{R}^d} = 1\right\}$. In this case, there exists a closed form expression for the eigenvectors of the Laplacian. In the 3-D case $d = 3$, they are indexed by $n = (\ell, m)$

$$\forall\, \ell \in \mathbb{N}, \quad \forall\, m = -\ell, \ldots, \ell, \quad \varphi_{\ell,m}(\theta, \varphi) = e^{\mathrm{i} m \varphi} P_\ell^m(\cos(\theta))$$

and then the eigenvalue of the Laplacian is $\lambda_{\ell,m} = -\ell(\ell+1)$. Here P_ℓ^m are associated Legendre polynomials, and we used spherical coordinates $x = (\cos(\varphi), \sin(\varphi)\sin(\theta), \sin(\varphi)\cos(\theta)) \in \mathbb{S}^3$ for $(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]$. The index ℓ is analogous to the amplitude of Fourier frequencies in 2-D. For a fixed ℓ, the space $V_\ell = \mathrm{span}(\varphi_{\ell,m})$ is an eigenspace of Δ, and is also invariant under rotation.

Figure 2.20: Spherical coordinates.

2.8.3 On a Graph

We assume $\mathbb{X}$ is a graph of N vertices, simply indexed $\{1, \ldots, N\}$. Its "geometry" is described by a connectivity matrix of weights $W = (w_{i,j})_{i \sim j}$ where we denote $i \sim j$ to indicate that (i, j) is an edge of the graph for $(i, j) \in \mathbb{X}^2$. We assume that this weight matrix and the connectity is symmetric, $w_{i,j} = w_{j,i}$.

The graph Laplacian $\Delta : \mathbb{R}^N \to \mathbb{R}^N$ is computing the difference between the average of values around a point and the value at this point

$$\forall\, f \in \mathbb{R}^N, \quad (\Delta f)_i \overset{\text{def.}}{=} \sum_{j\sim i} w_{i,j} f_j - (\sum_{j\sim i} w_{i,j}) f_i \quad \Longrightarrow \quad \Delta = W - D$$

where $D \overset{\text{def.}}{=} \mathrm{diag}_i(\sum_{j\sim i} w_{i,j})$. In particular, note $\Delta \mathbb{1} = 0$

For instance, if $\mathbb{X} = \mathbb{Z}/N\mathbb{Z}$ with the graph $i \sim i-1$ and $i \sim i+1$ (modulo N), then Δ is the finite difference Laplacian operator $\Delta = D_2$ defined in (2.17). This extends to any dimension by tensorization.

Figure 2.21: Weighted graph.

Proposition 8. *Denoting $G : f \in \mathbb{R}^N \mapsto (\sqrt{w_{i,j}}(f_i - f_j))_{i<j}$ the graph-gradient operator, one verifies that*

$$-\Delta = G^\top G \quad \Longrightarrow \quad \forall\, f \in \mathbb{R}^N, \quad \langle \Delta f,\, f\rangle_{\mathbb{R}^N} = -\langle Gf,\, Gf\rangle_{\mathbb{R}^P}.$$

where P is the number of (ordered) edges $E = \{(i,j)\ ;\ i \sim j, i < j\}$.

Proof. One has

$$\begin{aligned}
\|Gf\|^2 &= \sum_{(i,j)\in E} w_{i,j}|f_i - f_j|^2 = \sum_{i<j} w_{i,j} f_i^2 + \sum_{i<j} w_{i,j} f_j^2 - 2\sum_{i<j} w_{i,j} f_i f_j \\
&= \sum_{i<j} w_{i,j} f_i^2 + \sum_{i>j} w_{i,j} f_i^2 - \sum_{i,j} w_{i,j} f_i f_j = \sum_j f_i^2 \sum_{i,j} w_{i,j} - \sum_i f_i \sum_j w_{i,j} f_j \\
&= \langle Df,\, f\rangle - \langle Lf,\, f\rangle = -\langle Lf,\, f\rangle.
\end{aligned}$$

□

This proposition shows that Δ is a negative semi-definite operator, which thus diagonalizes in an ortho-basis $(\varphi_n)_{n=1}^N$, with $\varphi_1 = 1$, with eigenvalues $0 \geqslant \lambda_1 \geqslant \lambda_N$. If $\mathbb{X}$ is connected, one can show that $\lambda_1 < 0$. In the case of a regular graph associated to a uniform grid, one retrieves the discrete Fourier basis (2.8).

More details and application of Laplacians on graphs can be found in Chapter 3, see in particular Section 3.2.5.

Bibliography

[1] P. Alliez and C. Gotsman. Recent advances in compression of 3d meshes. In N. A. Dodgson, M. S. Floater, and M. A. Sabin, editors, *Advances in multiresolution for geometric modelling*, pages 3–26. Springer Verlag, 2005.

[2] P. Alliez, G. Ucelli, C. Gotsman, and M. Attene. Recent advances in remeshing of surfaces. In *AIM@SHAPE repport*. 2005.

[3] Amir Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. SIAM, 2014.

[4] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[6] E. Candès and D. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Commun. on Pure and Appl. Math.*, 57(2):219–266, 2004.

[7] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Compte Rendus de l'Académie des Sciences*, Serie I(346):589–592, 2006.

[8] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying. Fast discrete curvelet transforms. *SIAM Multiscale Modeling and Simulation*, 5:861–899, 2005.

[9] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20:89–97, 2004.

[10] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery*, 9(263-340):227, 2010.

[11] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[12] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

[13] F. R. K. Chung. Spectral graph theory. *Regional Conference Series in Mathematics, American Mathematical Society*, 92:1–212, 1997.

[14] Philippe G Ciarlet. Introduction à l'analyse numérique matricielle et à l'optimisation. 1982.

[15] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Multiscale Modeling and Simulation*, 4(4), 2005.

[16] P. Schroeder et al. D. Zorin. Subdivision surfaces in character animation. In *Course notes at SIGGRAPH 2000*, July 2000.

[17] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. on Pure and Appl. Math.*, 57:1413–1541, 2004.

[18] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.*, 4(3):245–267, 1998.

[19] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, Dec 1994.

[20] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[21] M. Figueiredo and R. Nowak. An EM Algorithm for Wavelet-Based Image Restoration. *IEEE Trans. Image Proc.*, 12(8):906–916, 2003.

[22] M. S. Floater and K. Hormann. Surface parameterization: a tutorial and survey. In N. A. Dodgson, M. S. Floater, and M. A. Sabin, editors, *Advances in multiresolution for geometric modelling*, pages 157–186. Springer Verlag, 2005.

[23] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Birkhäuser Basel, 2013.

[24] I. Guskov, W. Sweldens, and P. Schröder. Multiresolution signal processing for meshes. In Alyn Rockwood, editor, *Proceedings of the Conference on Computer Graphics (Siggraph99)*, pages 325–334. ACM Press, August8–13 1999.

[25] A. Khodakovsky, P. Schröder, and W. Sweldens. Progressive geometry compression. In *Proceedings of the Computer Graphics Conference 2000 (SIGGRAPH-00)*, pages 271–278, New York, July 23–28 2000. ACMPress.

[26] L. Kobbelt. $\sqrt{3}$ subdivision. In Sheila Hoffmeyer, editor, *Proc. of SIGGRAPH'00*, pages 103–112, New York, July 23–28 2000. ACMPress.

[27] M. Lounsbery, T. D. DeRose, and J. Warren. Multiresolution analysis for surfaces of arbitrary topological type. *ACM Trans. Graph.*, 16(1):34–73, 1997.

[28] S. Mallat. *A Wavelet Tour of Signal Processing, 3rd edition.* Academic Press, San Diego, 2009.

[29] Stephane Mallat. *A wavelet tour of signal processing: the sparse way.* Academic press, 2008.

[30] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Commun. on Pure and Appl. Math.*, 42:577–685, 1989.

[31] Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.

[32] Gabriel Peyré. *L'algèbre discrète de la transformée de Fourier.* Ellipses, 2004.

[33] Gabriel Peyré and Marco Cuturi. Computational optimal transport. 2017.

[34] J. Portilla, V. Strela, M.J. Wainwright, and Simoncelli E.P. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Proc.*, 12(11):1338–1351, November 2003.

[35] E. Praun and H. Hoppe. Spherical parametrization and remeshing. *ACM Transactions on Graphics*, 22(3):340–349, July 2003.

[36] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.

[37] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 2015.

[38] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, Frank Lenzen, and L Sirovich. *Variational methods in imaging.* Springer, 2009.

[39] P. Schröder and W. Sweldens. Spherical Wavelets: Efficiently Representing Functions on the Sphere. In *Proc. of SIGGRAPH 95*, pages 161–172, 1995.

[40] P. Schröder and W. Sweldens. Spherical wavelets: Texture processing. In P. Hanrahan and W. Purgathofer, editors, *Rendering Techniques '95.* Springer Verlag, Wien, New York, August 1995.

[41] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[42] A. Sheffer, E. Praun, and K. Rose. Mesh parameterization methods and their applications. *Found. Trends. Comput. Graph. Vis.*, 2(2):105–171, 2006.

[43] Jean-Luc Starck, Fionn Murtagh, and Jalal Fadili. *Sparse image and signal processing: Wavelets and related geometric multiscale analysis.* Cambridge university press, 2015.

[44] W. Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computation Harmonic Analysis*, 3(2):186–200, 1996.

[45] W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal.*, 29(2):511–546, 1997.