Wavelets and Fourier Transforms, WISM453
Part 1: Fourier Theory

G.L.G. Sleijpen
Department of Mathematics
Utrecht University

August 18, 2008

Preface

Fourier Theory belongs to the basic mathematical prerequisites in many technical and physical disciplines as Astrophysics (radar technology), Electronics, Geophysics, Information Theory, Optics, Quantum Mechanics, Spectroscopy, etc., etc.. Everybody who works in such a field should feel comfortable with Fourier transforms. But Fourier Theory also plays a fundamental role in Mathematics, in subjects as Partial Differential Equations, Numerical Analysis, Stochastics, etc.. Fourier Theory is considered to be a hard subject in Mathematics that can not be properly presented to math students in their first or second year. Some familiarity with Measure and Integration Theory and Functional Analysis is required. Mathematical textbooks spend one hundred pages or more to discuss the basic principles of Fourier Theory (see, for instance, [5, 11]), while engineering textbooks never use more than ten pages (see, for instance, [9, 8, 2]).

The order of integrals and limits is routinely exchanged in the exposition of the theory. In addition, many sequences of functions show up that are supposed to converge but of which convergence is not obvious, and often it is even not clear in what sense they converge. Physicists and engineers appear to have less problems with these type of complications: based on physical arguments, they often 'know' whether a limit function exists.

Thorough mathematical discussions of Fourier Theory often do not pass the treatment of the basics. In these notes, we want to demonstrate that there are also interesting mathematical aspects associated with applications of this theory. We can only address a few applications and a few fundamental mathematical aspects. For more interesting mathematical discussions, we refer to [10, 7].

Firstly, we collect some basic properties of the theory. At a number of places, we describe where the mathematical problems are and we indicate how they can be solved. The discussion will not be completely rigorous, but it will be more fundamental than in many technical textbooks.

Contents

1 Preliminaries

For ease of presentation, we assume that the functions in this section are complex-valued and that the scalars are complex numbers. However, the definitions and statements can also be formulated for the 'real' case.

1.1 Norms. Let $\mathcal{V}$ be a vector space (function space).
A map $\|\cdot\|$ from $\mathcal{V}$ to $\mathbb{R}$ is a *norm* if the following three properties hold:
1) $\|f\| \geq 0$; $\quad \|f\| = 0$ if and only if $f = 0 \quad (f \in \mathcal{V})$,
2) $\|f + g\| \leq \|f\| + \|g\| \quad (f, g \in \mathcal{V})$ (*triangle inequality*),
3) $\|\lambda f\| = |\lambda|\,\|f\| \quad (f \in \mathcal{V}, \lambda \in \mathbb{C})$.

A sequence (f_n) in $\mathcal{V}$ is a *Cauchy* sequence if $\lim_{n>m\to\infty} \|f_n - f_m\| = 0$.
The normed space $\mathcal{V}$ is *complete* if each Cauchy sequence (f_n) in $\mathcal{V}$ *converges* to some f in $\mathcal{V}$: $\lim_{n\to\infty} \|f_n - f\| = 0$.

The following variant of the triangle inequality is often useful (see Exercise 1.1)

$$\Big|\|f\| - \|g\|\Big| \leq \|f - g\| \qquad (f, g \in \mathcal{V}).$$

1.2 Examples.
Sup-norm. For complex-valued functions f defined on some set $\mathbf{I}$, let $\|f\|_\infty$ be defined by

$$\|f\|_\infty \equiv \sup\{|f(x)| \,\Big|\, x \in \mathbf{I}\}. \tag{1}$$

Then, $\|\cdot\|_\infty$ defines a norm on the space $\mathcal{V} = C([a, b])$ of all complex-valued continuous functions on the interval $[a, b]$. Here a, b are reals and $b > a$. $\|\cdot\|_\infty$ is called the *sup-norm* or ∞-*norm*. The space $C([a, b])$ is complete with respect to the sup-norm. If (f_n) converges to f in sup-norm, then we will also say that (f_n) converges *uniformly* on $[a, b]$ to f.[1]

The sup-norm also forms a norm on the space $L^\infty([a, b])$ of all bounded integrable functions in $[a, b]$. This space is complete with respect to the sup-norm.[2]

1-norm. For functions f on $[a, b]$ for which $|f|$ is integrable, let $\|f\|_1$ be defined by

$$\|f\|_1 \equiv \int_a^b |f(t)|\, \mathrm{d}t. \tag{2}$$

Then $\|\cdot\|_1$ defines a norm on the space $\mathcal{V} = C([a, b])$; $\|\cdot\|_1$ is the 1-*norm*.

[1] More formally: if $\varepsilon > 0$, then there exists an $N \in \mathbb{N}$ such that $\|f - f_n\|_\infty < \varepsilon$ whenever $n > N$. This implies that $|f(x) - f_n(x)| < \varepsilon$ for all $n > N$ and *all* $x \in [a, b]$. The value of N depends on ε but not on x and that is where 'uniformly' refers to: we can use the same N for all x (in formula: $\forall\varepsilon\exists N$ such that $\forall x \ldots$). The statement 'the sequence of functions (f_n) converges to the function f' may also refer to *point-wise convergence*: $\lim_{n\to\infty} f_n(x) = f(x) \ \ (x \in [a, b])$. In such a case, we have to find an N for each ε and each x (in formula: $\forall\varepsilon, x\exists N \ldots$): the N is allowed to depend on x.

In case of ambiguity, we will use the words 'uniformly' and 'point-wise' to specify in what sense the convergence has to be understood.

[2] Our functions f will be complex-valued and defined on $\mathbf{I} = \mathbb{R}$ or on some interval $\mathbf{I}$ of $\mathbb{R}$. Integrability and integrals have to be understood in the Lebesgue sense. But, for ease of interpretation, one may think of integrable functions f as functions that are continuous or continuously differentiable in all points in the domain $\mathbf{I}$ of f with the exception of finitely or countably many and for which $\int_{\mathbf{I}} |f(t)|\, \mathrm{d}t$ exists. The integral $\int_a^b |f(t)|\, \mathrm{d}t$ over $(a, b) \subset \mathbb{R}$ can be understood in a Riemannian sense. If, for instance, $f(t) = \frac{1}{\sqrt{t}}$ for $t \neq 0$ and $f(0) = 0$ then f is continuous everywhere except in 0 and f is integrable over each bounded interval.

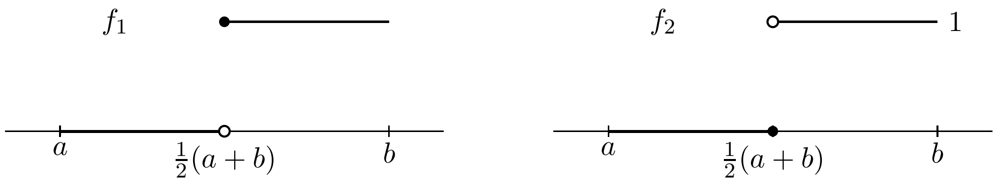


FIGURE 1. The functions f_1 (left graph) and the f_2 (right graph) coincide except at $\frac{1}{2}(a+b)$.

The space $C([a,b])$ is not complete with respect to $\|\cdot\|_1$ (see Exercise 1.4).

Let $L^1([a,b])$ be the space of all complex-valued functions f on $[a,b]$ for which $|f|$ is integrable (i.e., $\int_a^b |f(t)|\,\mathrm{d}t < \infty$). The function f, defined by $f(a) \equiv 1$ and $f(t) \equiv 0$ for all $t \in (a,b]$, shows that there are non-zero functions f in $L^1([a,b])$ for which $\|f\|_1 = 0$. Therefore, $\|\cdot\|_1$ is not a norm on $L^1([a,b])$. However, often the value of a function at a single point or at a few points is not of importance: for instance, in many application, functions f_1 and f_2 as in Fig. 1 provide the same information ($f_1(t) \equiv 0$ for $t < \frac{1}{2}(a+b)$, $f_1(t) \equiv 1$ elsewhere, and $f_2(t) \equiv 0$ for $t \leq \frac{1}{2}(a+b)$, $f_2(t) \equiv 1$ elsewhere ($t \in [a,b]$)). In these cases, there is no objection to identify functions f and g that coincide *almost everywhere*, i.e., for which the set $\mathcal{N} \equiv \{t \in [a,b] \,\big|\, f(t) \neq g(t)\}$ is *negligible*.[3] In particular, we can identify the the functions f_1 and f_2 of Fig. 1: they differ on a set that consists of one point only: $\mathcal{N} = \{\frac{1}{2}(a+b)\}$.

If, in $L^1([a,b])$, we identify functions that coincide *almost everywhere*, then $\|\cdot\|_1$ is a norm on $L^1([a,b])$ for which $L^1([a,b])$ is complete (cf., Exercise 1.4).

2-norm. For functions f on $[a,b]$ for which $|f|^2$ is integrable, let $\|f\|_2$ be defined by

$$\|f\|_2 \equiv \sqrt{\int_a^b |f(t)|^2\,\mathrm{d}t}. \tag{3}$$

Then $\|\cdot\|_2$ defines a norm on $\mathcal{V} = C([a,b])$; $\|\cdot\|_2$ is the 2-*norm*.

Let $L^2([a,b])$ be the space of functions f on $[a,b]$ that are absolute square integrable (i.e., $\int_a^b |f(t)|^2\,\mathrm{d}t < \infty$). If, in $L^2([a,b])$, we identify function that coincide almost everywhere, then $\|\cdot\|_2$ is a norm for which $L^2([a,b])$ is complete.

Quadratically summable sequences. For sequence $(\mu_k)_{k\in\mathbb{Z}}$ of complex numbers,

$$\|(\mu_k)\|_2 \equiv \sqrt{\sum_{k=-\infty}^{\infty} |\mu_k|^2} \tag{4}$$

defines a norm on the space $\ell^2(\mathbb{Z})$ of all sequences of complex numbers that, in absolute value, are quadratically summable: $\ell^2(\mathbb{Z}) \equiv \{(\mu_k)_{k\in\mathbb{Z}} \,\big|\, \sum_k |\mu_k|^2 < \infty\}$.

Some norms are induced by 'inner products'. These norms have additional properties that make them of special interest for theoretical analysis.

[3]A subset $\mathcal{N}$ of $\mathbb{R}$ is *negligible* if it has measure 0, i.e., $\int \chi_{\mathcal{N}}(t)\,\mathrm{d}t = 0$, where $\chi_{\mathcal{N}}(t) \equiv 1$ if $t \in \mathcal{N}$ and $\chi_{\mathcal{N}}(t) \equiv 0$ if $t \notin \mathcal{N}$. The negligible sets that we encounter will often consists of one or a few points only. But sets like $\{\frac{1}{n} \,\big|\, n \in \mathbb{N}\}$ are also negligible.

1.3 Inner products. Let $\mathcal{V}$ be a vector space (function space).
A map $(\cdot,\cdot)$ from $\mathcal{V}\times\mathcal{V}$ to $\mathbb{C}$ is an *inner product* if the following three properties hold:

1) $(f,f)\geq 0$; $\quad (f,f)=0$ if and only if $f=0 \quad (f\in\mathcal{V})$,
2) $(f,g)=\overline{(g,f)} \quad (f,g\in\mathcal{V})$,
3) the map $f\rightsquigarrow(f,g)$ from $\mathcal{V}$ to $\mathbb{C}$ is linear $\quad (g\in\mathcal{V})$,

that is, $(\alpha_1 f_1+\alpha_2 f_2,g)=\alpha_1(f_1,g)+\alpha_2(f_2,g) \quad (\alpha_1,\alpha_2\in\mathbb{C},\ f_1,f_2,g\in\mathcal{V})$.

If $(\cdot,\cdot)$ is an inner product on $\mathcal{V}$, then

$$\|f\|\equiv\sqrt{(f,f)} \qquad (f\in\mathcal{V}) \tag{5}$$

defines a norm on $\mathcal{V}$ (see Exercise 1.5).

We say that f is *orthogonal* to g, denoted by $f\perp g$, if $(f,g)=0$.
Pythagoras' Theorem holds for spaces with an inner product (see Exercise 1.6):

$$\|f+g\|^2=\|f\|^2+\|g\|^2 \qquad (f,g\in\mathcal{V},\ f\perp g). \tag{6}$$

Pythagoras' theorem together with Cauchy–Schwartz' inequality (below) makes the inner product to a powerful object in theoretical arguments.

1.4 Cauchy–Schwartz' inequality. *If $(\cdot,\cdot)$ is an inner product on $\mathcal{V}$ with associated norm $\|\cdot\|$, then*

$$|(f,g)|\leq\|f\|\cdot\|g\| \qquad (f,g\in\mathcal{V}). \tag{7}$$

We have equality only if f is a multiple of g.

Proof. Assume that $\|f\|=1$ and $\|g\|=1$. To show that $|(f,g)|\leq 1$, consider $\zeta\equiv(f,g)/|(f,g)|$. Then, $\overline{\zeta}\,(f,g)=|(f,g)|$ and

$$0\leq(f-\zeta g,f-\zeta g)=\|f\|^2-\overline{\zeta}\,(f,g)-\zeta\,\overline{(f,g)}+\|g\|^2=2(1-|(f,g)|).$$

For the proof of the last statement, see Exercise 1.7. □

1.5 Examples. For functions $f,g\in L^2([a,b])$, define

$$(f,g)\equiv\int_a^b f(t)\,\overline{g(t)}\,\mathrm{d}t. \tag{8}$$

If we identify functions that coincide almost everywhere, then (8) defines an inner product on $L^2([a,b])$ that is associated with the norm in (3).

Note that $\|f\|_1=\int_a^b|f(t)|\,\mathrm{d}t=(|f|,\mathbf{1})\leq\|f\|_2\,\|\mathbf{1}\|_2=\sqrt{b-a}\,\|f\|_2$. Here, $\mathbf{1}$ is the constant function $t\rightsquigarrow 1$ and we used Cauchy–Schwartz.
Moreover, $\|f\|_2^2=\int_a^b|f(t)|^2\,\mathrm{d}t\leq\int_a^b\|f\|_\infty^2\,\mathrm{d}t=(b-a)\,\|f\|_\infty^2$. In summary,

$$\|f\|_1\leq\sqrt{b-a}\,\|f\|_2, \qquad \|f\|_2\leq\sqrt{b-a}\,\|f\|_\infty. \tag{9}$$

Note that the inequalities are also correct if, say, $\|f\|_2=\infty$.

For sequences $(\mu_k),(\nu_k)\in\ell^2(\mathbb{Z})$,

$$<(\mu_k),(\nu_k)>\equiv\sum_{k=-\infty}^{\infty}\mu_k\,\overline{\nu_k} \tag{10}$$

defines an inner product on $\ell^2(\mathbb{Z})$ that is associated with the norm from (4).

1.6 We learn from the above examples that a space can be equipped with more than one norm. It depends on the application what norm is the most convenient one. For instance, in actual numerical computations in which one approximates complicated functions on $[a,b]$ with simple ones (as polynomials or finite linear combinations of sines and cosines), the sup-norm is the most popular norm for measuring the error, while the 2-norm is preferred in theoretical considerations.

The estimate $\|f\|_2 \leq \kappa\,\|f\|_\infty \quad (f \in C([a,b]))$, for $\kappa = \sqrt{b-a}$, implies

$$\lim_{n\to\infty} \|f - f_n\|_\infty = 0 \quad \Rightarrow \quad \lim_{n\to\infty} \|f - f_n\|_2 = 0. \tag{11}$$

The norms would be *equivalent* on $C([a,b])$ if we also could find a $\widetilde{\kappa}$ such that $\|f\|_\infty \leq \widetilde{\kappa}\,\|f\|_2$ for all $f \in C([a,b])$. Then convergence with respect to (w.r.t.) the 2-norm would imply convergence w.r.t. the sup-norm (i.e., the converse of (11)). Unfortunately, the norms are not equivalent and a sequence of functions that converges in the 2-norm may diverge in the sup-norm: on $[a,b] = [0,1]$, with $f_n(t) \equiv t^n \quad (t \in [0,1])$, we have that $\lim_{n\to\infty} \|f_n - \mathbf{0}\|_2 = 0$, while there is no continuous function f on $[0,1]$ for which $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$. See also Exercise 1.9.

Conditions in Fourier theory that ensure convergence w.r.t. sup-norm often involve f'. This is essentially for the following reason: if $f \in C([a,b])$ such that $f' \in L^2([a,b])$, then $f(t) = f(a) + \int_a^t f'(s)\,\mathrm{d}s$. Hence,

$$\|f\|_\infty \leq |f(a)| + \int_a^b |f'(s)|\,\mathrm{d}s \leq |f(a)| + \|f'\|_2\,\|\mathbf{1}\|_2 = |f(a)| + \sqrt{b-a}\,\|f'\|_2.$$

If $S_n(t) = f(a) + \int_a^t \widetilde{S}_n(s)\,\mathrm{d}s$, then

$$\|f - S_n\|_\infty \leq \sqrt{b-a}\,\|f' - \widetilde{S}_n\|_2.$$

In particular, $\lim_{n\to\infty} \|f' - \widetilde{S}_n\|_2 = 0 \quad \Rightarrow \quad \lim_{n\to\infty} \|f - S_n\|_\infty = 0$: we can deduce sup-norm convergence from 2-norm convergence of the derivatives.

1.7 Norms on unbounded intervals. In the above examples, we considered functions on some bounded interval $[a,b]$ in $\mathbb{R}$. However, we will also be interested in functions defined on the whole $\mathbb{R}$ (or on semi-bounded intervals as $[a,\infty)$). For these functions, natural norms are

$$\|f\|_1 = \int_{-\infty}^{\infty} |f(t)|\,\mathrm{d}t, \quad \|f\|_2 = \sqrt{\int_{-\infty}^{\infty} |f(t)|^2\,\mathrm{d}t}, \text{ and } \|f\|_\infty = \sup\{|f(t)| \,\big|\, t \in \mathbb{R}\}.$$

For simplicity of notation, we use the same notation as for norms of functions on a bounded interval $[a,b]$.

1.8 Integrable functions need not to be continuous. However, they are close to continuous functions provided that the distance is measured in 1-norm or 2-norm and not in the sup-norm. Without proof, we mention the following result.

Density theorem. [4] *Let $p = 1$ or $p = 2$. Let $f \in L^p([a,b])$.*
Then, for each $\varepsilon > 0$, there is a function $g \in C([a,b])$ such that $\|f - g\|_p < \varepsilon$. □

[4] In mathematical textbooks, the space $L^1([a,b])$ is often introduced as the *closure* in 1-norm of the space of continuous functions, i.e., as the space of functions f for which, for some sequence (f_n) of continuous functions, $\|f - f_n\|_1 \to 0 \quad (n \to \infty)$. Then the density theorem is correct by definition.

We can also find smooth functions g that are close to f: for each $\varepsilon > 0$ and each $k > 0$, there is a $g \in C^{(k)}([a,b])$ such that $\|f-g\|_p < \varepsilon$, see Exercise 1.17. Apparently, we can put stricter conditions on the smoothness of g, but we can not take $p = \infty$: for instance, for f_0, defined on $[-1,+1]$ by $f_0(t) = -1$ if $t \leq 0$ and $f_0(t) = 1$ if $t > 0$, we have that $\|f_0 - g\|_\infty \geq 1$ whenever g is continuous at 0.

The theorem is also correct if we replace $[a,b]$ by some (semi-)unbounded interval as $[a,\infty)$ or $\mathbb{R}$. Then we can find a g that is uniformly continuous and vanishes at ∞.

1.9 Point-wise convergence. Let I be a bounded or (semi-)unbounded interval in $\mathbb{R}$ (I is $[a,b]$, $[a,\infty)$, $(-\infty,b]$, or $\mathbb{R}$). Consider a sequence (f_n) of real or complex-valued functions f_n on I and a function f such that

$$\lim_{n\to\infty} f_n(t) = f(t) \qquad \text{for all} \quad t \in I:$$

the sequence (f_n) *converges point-wise* to f. The sequence (f_n) does not necessarily converge to f w.r.t. the $\|\cdot\|_2$ (or $\|\cdot\|_1$) norm even if (f_n) and f are in $L^1(I)$ (resp. $L^2(I)$); see (d) of Exercise 1.9. The following gives a condition under which we have $\|\cdot\|_1$ convergence.

Lebesgue's Dominated Convergence Theorem. *If there is a g such that*

$$|f_n(t)| \leq |g(t)| \quad \textit{for all } t \in I,\ n \in \mathbb{N} \quad \textit{and} \quad g \in L^1(I)$$

then $\displaystyle\lim_{n\to\infty} \|f_n - f\|_1 = 0.$ □

For extensions and applications, see Exercise 1.10, Exercise 1.11, and Exercise 1.12.

1.10 Integration by parts. If F and G are differentiable, $f = F'$, and $g = G'$, then $(FG)' = fG + Fg$ (*the product rule*). Integration of this expression leads to

$$\int_a^b f(t)G(t)\,\mathrm{d}t = F(b)G(b) - F(a)G(a) - \int_a^b F(t)g(t)\,\mathrm{d}t. \tag{12}$$

Here, we implicitly assumed that F and G are differentiable everywhere on $[a,b]$. However, it is often convenient to integrate by parts also if the functions fail to be differentiable in some points. Formula (12) is also correct if $F(t) = \alpha + \int_c^t f(s)\,\mathrm{d}s$, $G(t) = \beta + \int_c^t g(s)\,\mathrm{d}s$ $(t \in [a,b])$ and $f, g \in L^1([a,b])$. Here, α, β are suitable constants and c is in $[a,b]$: *integration by parts* only requires F and G to be primitives of L^1 functions. Note that such F and G are continuous. Primitives of L^1 functions are said to be *absolutely continuous*. Continuity of F and G is essential for integrating by parts! See Exercise 1.19. See also Exercise 1.18.

Example. Consider $F_0(t) \equiv |t|$ for $t \in [-1,+1]$. Then $f_0(t) = F_0'(t) = -1$ for $t < 0$ and $f_0(t) = F_0'(t) = +1$ for $t > 0$. F_0 is not differentiable at $t = 0$. However, if we define $f_0(0) = 0$ (or any other value), then $f_0 \in L^1([-1,+1])$ and $F_0(t) = \int_0^t f_0(s)\,\mathrm{d}s$.

As for F_0, the derivative of f_0 is properly defined everywhere in $[-1,+1]$ except at $t = 0$: $h_0(t) \equiv f_0'(t) = 0$ for $t \neq 0$. Note, however, that assigning some finite value to $h_0(0)$ is not helpful now: the primitive of such an h_0 will be a constant function and not a step function as f_0.

Note that F_0 is continuous at 0, in contrast to f_0.

In the sequel, we will use expression as

$$\text{`}\textit{f is differentiable and } f' \in L^1([a,b])\text{'}. \tag{13}$$

Then, we will mean that

$$g = f' \in L^1([a,b]) \quad \text{and} \quad f(t) = \alpha + \int_c^t g(s)\;\mathrm{d}s \quad (t \in [a,b]) \tag{14}$$

for some α and $c \in [a,b]$. In particular, we then have that f is continuous.

Note that the Expression (13) applies to $f = F_0$, with F_0 as in the example above.

Exercises

Exercise 1.1. Consider a space $\mathcal{V}$ equipped with a norm $\|\cdot\|$. Prove that

$$\Big|\|f\| - \|g\|\Big| \leq \|f - g\| \qquad (f, g \in \mathcal{V}). \tag{15}$$

Exercise 1.2. Prove that $\|\cdot\|_\infty$ defines a norm on the space of all bounded integrable functions on $[a,b]$.

Exercise 1.3. Which of the following maps defines a norm on the mentioned space?
(a) $f \rightsquigarrow \int_a^b |f'(t)|\;\mathrm{d}t$ on $C^{(1)}([a,b])$.
(b) $f \rightsquigarrow \int_a^b |f'(t)|\;\mathrm{d}t$ on $\{f \in C^{(1)}([a,b]) \,\big|\, f(a) = 0\}$.
(c) $f \rightsquigarrow \max(\|f\|_\infty, \|f'\|_\infty)$ on $C^{(1)}([a,b])$.

Exercise 1.4. Let f_n be defined by

$$f_n(t) = 1 \quad \text{for } t \in [1,2] \quad \text{and} \quad f_n(t) = t^n \quad \text{for } t \in [0,1].$$

Show that (f_n) is a Cauchy sequence in $C([0,2])$ with respect to $\|\cdot\|_1$. Does (f_n) converge to a function in $C([0,2])$? Does (f_n) converge to a function in $L^1([0,2])$?

Exercise 1.5.
Prove that $\|f\| \equiv \sqrt{(f,f)}$ defines a norm on $\mathcal{V}$ if $(\cdot,\cdot)$ is an inner product on $\mathcal{V}$.

Exercise 1.6. $(\,\cdot\,,\cdot\,)$ is an inner product on $\mathcal{V}$, $\|\cdot\|$ is its associated norm.
Prove Pythagoras' Theorem: $\|f+g\|^2 = \|f\|^2 + \|g\|^2$ for all $f, g \in \mathcal{V}$, $f \perp g$.

Exercise 1.7. $(\,\cdot\,,\cdot\,)$ is an inner product on $\mathcal{V}$, $\|\cdot\|$ is its associated norm.
(a) Why may we assume that both f and g are normalized in the proof of the Cauchy–Schwartz' inequality?
(b) Suppose that $\|f\| = \|g\| = 1$ and $|(f,g)| = 1$. Consider $h \equiv f - (f,g)g$.
Show that $h \perp g$. Now, show that $\|h\|^2 = (h,f) = 0$. Conclude that $f = (f,g)g$.

Exercise 1.8. $(\,\cdot\,,\cdot\,)$ is an inner product on $\mathcal{V}$, $\|\cdot\|$ is its associated norm.
Let $\phi_1, \ldots, \phi_N$ be non-zero elements in $\mathcal{V}$ that form an *orthogonal system*, i.e., $\phi_n \perp \phi_m$ if $n \neq m$. Let $\mathcal{W}$ be the subspace of $\mathcal{V}$ spanned by ϕ_n: $\mathcal{W} = \{\sum_{n=1}^N \alpha_n \phi_n \,\big|\, \alpha_n \in \mathbb{C}\}$.
(a) Show that

$$g = \sum_{n=1}^N \frac{(g,\phi_n)}{(\phi_n,\phi_n)} \phi_n \quad \text{for all } g \in \mathcal{W}. \tag{16}$$

(b) Show that $\mathcal{V} = \mathcal{W}$ if the dimension of $\mathcal{V}$ is equal to N.
(c) Let $f \in \mathcal{V}$ and $g \in \mathcal{W}$. Prove the following

$$f - g \perp \mathcal{W} \qquad \Rightarrow \qquad \|f - g\| \leq \|f - h\| \quad \text{for all } h \in \mathcal{W}. \tag{17}$$

Here, $f-g \perp \mathcal{W}$ is short hand for $f-g \perp h$ for all $h \in \mathcal{W}$. Apparently, g is the *best approximation* of f in $\mathcal{W}$ with respect to the norm $\|\cdot\|$ if $f-g$ is orthogonal to $\mathcal{W}$.

(d) Let $f \in \mathcal{V}$ and $g \in \mathcal{W}$. Prove the following

$$f-g \perp \mathcal{W} \qquad \Leftarrow \qquad \|f-g\| \leq \|f-h\| \quad \text{for all } h \in \mathcal{W}. \tag{18}$$

(Hint: Use that $\|f-g\| \leq \|f-g+\varepsilon h\|$ for all $h \in \mathcal{W}$ and all (small) $\varepsilon \in \mathbb{C}$.)

(e) Let $f \in \mathcal{V}$. Prove that the best approximation of f in $\mathcal{W}$ is unique.

(f) Let $f \in \mathcal{V}$ and $g \in \mathcal{W}$. Prove that

$$g = \sum_{n=1}^{N} \frac{(f,\phi_n)}{(\phi_n,\phi_n)}\phi_n \qquad \Leftrightarrow \qquad \|f-g\| \leq \|f-h\| \quad \text{for all } h \in \mathcal{W}. \tag{19}$$

Exercise 1.9. Let (f_n) be a sequence of complex-valued functions on $[a,b]=[0,1]$ and let f be a complex-valued function on $[0,1]$. Consider the following statements:

(1) $\lim_{n\to\infty} \|f-f_n\|_2 = 0$
(2) $\lim_{n\to\infty} \|f-f_n\|_\infty = 0$
(3) $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in [0,1]$.
Then (1) $\Leftarrow$ (2) $\Rightarrow$ (3), while all other implications are incorrect.

(a) Prove that (1) $\Leftarrow$ (2) and (2) $\Rightarrow$ (3).

(b) Show that (1) $\not\Rightarrow$ (2) and (1) $\not\Rightarrow$ (3).
(Hint: consider $f_n(t) \equiv 1-nt \ \ (nt<1)$, $f_n(t) \equiv 0 \ \ (1 \leq nt)$, and $f \equiv 0$.)

(c) Show that (2) $\not\Leftarrow$ (3). (Hint: consider $f_n(t) \equiv nt \ \ (nt<1)$, $f_n(t) \equiv 2-nt \ \ (1 \leq nt < 2)$, $f_n(t) \equiv 0 \ \ (2 \leq nt)$, and $f \equiv 0$.)

(d) Show that (1) $\not\Leftarrow$ (3). (Hint: scale the functions f_n in the hint of (c).)

Exercise 1.10. **Lebesgue's theorem.**

Let I be $[a,b]$ or $\mathbb{R}$. Let (f_n) be a sequence of functions on I that converges point-wise to a function f (see §1.9), let g be a function such that $|f_n(t)| \leq |g(t)|$ for all $t \in I$ and $n \in \mathbb{N}$.

(a) Does Lebesgue's Theorem hold w.r.t. $\|\cdot\|_2$ convergence if $g \in L^2(\mathbb{R})$
(i.e. $\|f_n - f\|_2 \to 0$ for $n \to \infty$ if $g \in L^2(\mathbb{R})$)?

(b) Does Lebesgue's Theorem hold w.r.t. $\|\cdot\|_\infty$ convergence if $g \in L^\infty(\mathbb{R})$?

Exercise 1.11. Let $f \in L^1(\mathbb{R})$ such that the function $t \rightsquigarrow t\,f(t)$ also is in $L^1(\mathbb{R})$.

(a) Prove that $g(\omega) \equiv \int_{-\infty}^{\infty} f(t)\,\sin(2\pi t\omega)\,\mathrm{d}t \quad (\omega \in \mathbb{R})$ defines a bounden function on $\mathbb{R}$.

(b) Use Lebesgue's Theorem (see §1.9) to prove that g is differentiable and that

$$g'(\omega) = 2\pi \int_{-\infty}^{\infty} t\,f(t)\,\cos(2\pi t\omega)\,\mathrm{d}t \quad (\omega \in \mathbb{R}).$$

Exercise 1.12. **Riemann Sums for functions on $\mathbb{R}$.**

Let $g \in C(\mathbb{R})$. For $h>0, T>0$, consider the Riemann Sums

$$R_{h,T} \equiv \sum hg(hk), \quad \text{where we sum over all } k \in \mathbb{Z}, kh \in [-T,T).$$

For $h>0$, let g_h be defined by $g_h(t) \equiv kh$ if $t \in [kh, kh+h)$.

(a) Show that $g_h(t) \to g(t) \ \ (h \to 0)$ for all $t \in \mathbb{R}$. Show that, for each $T>0$

$$\int_{-T}^{T} g_h(t)\,\mathrm{d}t = R_{h,T}(g) \quad (h>0) \qquad \text{and} \qquad \lim_{h\to 0} R_{h,T}(g) = \int_{-T}^{T} g(t)\,\mathrm{d}t.$$

(b) Suppose there is an $\alpha > 0$ such that $R_{h,T}(|g|) \leq \alpha$ for all $h > 0$ that are sufficiently small and all T that are suffciently large. Show that $\int_{-\infty}^{\infty} |g(t)|\ \mathrm{d}t < \alpha$ and $g \in L^1(\mathbb{R})$.

Can we apply Lebesgue's Convergence Theorem at this point to prove that the Riemann Sums converge to $\int g$ if, in addition, we also have that g and g_h are in $L^1(\mathbb{R})$?

(c) Assume that g is differentiable and that both g and g' are in $L^1(\mathbb{R})$. Prove that $\int_0^h |g(t) - g(0)|\ \mathrm{d}t \leq h \int_0^h |g'(t)|\ \mathrm{d}t$. Conclude that

$$|\int_{-T}^{T} g(t)\ \mathrm{d}t - R_{h,T}(g)| \leq \int_{-\infty}^{\infty} |g(t) - g_h(t)|\ \mathrm{d}t \leq h \int_{-\infty}^{\infty} |g'(t)|\ \mathrm{d}t = h\|g'\|_1.$$

In particular, this leads to

$$\int_{|t|>T} |g_h(t)|\ \mathrm{d}t \leq \int_{|t|>T} |g(t)|\ \mathrm{d}t + h\|g'\|. \tag{20}$$

(d) Prove that, if $g, g' \in L^1(\mathbb{R})$, then

$$\mathrm{Error}_{h,T} \equiv |\int_{-\infty}^{\infty} g(t)\ \mathrm{d}t - R_{h,T}(g)| \leq \int_{|t|>T} |g(t)|\ \mathrm{d}t + h\|g'\|_1.$$

Conclude that

$$|\int_{-\infty}^{\infty} g(t)\ \mathrm{d}t - R_{h,T}(g)| \to 0 \quad \text{for } h \to 0, T \to \infty,$$

where the limit is independent of the order (that is, for each $\varepsilon > 0$, there is a $h_0 > 0$ and a $T_0 > 0$ such that for all $h \in (0, h_0]$ and $T > T_0$ we have that $\mathrm{Error}_{h,T} \leq \varepsilon$).

(e) Assume that $g \in C^{(1)}(\mathbb{R})$ and both g and g' are in $L^2(\mathbb{R})$. Prove that

$$|\int_{-\infty}^{\infty} |g(t)|^2\ \mathrm{d}t - R_{h,T}(|g|^2)| \to 0 \quad \text{for } h \to 0, T \to \infty,$$

where the limit is independent of the order.

Exercise 1.13. For functions on a bounded interval $[a, b]$, there are $\kappa > 0$ and $\widetilde{\kappa} > 0$ (namely, $\kappa = \sqrt{b-a}$ and $\widetilde{\kappa} = b - a$) such that

$$\|f\|_1 \leq \kappa\, \|f\|_2 \leq \widetilde{\kappa}\, \|f\|_\infty \quad \text{for all } f.$$

Is this also correct for functions on $\mathbb{R}$ (with $\|\cdot\|_1$ and $\|\cdot\|_2$ defined as in §1.7)?

Exercise 1.14. Prove that, for $p = 1$ and for $p = 2$, we have that

$$\|fg\|_p \leq \|f\|_p\, \|g\|_\infty \quad \text{for all } f \in L^p([a, b]) \text{ and } g \in L^\infty([a, b]). \tag{21}$$

Exercise 1.15. Prove that $L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}) \subset L^2(\mathbb{R})$.

Exercise 1.16. Consider the function f on $[-1, +1]$ defined by

$$f(t) \equiv -1 \text{ if } t \leq 0 \quad \text{and} \quad f(t) \equiv 1 \text{ if } t > 0 \quad (t \in [-1, +1]).$$

(a) Let $\varepsilon > 0$.
Show, by explicit construction, that there is a $g \in C([-1, +1])$ such that $\|f - g\|_1 < \varepsilon$.
(b) Prove that $\|f - g\|_\infty \geq 1$ for each $g \in C([-1, +1])$.

Exercise 1.17. Let $g \in C^{(\ell)}(\mathbb{R})$ and $\varepsilon > 0$. Here, $\ell \in \mathbb{N}_0$ and $C^{(0)}([a, b]) = C([a, b])$. Suppose $\delta > 0$ is such that $|g(s) - g(t)| < \varepsilon$ as soon as $|t - s| < \delta$. Define

$$h(t) = \frac{1}{2\delta} \int_{t-\delta}^{t+\delta} g(s)\ \mathrm{d}s \qquad (t \in \mathbb{R}). \tag{22}$$

(a) Show that $h \in C^{(\ell+1)}(\mathbb{R})$. (Hint: $h'(t) = (g(t+\delta) - g(t-\delta))/(2\delta)$.)

(b) Prove that $\|h - g\|_\infty < \varepsilon$. (Hint: $h(t) - g(t) = \frac{1}{2\delta}\int_{t-\delta}^{t+\delta} g(s) - g(t)\ \mathrm{d}s$.)

(c) Let $g \in C([a,b])$, $k \in \mathbb{N}$, and $\varepsilon > 0$.
Prove that $\|g - h\|_\infty < \varepsilon$ for some $h \in C^{(k)}([a,b])$.

Exercise 1.18. A function F on $[a,b]$ is said to be *absolutely continuous* if it is the primitive of some $L^1([a,b])$ function f: $F(t) = \alpha + \int_c^t f(t)\ \mathrm{d}t$ for some $\alpha \in \mathbb{C}$ and $c \in [a,b]$.

(a) Suppose F is absolutely continuous with L^1-derivative f. Let $\varepsilon > 0$ and $f_c \in C([a,b])$ such that $\|f - f_c\|_1 < \varepsilon$. Define $F_c(t) \equiv \alpha + \int_c^t f_c(s)\ \mathrm{d}s$. Show that F_c is continuously differentiable and that $\|F - F_c\|_\infty \leq \varepsilon$.

(b) Prove that each absolutely continuous function F is continuous.

(c) Prove that each continuous piecewise continuously differentiable function F on $[a,b]$ is absolutely continuous (F is *piecewise continuously differentiable* if for some finite increasing sequence $(c_0 = a, c_1, \ldots, c_{n-1}, c_n = b)$, the restriction of F to (c_{i-1}, c_i) is differentiable with continuous and bounded derivative for each $i = 1, \ldots, n$).
In particular, each function in $C^{(1)}([a,b])$ is absolutely continuous.

(d) Consider the function

$$F(x) \equiv x \sin\frac{\pi}{x} \quad \text{for } x \in [-1,0), \quad \text{and} \quad F(0) \equiv 0. \tag{23}$$

Show that F is continuous on $[0,1]$ but not absolutely continuous.
(Hint: f, given by $f(x) \equiv \frac{\pi}{x}\cos\frac{\pi}{x}$ for $x \in [-1,0)$ and $f(0) \equiv 0$, is not in $L^1([-1,0])$.)

Exercise 1.19. Suppose that $F(t) = \alpha + \int_c^t f(s)\ \mathrm{d}s$ and $G(t) = \beta + \int_c^t g(s)\ \mathrm{d}s \quad (t \in [a,b])$ for some $f, g \in L^1([a,b])$. Here, $\alpha, \beta \in \mathbb{C}$ and $c \in [a,b]$.

(a) Prove (12) for the present F and G using the fact that (12) is correct for continuously differentiable functions F and G. (Hint: see Exercise 1.18(a).)

Exercise 1.20. Suppose that $F(t) = \alpha + \int_c^t f(s)\ \mathrm{d}s \quad (t \in \mathbb{R})$ for some $f \in L^1(\mathbb{R})$. Here, $\alpha \in \mathbb{C}$ and $c \in \mathbb{R}$.

(a) Show that F is uniformly continuous.

(b) *F vanishes at infinity* if $F(t) \to 0$ for $|t| \to \infty$.
Show that F vanishes at infinity if both F and F' $(= f)$ are in $L^1(\mathbb{R})$.

(c) Does $F' \in L^1(\mathbb{R})$ (i.e., $f \in L^1(\mathbb{R})$) imply that F vanishes at infinity?

(d) Does F vanishes at infinity if $F \in L^1(\mathbb{R})$, but $F' \notin L^1(\mathbb{R})$?

Exercise 1.21. Let $f \in L^1(\mathbb{R})$. Let $\varepsilon > 0$. Prove that there is a $\delta > 0$ such that

$$\int_{-\infty}^{\infty} |f(t) - f(t+s)|\ \mathrm{d}t < 2\varepsilon \quad \text{for all} \quad s, |s| < \delta. \tag{24}$$

(Hint: use that there is a *uniformly continuous* function g on $\mathbb{R}$ such that $\|f - g\|_1 < \varepsilon$, i.e., g is such that for each $\varepsilon' > 0$ there is a $\delta > 0$ such that $|g(t) - g(s)| < \varepsilon'$ whenever $|s - t| < \delta$.)

Exercise 1.22. The *total variation* of a function F on $[a,b]$ is the amount by which $F(x)$ (the graph of F along the vertical axis) varies if we move x from a to b.

(a) The total variance of $F(t) \equiv \sqrt{|t|}$ on $[-1,1]$ is 2 (why?), whereas the total variance on $[-1,0]$ of F in (23) is infinite (why?).

The formal definition of *total variance* $\mathrm{Var}(F)$ is as follows

$$\mathrm{Var}(F) \equiv \sup \sum_{k=1}^{n} |F(c_i) - F(c_{i-1})|$$

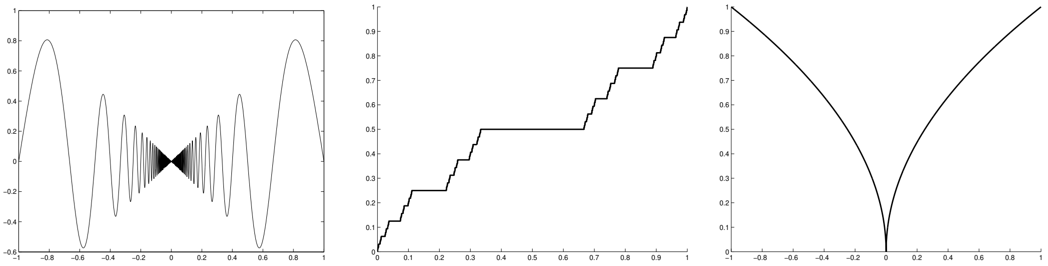


FIGURE 2. All functions of which the graph is shown here are continuous. The function at the left, $x \rightsquigarrow x\sin(\pi/x)$, does not have a bounded variation. The middle function, Cantor's stair, is of bounded variation, but its derivative is equal to 0 almost everywhere and the function is not absolutely continuous. The function at the right, $x \rightsquigarrow \sqrt{|x|}$, is absolutely continuous, but is not $C^{(1)}$.

where the supremum is taken over all $n \in \mathbb{N}$ and all increasing sequences $(c_0 = a, c_1, \ldots, c_{n-1}, c_n = b)$. F is of *bounded variation* (BV) if $\mathrm{Var}(F) < \infty$.

F is *non-decreasing* on $[a, b]$ if $-\infty < F(x) \leq F(y) < \infty$ for all $x, y \in [a, b]$, $x < y$. In particular, non-decreasing functions are real-valued.

(b) Show that non-decreasing functions are of BV. Show that any finite linear combination of non-decreasing functions is of BV. Actually, any function F that is of BV is of the form $F = F_1 - F_2 + i(F_3 - F_4)$ with F_j non-decreasing (but you do not have to show that). Conclude that F is of BV if F has at most finitely many local extrema.

(c) Prove that

$$\mathrm{Var}(F) = \|f\|_1$$

if F is absolutely continuous with 'derivative' $f \in L^1([a, b])$. In particular, we see that each absolutely continuous (AC) function F is of *bounded variation*, i.e., $\mathrm{Var}(F) < \infty$.

(d) In Exercise 1.23, we will construct a function F that is continuous, and of bounded variation, but that is not absolutely continuous. Use also this fact to conclude that

$$C^{(1)}([a, b]) \subset \{f \,\big|\, f \text{ is AC on } [a, b]\} \subset \{f \in C([a, b]) \,\big|\, f \text{ is of BV}\} \subset C([a, b])$$

and that all inclusions are strict (see Fig. 2).

Exercise 1.23. **Cantor's stair.**

Suppose that the function values $F(a)$ and $F(b)$ at the ends a and b of the interval $I = [a, b]$ are known. Then we obtain the values of F in the other points of $[a, b]$ as follows. We devide the interval $[a, b]$ in three subintervals, I_1, I_2, and I_3, of equal length, $I_1 \equiv a + (b - a)[0, \frac{1}{3}]$, $I_2 \equiv a + (b - a)[\frac{1}{3}, \frac{2}{3}]$, and $I_3 \equiv a + (b - a)[\frac{2}{3}, 1]$, and we define F on the middle interval by $F(x) \equiv \frac{1}{2}(F(a) + F(b))$ for all $x \in I_2$. Note that now F has been defined on both ends of both intervals I_1 and I_3. Therefore, the procedure can be repeated on these two remaining intervals, i.e., we take $I = I_1$ and then $I = I_3$, etc.. We start our construction on the interval $[0, 1]$ with the values $F(0) = 0$ and $F(1) = 1$. A graphical display of this function is given in Fig. 2.

(a) Write a MATLAB code of at most 25 commands that for any given resolution $h > 0$ along the horizontal axis produces the graph of F as in center picture of Fig. 2 (i.e., if $h > 0$ is given, then for each $y \in [0, 1]$, $F(x)$ is computed by the program for at least one $x \in [y, y + h]$).

(b) In the first step, F has *not* been defined on (two) intervals of total length $\frac{2}{3}$, in the second step, the total length of this area of 'undefinedness' has been reduced to $(\frac{2}{3})^2$, etc.. Check this. The set of points at which F has not been defined is negligible. Argue that in the remaining points of $[0, 1]$ F can be defined such that F is continous on $[0, 1]$.

Compute the total variation of F. Show that F' exists and equals to 0 almost everywhere. Conclude that F is not absolutely continuous.

2 Fourier series

The functions in this section are defined on $\mathbb{R}$, have complex values and are *locally* integrable (i.e. integrable on each bounded interval), unless stated otherwise.

Let $T > 0$.

2.1 Periodic functions. A function $f : \mathbb{R} \to \mathbb{C}$ is T-*periodic* (that is, periodic with *period* T) if

$$f(t+T) = f(t) \text{ for all } t \in \mathbb{R}.$$

If the function f is T-periodic, then

$$\int_0^T f(t)dt = \int_\tau^{T+\tau} f(t)\,\mathrm{d}t \quad \text{for all} \quad \tau \in \mathbb{R}. \tag{25}$$

The space of all T periodic functions f for which

$$\|f\|_1 \equiv \frac{1}{T}\int_0^T |f(t)|\,\mathrm{d}t < \infty$$

is denoted by $L^1_T(\mathbb{R})$; $L^2_T(\mathbb{R})$ is the space of T-periodic functions f for which

$$\|f\|_2 \equiv \sqrt{\frac{1}{T}\int_0^T |f(t)|^2\,\mathrm{d}t} < \infty.$$

2.2 Fourier series. Consider an $f \in L^1_T(\mathbb{R})$.
We define the *exponential Fourier coefficients*:

$$\gamma_k(f) \equiv \frac{1}{T}\int_0^T f(t)\exp(-2\pi i t\frac{k}{T})\,\mathrm{d}t \qquad (k \in \mathbb{Z}). \tag{26}$$

With

$$S_n(f)(t) \equiv \sum_{k=-n}^{n} \gamma_k(f)\exp(2\pi i t\frac{k}{T}) \qquad (t \in \mathbb{R}), \tag{27}$$

$S_n(f)$ is the nth *partial Fourier series* of f. The formal sum $\sum_{k\in\mathbb{Z}} \gamma_k(f)\exp(2\pi i t\frac{k}{T})$ is the *Fourier series* of f. We write

$$f \sim \sum_{k=-\infty}^{\infty} \gamma_k(f)\exp(2\pi i t\frac{k}{T}).$$

Fourier series can also be defined in terms of the *trigonometric Fourier coefficients*

$$\begin{cases} \alpha_k(f) \equiv \dfrac{2}{T}\displaystyle\int_0^T f(t)\cos(2\pi t\frac{k}{T})\,\mathrm{d}t \qquad (k \in \mathbb{N}_0), \\[2ex] \beta_k(f) \equiv \dfrac{2}{T}\displaystyle\int_0^T f(t)\sin(2\pi t\frac{k}{T})\,\mathrm{d}t \qquad (k \in \mathbb{N}). \end{cases} \tag{28}$$

Then, $2\gamma_k(f) = \alpha_k(f) - i\beta_k(f)$ and

$$S_n(f)(t) = \tfrac{1}{2}\alpha_0(f) + \sum_{k=1}^{n}\left(\alpha_k(f)\cos(2\pi t\frac{k}{T}) + \beta_k(f)\sin(2\pi t\frac{k}{T})\right) \qquad (t \in \mathbb{R}). \tag{29}$$

Here we used the fact that $\exp(i\phi) = \cos(\phi) + i\sin(\phi)$.

2.3 If f is *odd* (i.e., $f(t) = -\overline{f(-t)}$ for all t), and real-valued, then

$$f \sim \sum_k \beta_k(f) \sin(2\pi t \frac{k}{T}),$$

and if f is *even* (i.e., $f(t) = \overline{f(-t)}$ for all t), and real-valued, then

$$f \sim \frac{1}{2}\alpha_0(f) + \sum_k \alpha_k(f) \cos(2\pi t \frac{k}{T}).$$

Usually, only the real parts have a meaning in Physics. We could restrict ourselves to sines and cosines. However, mathematical manipulations with powers of e are much more convenient.

A popular choice for T in many textbooks is $T = 2\pi$. This choice simplifies the formulas $\exp(2\pi i t \frac{k}{T})$ to $\exp(itk)$. For ease of future reference, we do not follow this simplification.

The coefficients $\gamma_k(f)$ are often denoted by $\widehat{f}(k)$ and $\widehat{f}$ is called the Fourier transform of f. In the next section, we will see that the notation $\frac{1}{T}\widehat{f}(\frac{k}{T})$ would be more appropriate.

The Fourier series of f has been introduced as a formal expression. Of course, we would like to know whether it coincides with f: does $(S_n(f))$ converge to f? For many functions f it does. However, according to a construction by du Bois–Reymond (from 1872), there is a T-periodic continuous function f whose Fourier series is divergent at some point x (i.e., $(S_n(f)(x))$ is divergent; see Exercise 2.22). Convergence appears to depend on the 'smoothness' of f and on the way convergence is measured. The following theorem gives some conditions. For proofs, we refer to the literature. See also Exercise 1.9.

2.4 Theorem. *Let* $f \in L^1_T(\mathbb{R})$.

(a) $\lim_{n\to\infty} \|f - S_n(f)\|_2 = 0$ *if* $f \in L^2_T(\mathbb{R})$.

(b) *Let* $x \in \mathbb{R}$ *and* $\delta > 0$. *If both* $f(x+)$ *and* $f(x-)$ *exist,*[5] *and* f *is* $C^{(1)}$ *on* $[x-\delta, x)$ *and on* $(x, x+\delta]$ *with bounded derivative, then*

$$\lim_{n\to\infty} S_n(f)(x) = \tfrac{1}{2}[f(x+) + f(x-)].$$

(c) $\lim_{n\to\infty} \|f - S_n(f)\|_\infty = 0$ *if* $f \in C^{(1)}(\mathbb{R})$. □

If $(S_n(f)(t))$ converges to $f(t)$, then we may write $f(t) = \sum_{k=-\infty}^{\infty} \gamma_k(f) \exp(2\pi i t \frac{k}{T})$.

Convergence in the square integral sense (as in (a)) does not imply point-wise convergence. The du Boi–Reymond example shows that a Fourier series may diverge at some point even if f is continuous (even for continous functions there may be infitely many points of divergence). However, it can not diverge everywhere: a result by Carleson (from 1966) shows that set of points of divergence for a function f in $L^2_T(\mathbb{R})$ is negligible (has measure 0). It is crucial here that f is in $L^2_T(\mathbb{R})$: there are functions f in $L^1(\mathbb{R})$ for which the Fourier series diverges in almost all points (Kolmogorov, 1923).

[5] $f(x+)$ exists if the limit of $f(x+\varepsilon)$ for $\varepsilon \to 0$ with $\varepsilon > 0$ exists. Then $f(x+)$ is the limit value. Similarly, $f(x-) = \lim_{\varepsilon>0,\varepsilon\to 0} f(x-\varepsilon)$.

The Fourier series may diverge at x even if f is continuous. However, if f is continuous at x and its series converges at x, then it converges to the 'correct' value: either $(S_n(f)(x))$ converges to $f(x)$ or it diverges (Cantor, 1924; see also Exercise 2.13.)

The result in (b) (Jordan's test; [5, §10.1.1]) is remarkable: the condition only involves values of f in some (arbitrarily) small neighborhood of x, while each $\gamma_k(f)$ depends on all values of f. This is a general fact and is known *the Riemann localization principle*: the convergence of the Fourier series to $f(x)$ only depends on the values of f in a neighborhood of x.

The conditions in (b) and in (c) can be relaxed.
For instance, in (b), it is sufficient if both $f(x+)$ and $f(x-)$ exists and f is a linear combination of finitely many functions that are non-decreasing[6] on $[x-\delta, x+\delta]$ (f is of *bounded variation* on $[x-\delta, x+\delta]$, see Exercise 6.20. Note that no continuity is required). For more theory on point-wise convergence, see, e.g., [1] and [5, §10].
In Exercise 6.20 we will see that (cf. (c)) we have uniform convergence of the Fourier series as soon as the T-periodic function f is continous and a linear combination of finitely many non-decreasing functions on $(-\frac{1}{2}T, \frac{1}{2}T]$ (which is the case if, for instance, f is of the form $f(t) = a + \int_0^t g(s)\,\mathrm{d}s$ with $\int_0^T |g(s)|\,\mathrm{d}s < \infty$; cf., Exercise 1.18), but, as we already know, only requiring continuity for f is not sufficient for (c).

Averaging the Fourier series (*Cesàro sums*) turns the Fourier series of any T-periodic continous function f into a uniformly converging one (see Exercise 6.19):

$$\lim_{n\to\infty} \|\tfrac{1}{n+1}\left(\sum_{j=0}^{n} S_j(f)\right) - f\|_\infty = 0 \quad \text{if} \quad f \in L^1_T(\mathbb{R}) \cap C(\mathbb{R}). \tag{30}$$

This result is elegant, because the convers is true as well: if the (average of the) Fourier series of f converges uniformly then f is continuous (see Exercise 2.24). In addition, it has interesting theoretical mathematical consequences (see Exercise 2.23). However, averaging turns fastly converging Fourier series into slowly converging ones (see (d) of Exercise 2.23) and its practical interest is therefore limited.

We will write

$$f = \sum_{k=-\infty}^{\infty} \gamma_k(f) \exp(2\pi i t \frac{k}{T}) \quad \text{for functions } f \in L^2_T(\mathbb{R}), \tag{31}$$

but one should keep in mind that the meaning of this expression depends on the way convergence is measured ($\sum_{k=-\infty}^{\infty}\ldots$ assumes a limit, $\lim_{n\to\infty}\sum_{|k|\leq n}\ldots$!).

In practice, it is useful to have an estimate for the error in an approximation by $S_n(f)$. The theorem does not provide this information. Nevertheless, the theorem is important. It plays an essential role in theoretical arguments; the proof of Parseval's formula in 2.8 forms an example.

Formula (31) tells us that functions f in $L^2_T(\mathbb{R})$ can be viewed as a *superposition* of *harmonic oscillations* $t \rightsquigarrow \gamma_k(f)e^{2\pi i t\frac{k}{T}}$ with *frequency* $\frac{k}{T}$.[7] The contribution of the oscillation with frequence $\frac{k}{T}$ has *amplitude* $|\gamma_k(f)|$. If $\gamma_k(f) = |\gamma_k(f)|e^{-i\phi_k}$ with ϕ_k in $[0, 2\pi)$, then ϕ_k is the *phase* of f at frequency $\frac{k}{T}$.[8]

[6] A function f is *non-decreasing* on $[a,b]$ if $-\infty < f(x) \leq f(y) < \infty$ for all $x, y \in [a,b]$, $x<y$.

[7] If t is a point of time and $\mathbb{R}$ represents time, then $t \rightsquigarrow e^{2\pi i t\omega}$ runs through k periods in T seconds, that is, on average, $\omega \equiv \frac{k}{T}$ periods per second. The quantity ω is called the frequency then and is measured in Hz (*Hertz*), i.e., oscillations per second. We will use this terminology. Often the term 2π is absorbed in ω, i.e., $t \rightsquigarrow e^{it\omega}$ is considered. Then ω is called the *angular frequency*.

[8] To understand why the expression 'phase' is used, note that $\gamma_k \exp(2\pi i t\frac{k}{T}) = |\gamma_k| \exp(2\pi i t\frac{k}{T} - \phi_k)$: the harmonic oscillation with frequency $\frac{k}{T}$ is shifted by ϕ_k.

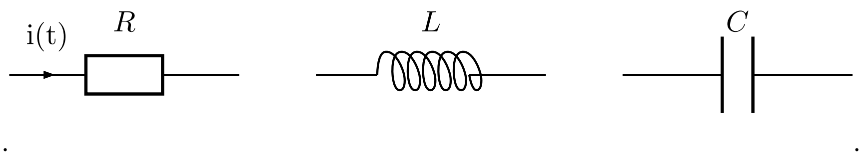


FIGURE 3. Elementary building blocks in an electrical network. From left to right, resistor, inductor (coil), capacitor.

The next theorem states that there is a very simple relation between Fourier coefficients of a function and its derivative. This theorem allows to translate 'complicated' problems from Analysis to 'simple' algebraic problems: differentiating is translated to multiplication. This property explains to some extent the popularity of Fourier transforms in technical sciences.

2.5 Theorem. *If f is differentiable and $f' \in L^1_T(\mathbb{R})$ then*

$$\gamma_k(f') = 2\pi i \tfrac{k}{T}\, \gamma_k(f) \quad \textit{for all} \quad k \in \mathbb{Z}.$$

Proof. Integrate by parts. The fact that both f and $t \rightsquigarrow \exp(2\pi i t \frac{k}{T})$ are T-periodic implies that the 'stock term' is 0. □

In the next application, we demonstrate how Fourier transformations can simplify the analysis of an electric current I in a circuit induced by a known varying electric potential V.

2.6 Application. At time t, let $V_C(t)$ be the electric potential at the two ends of a capacitor with capacitance C and let I_C be the electric current induced by this voltage difference. Then $\frac{d}{dt}V_C(t) = \frac{1}{C}I_C(t)$ for all t. If the voltage drop V_C is T-periodic and $Z_k(C) \equiv 1/(2\pi i \frac{k}{T} C)$, then $\gamma_k(V_C) = Z_k(C)\gamma_k(I_C)$.

If $V_L(t)$ is the potential at two ends of an inductor with inductance L and $I_L(t)$ is the induced current, then $V_L(t) = L\frac{d}{dt}I_L(t)$. If V_L T-periodic and $Z_k(L) \equiv 2\pi i \frac{k}{T} L$, then $\gamma_k(V_L) = Z_k(L)\gamma_k(I_L)$.

If $V_R(t)$ is the voltage drop at the two ends of a resistor with resistance R and $I_R(t)$ is the induced current, then $V_R = RI_R$ (Ohm's law): if V_R T-periodic and $Z_k(R) \equiv R$, then $\gamma_k(V_R) = Z_k(R)\gamma_k(I_R)$.

Now, it is easy to analyze an electric network that is constructed from capacitors, inductors, and resistors only. As an example we consider the following circuit.

We put a T-periodic potential V at two points that are connected with each other by a conducting wire via a capacitor, an inductor and a resistor (see Fig. 4). According to *Kirchhoff's laws*[9] we have that $V = V_C + V_L + V_R$ and $I \equiv I_C = I_L = I_R$.
These relations imply that $\frac{1}{C}I(t) + LI''(t) + RI'(t) = V'(t)$. We could employ analytical tools to solve this differential equation for I (assuming that I is periodic). The solution can much easier be obtained with a Fourier transform: $(Z_k(C)+Z_k(L)+Z_k(R))\gamma_k(I) = \gamma_k(V)$. With $Z_k \equiv Z_k(C) + Z_k(L) + Z_k(R)$ we obtain the Fourier coefficients of I from the ones of V: $\gamma_k(I) = \frac{1}{Z_k}\gamma_k(V)$.

[9]Kirchhoff's voltage law: the algebraic sum of all instantaneous voltage drops around any closed loop in a circuit is zero. Kirchhoff's current law: at any point of a circuit, the sum of the inflowing currents is equal to the sum of the outflowing currents.

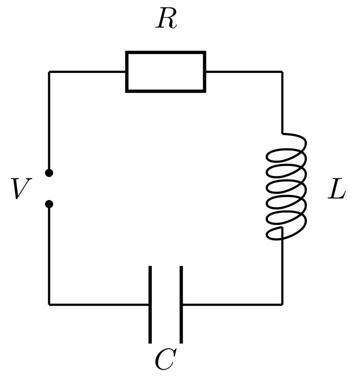


FIGURE 4. A simple electronic circuit.

2.7 Orthonormality. Define

$$(f,g)=\frac{1}{T}\int_0^T f(t)\cdot\overline{g(t)}\,\mathrm{d}t,\quad \|f\|_2=\sqrt{(f,f)}\quad (f,g\in L_T^2(\mathbb{R})):$$

$(\,\cdot\,,\cdot\,)$ defines an inner product with associated norm $\|\cdot\|_2$ on the space $L_T^2(\mathbb{R})$.

The (ϕ_k), with

$$\phi_k(t)\equiv\exp(2\pi i t\frac{k}{T})=(\phi_1(t))^k \qquad (t\in\mathbb{R},k\in\mathbb{Z}),$$

form an *orthogonal system*, i.e., $(\phi_k,\phi_\ell)=0$ for all $k,\ell\in\mathbb{Z},k\neq\ell$. Moreover, the ϕ_k are *normalized*, i.e., $\|\phi_k\|_2=1\quad(k\in\mathbb{Z})$: the system (ϕ_k) is said to be *orthonormal*.

Note that orthogonality (orthonormality) depends on the inner product. In the following sections, we will introduce other inner products for which, for simplicity of notation, we will use the same symbols. In particular, the dependency on T does not show in our notation.

Now, for $f\in L_T^2(\mathbb{R})$, the Fourier series can be represented as

$$\gamma_k=(f,\phi_k)\quad(k\in\mathbb{Z}),\qquad S_n(f)=\sum_{|k|\leq n}(f,\phi_k)\phi_k\quad(n\in\mathbb{N}) \tag{32}$$

and

$$f=\sum_{k=-\infty}^{\infty}(f,\phi_k)\phi_k, \tag{33}$$

where (33) has to be understood in the L_T^2-sense: $\|f-\sum_{|k|<n}(f,\phi_k)\phi_k\|_2\to 0$ for $n\to\infty$.

The representation in (33) is unique for functions $f\in L_T^2(\mathbb{R})$, that is, if $f=\sum_{k=-\infty}^{\infty}\widetilde{\gamma}_k\phi_k$ in the L_T^2-sense, then $\widetilde{\gamma}_k=\gamma_k(f)=(f,\phi_k)$ for all $k\in\mathbb{Z}$ (see Exercise 2.11). The (ϕ_k) form a so-called *Schauder* basis for $L_T^2(\mathbb{R})$.

The representation of $S_n(f)$ and of f as a linear combination of functions that are mutually orthogonal leads to an 'infinitely dimensional' variant of Pythagoras' theorem which is called the *Parseval's formula* (or also *Bessel's equality*) in Fourier theory.

2.8 Parseval's formula.

$$\sum_{k=-\infty}^{\infty} |\gamma_k(f)|^2 = \frac{1}{T}\int_0^T |f(t)|^2 \,\mathrm{d}t \qquad (f \in L_T^2(\mathbb{R})). \tag{34}$$

Proof. 1) We have that $f - S_n(f) \perp \phi_k$ $(|k| \leq n)$. This implies that $f - S_n(f) \perp S_n(f)$, since $S_n(f)$ is a finite linear combination of multiples of ϕ_k $(|k| \leq n)$.
2) Now, Pythagoras' theorem leads to $\|f\|_2^2 = \|f - S_n(f)\|_2^2 + \|S_n(f)\|_2^2$.
3) Since the ϕ_k are orthonormal, we can repeatedly apply Pythagoras' theorem to find that $\|S_n(f)\|_2^2 = \sum_{|k|\leq n} |\gamma_k(f)|^2$.
4) Finally, a combination of the results in 2) and 3) with a limit process using (a) of Theorem 2.4 proves Parseval's formula. □

Replacing f by $f + \zeta g$ in (34) and varying the scalar ζ, leads to the following variant of Parseval's formula

$$\sum_{k=-\infty}^{\infty} \gamma_k(f)\,\overline{\gamma_k(g)} = \frac{1}{T}\int_0^T f(t)\,\overline{g(t)}\,\mathrm{d}t \qquad (f, g \in L_T^2(\mathbb{R})). \tag{35}$$

2.9 According to Parseval's formula, the Fourier coefficients $(\gamma_k(f))$ of functions f in $L_T^2(\mathbb{R})$ are absolutely quadratically summable: $\sum |\gamma_k(f)|^2 < \infty$. Conversely, if a sequence $(\mu_k)_{k\in\mathbb{Z}}$ of scalars is absolutely quadratically summable, then it is a sequence of Fourier coefficients of some function in $L_T^2(\mathbb{R})$: to be more precise, $g \equiv \sum \mu_k\,\phi_k$ is in $L_T^2(\mathbb{R})$ and $\mu_k = \gamma_k(g)$ $(k \in \mathbb{Z})$. This result is the *Riesz–Fischer theorem.* Apparently, the Fourier transform $f \rightsquigarrow (\gamma_k(f))$ identifies the space $L_T^2(\mathbb{R})$ with the space $\ell^2(\mathbb{Z}) \equiv \{(\mu_k)_{k\in\mathbb{Z}} \,\big|\, \sum |\mu_k|^2 < \infty\}$. If we equip $\ell^2(\mathbb{Z})$ with the inner product $<(\mu_k), (\nu_k)> \equiv \sum \mu_k\,\overline{\nu_k}$ with associated norm $\|(\mu_k)\|_2 \equiv \sqrt{\sum |\mu_k|^2}$, then the identification by the Fourier transform also preserves norm and inner product.

The transform $(\gamma_k)_{k\in\mathbb{Z}} \rightsquigarrow \sum_k \gamma_k\,\phi_k$ for sequences $(\gamma_k) \in \ell^2(\mathbb{Z})$ is called the *discrete Fourier transform.* Note that this transform depends on T.

The orthogonality property that we deduced in step 1) of the proof of Parseval's formula implies also that the nth partial Fourier series is the best approximation in some sense:

2.10 Proposition. *If $f \in L_T^2(\mathbb{R})$ and $\widetilde{S}_n = \sum_{|k|\leq n} \widetilde{\gamma}_k\,\phi_k$ for some scalars $\widetilde{\gamma}_k$, then*

$$\|f - S_n(f)\|_2 \leq \|f - \widetilde{S}_n\|_2.$$

Proof. Use Pythagoras and the fact that $f - S_n(f) \perp S_n(f) - \widetilde{S}_n$ (see the first step in the proof of Parseval's formula). □

2.11 An immediate consequence of Parseval's formula is that, for $f \in L_T^2(\mathbb{R})$,

$$\lim_{k\to\infty} \gamma_k(f) = 0. \tag{36}$$

Although there are T-periodic functions f for which $\|f\|_1 < \infty$, while $\|f\|_2 = \infty$, it can be shown that (36) holds as soon as $\|f\|_1 < \infty$ (this is the *Riemann-Lebesgue lemma*; see Exercise 2.15).

For smoother T-periodic functions f, the speed with which $\gamma_k(f)$ converges to 0 ($n \to \infty$) can be estimated.

First note that

$$|\gamma_k(f)| \leq \|f\|_1.$$

As compared to (36), this estimate is not very exciting. However, in combination with Theorem 2.5, it leads to interesting estimates: $|\gamma_k(f)| = \frac{T}{2\pi|k|}|\gamma_k(f')| \leq \frac{T}{2\pi|k|}\|f'\|_1$ if f is differentiable, or, more general: if f is i-times differentiable, then

$$|\gamma_k(f)| = \left|\frac{T}{2\pi k}\right|^i |\gamma_k(f^{(i)})| \leq \left|\frac{T}{2\pi k}\right|^i \|f^{(i)}\|_1.$$

This estimate can be used to obtain an upper bound on the error in $S_n(f)$ (see Exercise 2.16): for some positive constants κ_2 and κ_∞ we have that

$$\|f - S_n(f)\|_2 \leq \kappa_2\, n^{-i+\frac{1}{2}}\, \|f^{(i)}\|_1 \quad \text{and} \quad \|f - S_n(f)\|_\infty \leq \kappa_\infty\, n^{-i+1}\|f^{(i)}\|_1$$

Note that the speed of convergence (or, actually, an upper bound of the error) depends on the smoothness of f.

2.12 Application: the wave equation. Consider the *partial differential equation* (PDE)

$$\frac{\partial^2 u}{\partial t^2}(x,t) = c^2 \frac{\partial^2 u}{\partial x^2}(x,t) \quad \text{for} \quad t \geq 0,\, x \in [0,1] \tag{37}$$

with *boundary conditions* (BC)

$$u(0,t) = u(1,t) = 0 \quad (t \geq 0), \tag{38}$$

and *initial conditions* (IC)

$$u(x,0) = \phi_1(x) \quad \text{and} \quad \frac{\partial u}{\partial t}(x,0) = \phi_2(x) \qquad (x \in [0,1]). \tag{39}$$

The real-valued functions $\phi_1, \phi_2 \in L^2([0,1])$ are known, u is real-valued and has to be solved for. c is a given real constant. Equation (37) is the *wave equation*: $u(t,x)$ describes the height of a vibrating string at time t and position x. The (idealized) one dimensional string is stretched between 0 and 1. The string is fixed at 0 and 1, which explains the boundary values $u(t,0) = u(t,1) = 0$ ($t \geq 0$). ϕ_1 and ϕ_2 describe the form and 'speed' of the string at release at time $t = 0$.

To solve this problem, we firstly concentrate on the solution of (37) and (38).
We try to find a solution u of the form $u(x,t) = f(t)g(x)$. Note that it is not clear that such a solution exists, nor that it leads to a solution that also satisfies (39). Nevertheless, we will try and we will see that this approach is fruitful.
Substitution in (37) yields $\frac{f''}{f}(t) = c^2\frac{g''}{g}(x)$. Since this equation should hold for all $x \in [0,1]$ and for all $t \geq 0$, it can only be correct if

$$f''(t) = \lambda^2 c^2 f(t) \quad (t \geq 0) \quad \text{and} \quad g''(x) = \lambda^2 g(x) \quad (x \in [0,1]),$$

for some constant λ (we used λ^2 only to simplify notation below). Since (38) implies that g also should satisfy the conditions $g(0) = g(1) = 0$, we find that g is a multiple of $\sin \pi k x$ for some $k \in \mathbb{Z}$ and that $\lambda^2 = -k^2\pi^2$. Hence, f is a multiple of $\exp(\pi i c k t)$ and $f(t)g(x)$ is a multiple of $\exp(\pi i c k t)\sin(\pi k x)$. Note that for g it suffices to take only k

in $\mathbb{N}$. However, both $\lambda = ik\pi$ and $\lambda = -ik\pi$ are solutions of $\lambda^2 = -k^2\pi^2$. By letting k range over $\mathbb{Z}$, we capture both solutions. The PDE and the BC are linear. Therefore, linear combinations of the solution are also solutions. Hence, for any complex sequence $(a_k) \in \ell^2(\mathbb{Z})$,

$$u(x,t) = \sum_{k\in\mathbb{Z}} a_k\, e^{\pi ikct}\, \sin(\pi kx) \tag{40}$$

is also a solution of (37) and (38).

We will now use Fourier theory to see that an appropriate selection of the a_k leads to a u that also satisfies (39). Note that ϕ_1 and ϕ_2 can be expressed as a sine-series:

$$\phi_1(x) = \sum_{k=1}^{\infty} \alpha_k\, \sin(\pi kx) \quad \text{and} \quad \phi_2 = \sum_{k=1}^{\infty} \beta_k\, \sin(\pi kx) \tag{41}$$

with real coefficients α_k and β_k. To see this, consider the Fourier series of the odd 2-periodic function that coincides with ϕ_i on $(0,1)$ ($i = 1,2$; see also Exercise 2.7). If u is as in (40), then (39) implies that $a_k - a_{-k} = \alpha_k$ and $\pi ikc(a_k + a_{-k}) = \beta_k$ $(k \in \mathbb{N})$. Hence, $2a_k = \alpha_k - i\frac{\beta_k}{\pi kc}$ and $2a_{-k} = -\alpha_k - i\frac{\beta_k}{\pi kc} = -2\overline{a_k}$: the sequence (a_k) is odd. Since the α_k and β_k are real, we find that

$$u(x,t) = \sum_{k=1}^{\infty} \left(\alpha_k \cos(\pi kct) + \frac{\beta_k}{\pi kc}\sin(\pi kct)\right)\sin(\pi kx). \tag{42}$$

Note that, with $\gamma_k \equiv -ia_k$, (40) can be rewritten to

$$u(x,t) = \tfrac{1}{2}[\psi(ct+x) - \psi(ct-x)], \quad \text{where} \quad \psi(s) \equiv \sum_{k=-\infty}^{\infty} \gamma_k\, e^{\pi iks}. \tag{43}$$

The function ψ is 2-periodic in $L^2_1(\mathbb{R})$. Since $\gamma_{-k} = \overline{\gamma_k}$, we have that ψ is real-valued: $\psi(s) = 2\mathrm{Re}(\sum_{k=1}^{\infty} \gamma_k \exp(\pi iks))$. At fixed time t, the graphs of the function $x \rightsquigarrow \psi(ct+x)$ and of $x \rightsquigarrow \psi(ct-x)$ are mirror images, with mirror at $x = 0$. The function u is the difference of a wave that moves to the left and one that moves to the right with speed $-c$ and c, respectively.

Exercises

Exercise 2.1. Consider T-periodic functions f and g.

(a) Prove (25).

(b) Prove that $\|f\|_1 \leq \|f\|_2 \leq \|f\|_\infty$. Conclude that $C_T(\mathbb{R}) \subset L^2_T(\mathbb{R}) \subset L^1_T(\mathbb{R})$. Here, $C_T(\mathbb{R})$ is the space of all continuous T-periodic functions on $\mathbb{R}$.

(c) Find f and g such that $\|f\|_2 < \infty$, $\|f\|_\infty = \infty$ and $\|g\|_1 < \infty$, $\|g\|_2 = \infty$.

Exercise 2.2. Prove the following statements for $f \in L^1_T(\mathbb{R})$, $n \in \mathbb{N}, k \in \mathbb{Z}$.

(a) The $S_n(f)$ in (27) and in (29) coincide.

(b) $f \rightsquigarrow S_n(f)$ is linear: $S_n(\alpha f + \beta g) = \alpha S_n(f) + \beta S_n(g)$ $(\alpha, \beta \in \mathbb{C}, f, g \in L^1_T(\mathbb{R}))$.

(c) $S_n(f)$ is T-periodic and continuous.

(d) $|\gamma_k(f)| \leq \|f\|_1$.

(e) Sketch the graphs of $t \rightsquigarrow \cos(2\pi t\frac{k}{T})$ and of $t \rightsquigarrow \sin(2\pi t\frac{k}{T})$ for $k = 0, 1, 2$.

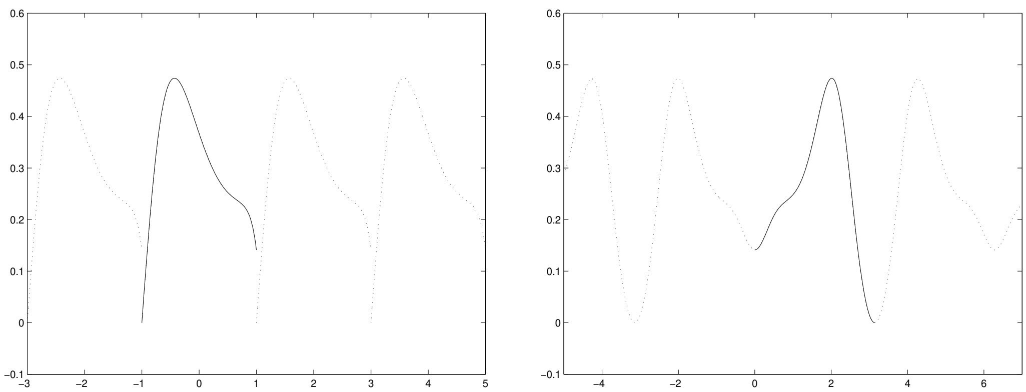


FIGURE 5. The left picture displays the graph of some f on $[-1,+1]$ (solid curve) and its 2-periodic associate f_* (the dotted curve). The picture at the right shows the graph of $t \rightsquigarrow f(\cos(t))$ for $t \in [0,\pi]$ (solid curve) and for $t \notin [0,\pi]$ (dotted curve). (See Exercise 2.6.)

Exercise 2.3. For a T-periodic function f and a real number a, the 'shifted' function f_a is defined by $f_a(t) \equiv f(t-a)$ $(t \in \mathbb{R})$. Show that
f_a is T-periodic, $\gamma_k(f_a) = \exp(-2\pi i a \frac{k}{T})\,\gamma_k(f)$ $(k \in \mathbb{Z})$, $S_n(f_a) = (S_n(f))_a$ $(n \in \mathbb{N})$.

Exercise 2.4.

(a) Show that $f \sim \frac{1}{2}\alpha_0 + \sum_{k=1}^{\infty} \alpha_k(f)\cos(2\pi t \frac{k}{T})$ if $f \in L^2_T(\mathbb{R})$ is even and real-valued.

(b) Show that $f \sim \sum_{k=1}^{\infty} \beta_k(f)\sin(2\pi t \frac{k}{T})$ if $f \in L^2_T(\mathbb{R})$ is odd and real-valued.

(c) Give a variant of the two properties above for the case where f is complex-valued.

(d) Show that each complex-valued function f on $\mathbb{R}$ can be written as $f = f_e + f_o$ with f_e even and f_o odd.

(e) Discuss the statements: 'f is even if and only if $\gamma_k(f)$ is real for all k' and 'f is real valued if and only if the sequence $(\gamma_k(f))_k$ is even, i.e., $\gamma_{-k}(f) = \overline{\gamma_k(f)}$'.

Exercise 2.5. Compute the Fourier coefficients of the following 2π-periodic functions f (i.e., $T = 2\pi$).

(a) $f(t) = 0$ for $t \in (-\pi, 0]$ and $f(t) = 1$ for $t \in (0, \pi]$ (block pulse).

(b) $f(t) = t$ for $t \in (-\pi, +\pi]$ (sawtooth function).

Exercise 2.6. **Periodic extensions.**

Consider a complex-valued function f on the interval $[-1,+1]$.

(a) Let f_* be defined by $f_*(t) \equiv f(t-2k)$ $\quad (t \in \mathbb{R})$, where $k \in \mathbb{Z}$ is such that $t - 2k \in (-1, 1]$ (see the left picture in Fig. 5). Show that f_* is 2-periodic. Is f_* continuous if f is continuous?

(b) For a complex-valued function g that is defined on the interval $[0,1]$, let h be defined on $[-1,+1]$ by $h(t) \equiv g(t)$, $h(-t) \equiv g(t)$ $\quad (t \in [0,1])$. Is h_* continuous if g is continuous? Is h_* continuously differentiable if that is the case for g?

(c) Show that $f_c(t) \equiv f(\cos(t))$, for $t \in \mathbb{R}$, 2π periodic (see the right picture in Fig. 5), and f is real-valued then f_c is even. For what $n \in \mathbb{N}_0$ is the following implication correct: if $f \in C^{(n)}([-1,+1])$ then $f_c \in C^{(n)}(\mathbb{R})$?

Exercise 2.7. **Sine series and cosine series.**

Let f be a real-valued function in $L^2([0,1])$. Convergence and equality below is in L^2-sense (integrating over $[0,1]$). Further, $\mathbb{N}_0$ and $\mathbb{N}$ are the collection of non-negative integers $(k \geq 0)$ and of positive integers $(k \geq 1)$, respectively.

(a) Show that there is a 1-periodic function F in $L^2_1(\mathbb{R})$ such that $F(t)=f(t)$ for all $t\in[0,1)$ and conclude that there are real sequences $(\alpha_k)\in\ell^2(\mathbb{N}_0)$, $(\beta_k)\in\ell^2(\mathbb{N})$ such that

$$f(t)=\tfrac{1}{2}\alpha_0+\sum_{k=1}^{\infty}\alpha_k\cos(2\pi tk)+\sum_{k=1}^{\infty}\beta_k\sin(2\pi tk). \tag{44}$$

(b) Show that there is an even 2-periodic function F in $L^2_2(\mathbb{R})$ such that $F(t)=f(t)$ for all $t\in(0,1)$ and conclude that there is a real sequence $(\alpha_k)\in\ell^2(\mathbb{N}_0)$ such that

$$f(t)=\tfrac{1}{2}\alpha_0+\sum_{k=1}^{\infty}\alpha_k\cos(\pi tk). \tag{45}$$

(c) Show that there is an odd 2-periodic function F in $L^2_2(\mathbb{R})$ such that $F(t)=f(t)$ for all $t\in(0,1)$ and conclude that there is a real sequence $(\beta_k)\in\ell^2(\mathbb{N})$ such that

$$f(t)=\sum_{k=1}^{\infty}\beta_k\sin(\pi tk). \tag{46}$$

Note that $\sin(\pi tk)$ is 0 for $t=0$ and $t=1$ $(k\in\mathbb{N})$. If f is smooth, but $f(0)\neq 0$ or $f(1)\neq 0$, then F is not smooth at $t=0$ or at $t=1$, and the sine-series will converge slowly.

(d) Show that there is a 4-periodic function F in $L^2_2(\mathbb{R})$ that is odd around $t=0$ ($F(t)=-F(-t)$), even around $t=1$ ($F(1+t)=F(1-t)$) and that coincides with f on $(0,1)$. Conclude that there is a real sequence $(\beta_k)\in\ell^2(\mathbb{N})$ such that

$$f(t)=\sum_{k=1}^{\infty}\beta_k\sin(\pi t(k+\tfrac{1}{2})). \tag{47}$$

Note that we can have fast convergence only if f is smooth, $f(0)=0$ and $f'(1)=0$ (why?).

Exercise 2.8. **Fourier–Chebyshev series.**

Let $f\in C([-1,+1])$ be real-valued, and let f_c be as defined in c) of Exercise 2.6.

(a) Show that $S_n(f_c)$ is of the form $S_n(f_c)(t)=\frac{1}{2}\alpha_0+\sum_{k=1}^{n}\alpha_k\cos(kt)$.

Consider the functions

$$T_k(x)\equiv\cos(k(\arccos(x))\quad\text{for }\ x\in[-1,+1]\quad(k\in\mathbb{N}_0)$$

and, with α_j as in (a),

$$C_n(f)(x)\equiv\tfrac{1}{2}\alpha_0+\sum_{k=1}^{n}\alpha_kT_k(x)\qquad(n\in\mathbb{N}_0,\ x\in[-1,+1]).$$

(b) Show that

$$\begin{cases} T_0(x)=1, T_1(x)=x, & (x\in[-1,+1])\\ T_{k+1}(x)=2xT_k(x)-T_{k-1}(x) & (x\in[-1,+1],\ k\in\mathbb{N})\end{cases}$$

(Hint: $\cos(\psi+\phi)=2\cos(\phi)\cos(\psi)-\cos(\psi-\phi)$), conclude that T_k is a polynomial of exact degree k and that $C_n(f)$ is a polynomial of degree $\leq n$: the T_k's are the so-called *Chebyshev polynomials* and $C_n(f)$ is the nth partial *Fourier–Chebyshev* series.

(c) Prove that the sequence of polynomials $C_n(f)$ converges uniformly on $[-1,+1]$ to f if $f\in C^{(1)}([-1,+1])$.

Exercise 2.9. Prove that $S_n(f)'=S_n(f')$ if f is differentiable and $f'\in L^1_T(\mathbb{R})$.

Exercise 2.10. Consider the inner product and the functions ϕ_k defined in §2.7.

(a) Show that the ϕ_k are in the space $L^2_T(\mathbb{R})$.

(b) Show that the system (ϕ_k) in §2.7 is orthonormal with respect to the inner product defined there.

(c) Prove (32).

Exercise 2.11.
Consider a sequence $(\mu_k)_{k\in\mathbb{Z}}$ of scalars. Let $g_n \equiv \sum_{|k|<n} \mu_k\,\phi_k \quad (n \in \mathbb{N})$.

(a) Show that (g_n) is a Cauchy sequence in $L^2_T(\mathbb{R})$ w.r.t. $\|\cdot\|_2$ if (μ_k) is in $\ell^2(\mathbb{Z})$.

(b) Suppose g is the limit in L^2_T sense of the sequence (g_n): $\lim_{n\to\infty} \|g - g_n\|_2 = 0$. Show that $(g, \phi_k) = \mu_k$. In particular, $\mu_k = 0 \;\; (k \in \mathbb{Z})$ if $g = 0$.

(c) Show that $\widetilde{\gamma}_k = (f, \phi_k) \;\; (k \in \mathbb{Z})$ if $f \in L^2_T(\mathbb{R})$ and $f = \sum_{k=-\infty}^{\infty} \widetilde{\gamma}_k\,\phi_k$ in L^2_T sense.

Exercise 2.12. Formulate Parseval's formula in terms of the α_k and β_k.

Exercise 2.13. Suppose $f \in L^2_T(\mathbb{R})$ is continuous and the sequence $(S_n(f))$ converges uniformly. Let g be the limit function of $(S_n(f))$.

(a) Prove that g is T-periodic and continuous.

(b) Prove that $f = g$.
(Hint: Use (a) of Theorem 2.4 and (a) of Exercise 2.1 to show that $\|f - g\|_2 = 0$.)

Exercise 2.14. Often a function is described by a table of function values. This type of description can be found in old table books, but also on new media like CDs and DVDs. For instance, on CDs, a function f that represents the sound (air pressure function represented by an electrical voltage function) is 'sampled' at a rate of 41 100 Hertz, that is, for $\Delta t \equiv \frac{1}{44\,100}$ second, the function values $f(k\Delta t)$ are stored: the $f(k\Delta t)$ are called 'sample values', $1/\Delta t$ is the *sample frequency*. Put $\Omega \equiv 1/(2\Delta t)$.

To analyze the function $k \rightsquigarrow f(k\Delta t) \;\; (k \in \mathbb{Z})$, Fourier transforms are employed. If $(f(k\Delta t))$ is in $\ell^2(\mathbb{Z})$, then the 'discrete' transform as mentioned in §2.9 can be used.

Define

$$F(\omega) \equiv \sum_{k=-\infty}^{\infty} \Delta t\, f(k\Delta t)\, e^{-2\pi i \omega k \Delta t} \quad (\omega \in \mathbb{R}).$$

(a) Show that $F \in L^2_{2\Omega}(\mathbb{R})$.

(b) Prove that $F \in C(\mathbb{R})$ if $(f(k\Delta t))$ is in $\ell^1(\mathbb{Z})$, where $\ell^1(\mathbb{Z})$ is the space of absolute summable sequences (μ_k): $|\!|\!|(\mu_k)|\!|\!|_1 \equiv \sum |\mu_k|$.

(c) Prove that $f(k\Delta t) = \int_{-\Omega}^{\Omega} F(\omega) e^{2\pi i \omega k \Delta t}\, \mathrm{d}\omega \quad (k \in \mathbb{Z})$.

Exercise 2.15. The T-periodic function f is such that $\|f\|_1 < \infty$.

(a) Prove that $|\gamma_k(f)| \leq \|f - g\|_1 + |\gamma_k(g)|$ for each T periodic function g.

(b) Prove the Riemann–Lebesgue lemma: $\lim_{k\to\infty} \gamma_k(f) = 0$.
(Hint: use the fact that there are smooth functions g for which $\|f - g\|_1 < \varepsilon$).

Exercise 2.16. Suppose $f \in L^1_T(\mathbb{R})$ is i-times differentiable, $i \geq 1$.

(a) Prove that $\|f - S_n(f)\|_2^2 = \sum_{|k|>n} |\gamma_k(f)|^2 \leq \sum_{|k|>n} |\frac{T}{2\pi k}|^{2i}\, \|f^{(i)}\|_1^2$.

(b) Prove that $\|f - S_n(f)\|_2 \leq \kappa\, \|f^{(i)}\|_1\, n^{-i+\frac{1}{2}}$, where $\kappa \equiv \sqrt{2}\,(T/(2\pi))^i$.

(c) Derive an upper bound for $\|f - S_n(f)\|_\infty$ in case $i > 1$.

(d) Let f be the 2-periodic function defined by $f(t) = |t| \;\; (t \in [-1, +1])$. Give an upper bound for $\|f - S_n(f)\|_2$.

Exercise 2.17. **The heat equation.**

Use Fourier theory to solve the one dimensional *heat equation*

$$\frac{\partial u}{\partial t}(x,t) = \gamma \frac{\partial^2 u}{\partial x^2}(x,t) \quad \text{for} \quad t \geq 0,\, x \in [0,1] \tag{48}$$

with boundary conditions (BC)

$$u(0,t) = u(1,t) = 0 \quad (t \geq 0), \tag{49}$$

and initial conditions (IC)

$$u(x,0) = \phi(x) \qquad (x \in [0,1]). \tag{50}$$

The real-valued function $\phi \in L^2([0,1])$ is known, u is real-valued and has to be solved for. γ is a given positive constant, $u(t,x)$ describes the temperature distribution at time t and position x in a (idealized) one dimensional string that is stretching between 0 and 1. The temperature of the string is fixed to 0 at 0 and 1, ϕ is the temperature of the string at initial time $t = 0$.

(a) Prove that multiples of $\exp(-(\pi k)^2\gamma t)\,\sin(\pi k x)$ solve (48) and (49) $(k \in \mathbb{N})$.

(b) Find an expression for the solution u of (48), (49), and (50) in terms of linear combinations of $\sin(\pi k x)$.

(c) Show that, for each $t > 0$, the function $x \rightsquigarrow u(x,t)$ is in $C^{(\infty)}([0,1])$.

NOTE. *Equations like the heat equation (and the wave equation) were the inspiration source for the development of Fourier theory. The result in (a) was known in the mid 18th century and it was also realized that linear combinations of solutions form a solution. However, it was not clear what kind of functions could be formed at $t = 0$ with linear combinations of* $\sin(\pi k x)$ *(even the concept of 'function' was not clear at that time). Although, Fourier published a book in 1822 with many particular instances of representations of functions in trigonometric series, it was Dirichlet who started the rigorous study of Fourier series in 1829 and who introduced a concept of 'function' in 1837.*

Exercise 2.18. **The wave equation 2.**

Use Fourier theory to solve the one-dimensional wave equation (see §2.12), but now with boundary conditions given by

$$u(0,t) = 0 \quad \text{and} \quad \frac{\partial u}{\partial x}(1,t) = 0 \qquad (t \geq 0). \tag{51}$$

(Hint: use (d) of Exercise 2.7).

Exercise 2.19. **The wave equation 3.**

We use the notations and results from §2.12. Consider (40) and note that, for fixed t, the function $w(x) \equiv u(x,t)$ is defined for all $x \in \mathbb{R}$.

(a) Prove that w is real-valued, 2-periodic, in $L^2_2(\mathbb{R})$, *odd around* 0 (i.e., $w(-x) = -w(x)$ $(x \in \mathbb{R})$), and *odd around* 1 (i.e., $w(1-x) = -w(1+x)$ $(x \in \mathbb{R})$).

(b) The value at a specific point x is not wel defined for L^2-functions. However, w is 0 at 0 and at 1 also in a L^2-sense: show that $\frac{1}{2\delta}\int_{-\delta}^{\delta} w(x)\,\mathrm{d}x = 0$ and $\frac{1}{2\delta}\int_{1-\delta}^{1+\delta} w(x)\,\mathrm{d}x = 0$.

Note that the Fourier series also allows to define the concept of derivative in some weak sense.

Exercise 2.20. **Dirichlet kernel.**

For $n \in \mathbb{N}$ consider the so-called *Dirichlet kernel*, the function D_n, given by

$$D_n(t) \equiv \sum_{k=-n}^{n} \exp(itk). \tag{52}$$

(a) Show that D_n is even, 2π-periodic, and

$$\frac{1}{2\pi}\int_0^{2\pi} D_n(t)\ \mathrm{d}t = 1. \tag{53}$$

(b) Show that, with $\zeta \equiv \exp(it)$,

$$D_n(t) = 1 + 2\sum_{k=1}^{n}\cos(tk) = 1 + 2\,\mathrm{Re}\left(\sum_{k=1}^{n}\zeta^k\right) = 1 + 2\,\mathrm{Re}\left(\zeta\frac{\zeta^n - 1}{\zeta - 1}\right) = \frac{\sin(t(n+\frac{1}{2}))}{\sin(\frac{1}{2}t)}. \tag{54}$$

(Hint: Use that $\frac{2i\zeta}{\zeta-1} = \zeta^{\frac{1}{2}}/\sin(\frac{1}{2}t)$. Why is this correct?)

(c) Show that

$$1 + \tfrac{2}{\pi}\log(n+\tfrac{1}{2}) \le \|D_n\|_1 = \frac{1}{\pi}\int_0^{\frac{1}{2}\pi} |D_n(t)|\ \mathrm{d}t \le 2 + \log(n+\tfrac{1}{2}). \tag{55}$$

Here $\|\cdot\|_1$ is the 1-norm for 2π-periodic functions.
(Hint: split the integral in two parts, integrate in the first part from 0 to the first positive zero $\frac{2\pi}{2n+1}$ of D_n, integrate in the second part from the first zero to $\frac{1}{2}\pi$ and use $\frac{t}{\pi} \le \sin(\frac{1}{2}t) \le \frac{1}{2}t$.)

(d) Put $\mathcal{D}_n(t) \equiv \int_0^t D_n(s)\ \mathrm{d}s$ for $|t| \le \pi$. Show that

$$\mathcal{D}_n(t) > 0 \quad (t \in (0,\pi]) \quad \text{and} \quad \|\mathcal{D}_n\|_\infty = |\mathcal{D}_n(\tfrac{2\pi}{2n+1})| \le 2\pi. \tag{56}$$

(Hint: inspect the graph of D_n.)

Exercise 2.21. **Sawtooth.**
Let f be 2π-periodic given by $f(t) = t - \pi$ for $t \in [0, 2\pi)$.

(a) Is f continuous? Is f even or odd? Show that

$$S_n(f)(t) = \sum_{k=1}^{n}\frac{2}{k}\sin(tk). \tag{57}$$

(b) Discuss the convergence of $(S_n(f))$ (w.r.t. $\|\cdot\|_2$, $\|\cdot\|_\infty$ and point-wise).

(c) Show that

$$\frac{\pi^2}{6} = \sum_{k=1}^{\infty}\frac{1}{k^2}. \tag{58}$$

(Hint: Use Parseval's formula.)

(d) Prove that

$$\|S_n(f)\|_\infty \le 2\pi. \tag{59}$$

(Hint: Show that, with $g \equiv S_n(f)$, $g' = D_n - 1$ (see (54)). Now, use (56) observing that $-t \le g(t) \le \mathcal{D}_n(t)$ for $t \in [0,\pi]$.)

Exercise 2.22. **Fejèr example.**
We will construct a continuous function with Fourier series that does not converge at 0.

For ease of notation, we consider $T = 2\pi$-periodic functions.

(a) For $p, q \in \mathbb{N}$, $q < p$, define

$$f(t) \equiv f_{p,q}(t) \equiv 2\sin(tp)\sum_{j=1}^{q}\frac{1}{j}\sin(tj). \tag{60}$$

Prove that

$$\alpha_{p+j}(f) = -\frac{1}{j} \quad \text{for} \quad 0 < |j| \le q \quad \text{and} \quad \alpha_{p+j}(f) = 0 \quad \text{elsewhere}. \tag{61}$$

(Hint: $2\sin(\phi)\sin(\phi) = \cos(\phi-\psi) - \cos(\phi+\psi)$.)

Show that

$$|S_{p+q}(f)(0) - S_p(f)(0)| = |\sum_{j=1}^{q} \alpha_{p+j}(f)| > \log(q). \tag{62}$$

Show that

$$\|f\|_\infty < 4\pi. \tag{63}$$

(Hint: use (57) and (59).)

(b) Now, select sequences $(p_k), (q_k)$ in $\mathbb{N}$ such that

$$p_k + q_k < p_{k+1} - q_{k+1} \quad \text{for all} \quad k \in \mathbb{N}, \qquad \frac{1}{k^2}\log q_k \to \infty \quad (k \to \infty) \tag{64}$$

(for instance, check that with $p_k \equiv 2q_k$ and $q_k \equiv e^{k^3}$, we have that $3q_k < q_{k+1}$ and $\log q_k = k^3$): the Fourier coefficients $\alpha_n(f_{p_k,q_k})$ do not 'overlap'. Put

$$F(t) = \sum_{k=1}^{\infty} \frac{1}{k^2} f_{p_k,q_k}. \tag{65}$$

Show that F T-periodic and continous.
Show that $|S_{p_k+q_k}(F)(0) - S_{p_k}(F)(0)| > \frac{1}{k^2}\log(q_k) \to \infty \quad (k \to \infty)$.
Conclude that $(S_n(F)(0))$ does not converge.

Exercise 2.23. **Cesàro sums and Weierstrass' theorem.**

Convergence can be 'improved' by using Cesàro sums.
If $s_n = \sum_{|k|\leq n} \gamma_n$, then the average $\sigma_n \equiv \frac{1}{n+1}\sum_{j=0}^{n} s_n$ is the *Cesàro sum*.

(a) Discuss the convergence behaviour of (s_n) and of (σ_n) in case $\gamma_k \equiv (-1)^k$ for $k \geq 0$ and $\gamma_k \equiv 0$ for $k < 0$.

For $f \in L^1_T(\mathbb{R})$, define

$$\sigma_n(f) \equiv \frac{1}{n+1}\sum_{j=0}^{n} S_j(f) \qquad (n \in \mathbb{N}). \tag{66}$$

It can be shown (see Exercise 6.19) that

$$\lim_{n\to\infty} \|\sigma_n(f) - f\|_\infty = 0 \quad \text{if} \quad f \in C(\mathbb{R}) \cap L^1_T(\mathbb{R}). \tag{67}$$

(b) Use this result to prove *Weierstrass' theorem*:
for each $f \in C([-1,+1])$ and each $\varepsilon > 0$ there is a polynomial p such that

$$\|f - p\|_\infty < \varepsilon. \tag{68}$$

(Hint: consider $f_c(t) \equiv f(\cos(t))$ $(t \in \mathbb{R})$ and $p_n(x) \equiv \sigma_n(f_c)(\arccos(x))$ $(x \in [-1,+1])$, and use (b) of Exercise 2.8.)

(c) Use (67) to prove (a) of Theorem 2.4.
(Hint: for $\varepsilon > 0$ there is a continous T-periodic function g such that $\|f-g\|_2 < \frac{1}{2}\varepsilon$. Now, observe that $\|f - S_n(f)\|_2 \leq \|f - \sigma_n(g)\|_2 \leq \|f-g\|_2 + \|g - \sigma_n(g)\|_\infty$ (why?).)

The following example gives one reason why we do not always work with $\sigma_n(f)$ instead of $S_n(f)$.

(d) Consider the T-periodic function $f(t) \equiv \cos(2\pi\frac{t}{T})$ $(t \in \mathbb{R})$. Show that $S_n(f) = f$ for all $n \in \mathbb{N}$ and $S_0(f) = 0$. Conclude that $\sigma_n(f) = \frac{n}{n+1}f$ and

$$\|S_n(f) - f\|_\infty = 0, \quad \text{while} \quad \|\sigma_n(f) - f\|_\infty = \tfrac{1}{n+1} \quad \text{for all } n \in \mathbb{N}.$$

Exercise 2.24. f is in $L^1_T(\mathbb{R})$ in this exercise.

(a) Prove that

$$\lim_{n\to\infty} \|f - S_n(f)\|_2 = 0 \quad \Leftrightarrow \quad f \in L^2_T(\mathbb{R}).$$

(b) Prove that

$$\lim_{n\to\infty} \|f - \sigma_n(f)\|_\infty = 0 \quad \Leftrightarrow \quad f \in C(\mathbb{R}),$$

with $\sigma_n(f)f$ as defined in (66).

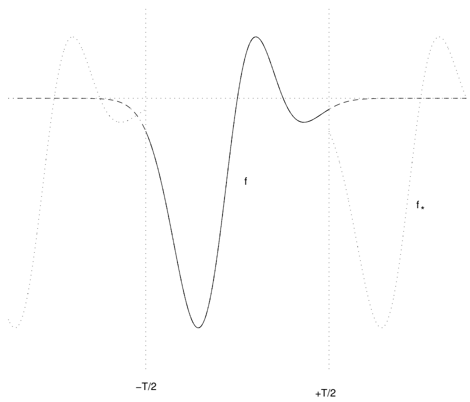


FIGURE 6. The picture displays the graph of some f. The graph of f restricted to $[-T/2, +T/2]$ is solid, the graph is dash-dotted elsewhere. The T-periodic extension of f restricted to $[-T/2, T/2]$ is dotted.

3 Fourier integrals

Functions as the V in the application above are never periodic in practice, not even approximately (think of the variance in voltage caused by a voice in a microphone).

We indicate how the theory in the preceding section can be formulated for non-periodic functions.

3.1 Fourier integrals, heuristics. Consider a function f for which

$$\|f\|_1 \equiv \int_{-\infty}^{+\infty} |f(t)| \ \mathrm{d}t < \infty.$$

If $f \in C^{(1)}(\mathbb{R})$, $T > 0$ and $\gamma_k(T) \equiv \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} f(t)\exp(-2\pi i t\frac{k}{T}) \ \mathrm{d}t$ then

$$f(t) = \sum_{k\in\mathbb{Z}} \gamma_k(T)\exp(2\pi i t\frac{k}{T}) \quad \text{for} \quad t \in (-\frac{T}{2}, \frac{T}{2})$$

(to see this, apply the theory from the preceding section to the function f_* that is T-periodic and for which $f_* = f$ on $[-\frac{T}{2}, +\frac{T}{2}]$; see Fig. 6).

In order to obtain an expression for f in terms of periodic 'exp-functions' that is correct for any $t \in \mathbb{R}$, it is tempting to drive T to ∞ in the above formulas. However, $\lim_{T\to\infty} \gamma_k(T) = 0$ for each k. If we select $\omega \in \mathbb{R}$ and converge both k and T to ∞ in such a way that $\frac{k}{T}$ converges to ω, then we see that $\lim_{T\to\infty} T\gamma_k(T)$ exists and is equal to $\int_{-\infty}^{+\infty} f(t)\exp(-2\pi i t\omega) \ \mathrm{d}t$.
If we note that, for $g \in C(\mathbb{R})$ for which $\int_{-\infty}^{+\infty} |g(\omega)| \ \mathrm{d}\omega$ exists,

$$\lim_{T\to\infty} \sum_{k\in\mathbb{Z}} \frac{1}{T} g(\frac{k}{T}) = \int_{-\infty}^{+\infty} g(\omega) \ \mathrm{d}\omega,$$

(Riemann sum) then we see that, with $\widehat{f}(\omega) \equiv \int_{-\infty}^{+\infty} f(t)\exp(-2\pi i t\omega) \ \mathrm{d}t \ \ (\omega \in \mathbb{R})$, we may expect that, for large T,

$$f(t) = \sum_{k\in\mathbb{Z}} \gamma_k(T)\exp(2\pi i t\frac{k}{T}) \approx \sum_{k\in\mathbb{Z}} \frac{1}{T}\widehat{f}(\frac{k}{T})\exp(2\pi i t\frac{k}{T}) \approx \int_{-\infty}^{+\infty} \widehat{f}(\omega)\exp(2\pi i t\omega) \ \mathrm{d}\omega$$

(here we used a Riemann sum approximation): we may expect that $f(t)=\int_{-\infty}^{+\infty}\widehat{f}(\omega)\exp(2\pi it\omega)\ \mathrm{d}\omega$ and $f(t)=\widehat{\widehat{f}}(-t)$.

We now give a more formal introduction.

3.2 Notations. The space of all functions f for which $\|f\|_1\equiv\int_{-\infty}^{+\infty}|f(t)|\ \mathrm{d}t<\infty$ is denoted by $L^1(\mathbb{R})$, while $L^2(\mathbb{R})$ denotes the space of all functions f for which $\|f\|_2\equiv\sqrt{\int_{-\infty}^{+\infty}|f(t)|^2\ \mathrm{d}t}<\infty$.

3.3 Fourier integrals for L^1-functions. For $f\in L^1(\mathbb{R})$ we define

$$\widehat{f}(\omega)\equiv\int_{-\infty}^{+\infty}f(t)e^{-2\pi it\omega}\ \mathrm{d}t\quad\text{for all}\quad\omega\in\mathbb{R}.$$

3.4 Theorem. *Let $f\in L^1(\mathbb{R})$.*

(a) *$\widehat{f}$ is bounded: $|\widehat{f}(\omega)|\leq\|f\|_1$ for all $\omega\in\mathbb{R}$.*

(b) *$\widehat{f}$ is a uniformly continuous function on $\mathbb{R}$.*

(c) *$\widehat{f}$ vanishes at $\pm\infty$: $\lim_{|\omega|\to\infty}\widehat{f}(\omega)=0$.*

Proof. (a) $|\widehat{f}(\omega)|\leq\int_{-\infty}^{+\infty}|f(t)|\,|\exp(-2\pi it\omega)|\ \mathrm{d}t=\|f\|_1$.
(b) For each $\omega\in\mathbb{R}$ and $\delta>0$ we have that

$$\widehat{f}(\omega)-\widehat{f}(\omega+\delta)=\int_{-\infty}^{+\infty}f(t)e^{-\pi it(2\omega+\delta)}(e^{+\pi it\delta}-e^{-\pi it\delta})\ \mathrm{d}t.$$

Hence, $|\widehat{f}(\omega)-\widehat{f}(\omega+\delta)|\leq\int_{-\infty}^{+\infty}|f(t)|\,|2\sin 2\pi t\delta|\ \mathrm{d}t$.

For $\varepsilon>0$, select a $T>0$ such that $\int_{|t|>T}|f(t)|dt<\frac{1}{4}\varepsilon$. Then select $\delta>0$ such that $|2\sin 2\pi t\delta|\leq\frac{1}{2\|f\|_1}\varepsilon$ for all t for which $|t|\leq T$, and we obtain that we have $|\widehat{f}(\omega)-\widehat{f}(\omega+\delta)|\leq\varepsilon$.
(c) Let $\omega\in\mathbb{R}$. Note that the substitution $t=s+\frac{1}{2\omega}$ leads to

$$\widehat{f}(\omega)=\int_{-\infty}^{+\infty}f(t)e^{-2\pi it\omega}\ \mathrm{d}t=-\int_{-\infty}^{+\infty}f(s+\frac{1}{2\omega})e^{-2\pi is\omega}\ \mathrm{d}s.$$

Hence, $\widehat{f}(\omega)=\frac{1}{2}(\widehat{f}(\omega)+\widehat{f}(\omega))=\int_{-\infty}^{+\infty}(f(t)-f(t+\frac{1}{2\omega}))e^{-2\pi it\omega}\ \mathrm{d}t$ and

$$|\widehat{f}(\omega)|\leq\int_{-\infty}^{+\infty}|f(t)-f(t+\frac{1}{2\omega})|\ \mathrm{d}t.$$

This estimate leads to (c); see Exercise 1.21. □

Note that (c) can be viewed as an analogue of the Riemann–Lebesgue lemma in §2.11.

As in §2.5, differentiation transforms to multiplication.

3.5 Theorem. *Let $f \in L^1(\mathbb{R})$.*
(a) *If f is differentiable and f' also belongs to $L^1(\mathbb{R})$, then*

$$\widehat{f'}(\omega) = 2\pi i\omega \widehat{f}(\omega) \quad (\omega \in \mathbb{R}).$$

(b) *If $tf \in L^1(\mathbb{R})$, then $\widehat{f} \in C^{(1)}(\mathbb{R})$ and $\widehat{f}^{(1)} = -2\pi i\widehat{(tf)}$.*
Here tf denotes the function $t \rightsquigarrow tf(t) \quad (t \in \mathbb{R})$.
Proof. (a) Integrate by parts and use the fact that f is continuous and vanishes at ∞ if both f and f' are in $L^1(\mathbb{R})$ (see Exercise 1.20).
(b) Apply Lebesgue's theorem to see that

$$\widehat{f}^{(1)}(\omega) = \lim_{\Delta\omega \to 0} \frac{\widehat{f}(\omega + \Delta\omega) - f(\omega)}{\Delta\omega} = -2\pi i \int tf(t)e^{-2\pi it\omega}\, \mathrm{d}t$$

□

Note that, for each $n \geq 0$, $t^n f$ is in $L^1(\mathbb{R})$ if f has a *bounded support*, i.e., if the *support* $\{t \in \mathbb{R} \,\big|\, f(t) \neq 0\}$ *of* f is contained in a interval $[-T, T]$ for some $T > 0$: the support of f is bounded by T.

3.6 Corollary. *If $f \in L^1(\mathbb{R})$ has a support bounded by $T > 0$, then*

$$\widehat{f} \in C^{(\infty)}(\mathbb{R}), \quad \|\widehat{f}^{(n)}\|_\infty \leq (2\pi T)^n \|f\|_1 \quad (n \in \mathbb{N}_0),$$

and f is an analytic function, i.e., the Taylor series converges on $\mathbb{R}$.
Proof. Inductive application of Theorem 3.5 shows that $\widehat{f} \in C^{(n)}(\mathbb{R})$ for each $n \in \mathbb{N}_0$, $\widehat{f}^{(n)}(\omega) = (-2\pi i)^n \widehat{(t^n f)}(\omega$, and $\|\widehat{f}^{(n)}\|_\infty \leq (2\pi)^n \|t^n f\|_1 \leq (2\pi T)^n \|f\|_1$.

For $\omega \in \mathbb{R}$ and $n \in \mathbb{N}$, the remainder term of the $(n-1)th$ order Taylor series at 0 of $\widehat{f}$ evaluated at ω is equal to $\frac{\omega^n}{n!}\widehat{f}^{(n)}(\xi)$. Here ξ is some real number between 0 and ω. This term can be bounded by $\frac{(2\pi\omega T)^n}{n!}\|f\|_1$. Since, for each $\kappa > 0$ $\frac{\kappa^n}{n!} \to 0$ if $n \to \infty$, we see that the Taylor series converges.

Note that the convergence is uniform on each interval $[-\Omega, +\Omega]$. □

The above results show that $\widehat{f}$ is a 'nice' function in case $f \in L^1(\mathbb{R})$. Actually, $\widehat{f}$ is so nice that we immediately see that the reverse formula suggested in 3.1 will not be obvious: because, if $\widehat{f} \in L^1(\mathbb{R})$ then $\widehat{\widehat{f}}$ is continuous and the relation $\widehat{\widehat{f}}(-t) = f(t)$ can not be correct for all t for L^1-functions f that have a discontinuity .

Before we discuss how to reconstruct f from $\widehat{f}$, we consider two examples.

3.7 Examples.
(a) *Consider the Gaussian function f defined by $f(t) \equiv \exp(-\pi t^2)$ for $t \in \mathbb{R}$. Then $f \in L^1(\mathbb{R})$ and*

$$\widehat{f}(\omega) = \exp(-\pi\omega^2) \quad \textit{for each} \quad \omega \in \mathbb{R}, \tag{69}$$

as we will show now. Note that

$$\int_{-\infty}^{+\infty} \exp(-\pi t^2 - 2\pi it\omega)\, \mathrm{d}t = \exp(-\pi\omega^2) \int_{-\infty}^{+\infty} \exp(-\pi(t + i\omega)^2)\, \mathrm{d}t.$$

To show that $g(\omega)\equiv\int_{-\infty}^{+\infty}\exp(-\pi(t+i\omega)^2)\,\mathrm{d}t=1$, we first note that

$$g(0)=\int_{-\infty}^{+\infty}\exp(-\pi t^2)\,\mathrm{d}t=1.$$

Moreover,

$$g'(\omega)=\int_{-\infty}^{+\infty}-2\pi i(t+i\omega)\exp(-\pi(t+i\omega)^2)\,\mathrm{d}t=i\exp(-\pi(t+i\omega)^2)\Big|_{t=-\infty}^{t=+\infty}=0.$$

Hence, $g(\omega)=g(0)+\int_0^\omega g'(\nu)\,\mathrm{d}\nu=1$ for all $\omega\in\mathbb{R}$.
(b) For $T>0$, consider the 'top-hat' function Π_T given by[10]

$$\Pi_T(t)\equiv 1\quad\text{if } |t|\leq T,\quad\text{and}\quad \Pi_T(t)\equiv 0\quad\text{if } |t|>T. \tag{70}$$

Then,

$$\widehat{\Pi_T}(\omega)=\int_{-T}^{+T}e^{-2\pi it\omega}\,\mathrm{d}t=\frac{1}{-2\pi i\omega}e^{-2\pi it\omega}\Big|_{t=-T}^{t=+T}=\frac{1}{\pi\omega}\sin(2\pi T\omega).$$

The function $t\rightsquigarrow\sin(\pi t)/(\pi t)$ plays an important role in Physics and is called the *sinc-function*:

$$\mathrm{sinc}(t)\equiv\frac{\sin(\pi t)}{\pi t}\qquad(t\in\mathbb{R}). \tag{71}$$

The Fourier transform of the scaled top-hat function $\frac{1}{2T}\Pi_T$ with height $1/(2T)$ and width $2T$ is equal to $\omega\rightsquigarrow\mathrm{sinc}(2T\omega)$:

$$\widehat{\frac{1}{2T}\Pi_T}(\omega)=\mathrm{sinc}(2T\omega)\quad(\omega\in\mathbb{R}); \tag{72}$$

see Fig. 7 for a graph of the top-hat function and the sinc-function.
Note that $\mathrm{sinc}(0)=1$ and $\mathrm{sinc}(k)=0$ for all $k\in\mathbb{Z}$, $k\neq 0$.

The first example supports the expectation that the reverse formula for Fourier integrals is correct. The second example is slightly discouraging. The transform function is uniformly continuous and vanishes at infinity, which is in line with the theorems. Unfortunately, sinc is not absolutely integrable: $\mathrm{sinc}\not\in L^1(\mathbb{R})$. However, observe that sinc is quadratically integrable: $\mathrm{sinc}\in L^2(\mathbb{R})$.

The following three lemmas allow us to define the Fourier transform also for functions in $L^2(\mathbb{R})$. First note that $f\in L^1(\mathbb{R})$ if $f\in L^2(\mathbb{R})$ and has a bounded support.

3.8 Lemma. *Let $f\in L^2(\mathbb{R})$ with support bounded by $L>0$. Then, for $T\geq 2L$,*

$$\|f\|_2^2=\sum_{k\in\mathbb{N}}\frac{1}{T}|\widehat{f}(\frac{k}{T})|^2\quad\textit{and}\quad\lim_{n\to\infty}\int_{-L}^{+L}|f(t)-\sum_{|k|\leq n}\frac{1}{T}\widehat{f}(\frac{k}{T})\exp(2\pi i\frac{k}{T}t)|^2\,\mathrm{d}t=0.$$

Proof. Note that $\widehat{f}(\omega)=\int_{-\frac{T}{2}}^{+\frac{T}{2}}f(t)e^{-2\pi it\omega}\,\mathrm{d}t$. With $\gamma_k(f)$ from §2.2, $\gamma_k(f)=\frac{1}{T}\widehat{f}(\frac{k}{T})$. The second statement of the lemma follows from (a) of Theorem 2.4. Parseval's formula

[10] The symbol Π is chosen as an obvious aid to memory.

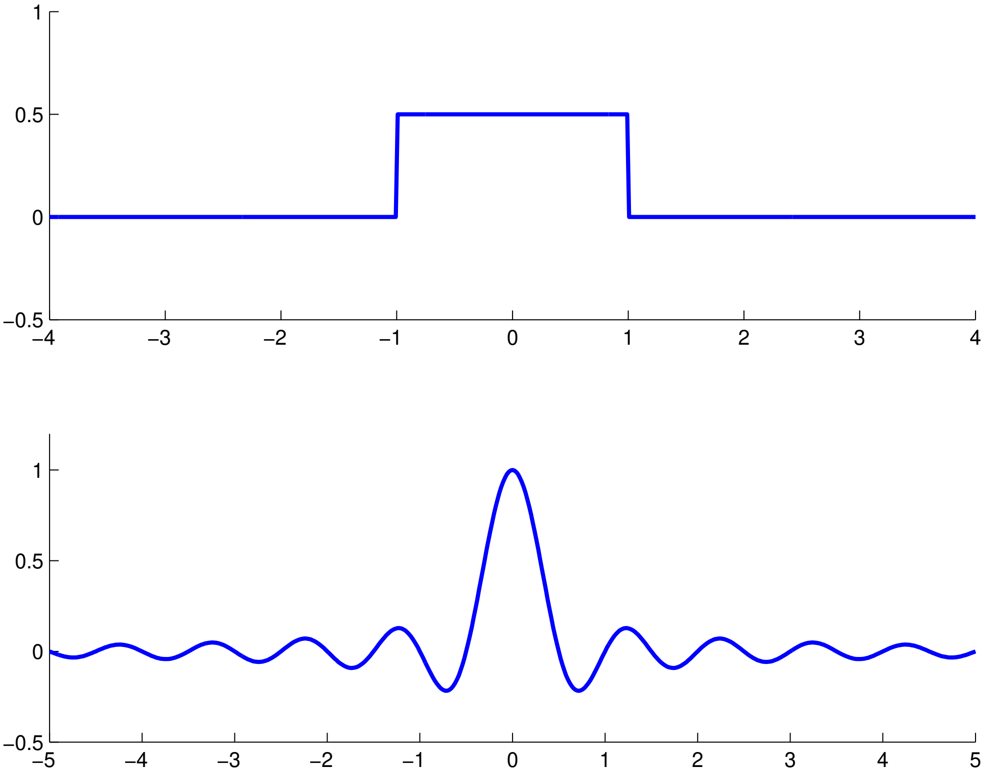


FIGURE 7. The top figure shows the graph of the scaled top-hat function $\frac{1}{2T}\Pi_T$ for $T=1$, the bottom figure displays its Fourier transform, the sinc function $\omega \rightsquigarrow \mathrm{sinc}(2T\omega)$ also for $T=1$.

implies that $\frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} |f(t)|^2\,\mathrm{d}t = \sum_{k\in\mathbb{Z}} |\frac{1}{T}\widehat{f}(\frac{k}{T})|^2$ and $\int_{-\infty}^{+\infty} |f(t)|^2\,\mathrm{d}t = \sum_{k\in\mathbb{Z}} \frac{1}{T}|\widehat{f}(\frac{k}{T})|^2$. $\square$

Apparently, $\lim_{T\to\infty}\sum_{k\in\mathbb{Z}}\frac{1}{T}|\widehat{f}(\frac{k}{T})|^2$ exists and, since $\widehat{f}$ is uniformly continuous, it is tempting to conclude that the limit equals $\int_{-\infty}^{+\infty}|\widehat{f}(\omega)|^2\,\mathrm{d}\omega = \|\widehat{f}\|_2^2$. This conclusion is correct (see Lemma 3.9 below) but the proof requires somewhat more work. The proof relies on the fact that $t^n f \in L^1(\mathbb{R})\cap L^2(\mathbb{R})$ if $f\in L^2(\mathbb{R})$ has a bounded support ($n=0,1$). Hence, $\widehat{f}\in C^{(\infty)}(\mathbb{R})$ (see Th. 3.5) and arguments for f and $\widehat{f}$ apply to tf and $\widehat{f}^{(1)}$ as well.

3.9 Lemma. *If $f\in L^2(\mathbb{R})$ has a bounded support, then $\widehat{f}\in L^2(\mathbb{R})$, $\|f\|_2=\|\widehat{f}\|_2$, and, $\|f-f_\Omega\|_2\to 0$ if $\Omega\to\infty$, where $f_\Omega(t)\equiv\int_{-\Omega}^{\Omega}\widehat{f}(\omega)\,e^{2\pi i t\omega}\,\mathrm{d}\omega$.*

Proof. Since $|\widehat{f}|^2\in C(\mathbb{R})$, the first formula in Lemma 3.8 and (b) of Exercise 1.12 implies that $|\widehat{f}|^2\in L^1(\mathbb{R})$, whence $\widehat{f}\in L^2(\mathbb{R})$. Replacing f by tf shows that $\widehat{f}^{(1)}\in L^2(\mathbb{R})$ and (e) of Exercise 1.12 tells us that $\|f\|_2=\|\widehat{f}\|_2$.

To prove the last claim in the lemma, assume the support of f is bounded by $L>0$, and, for $\Omega>0$, $T>2L$, consider the functions

$$f_{T,\Omega}(t)\equiv\frac{1}{T}\sum_{k\in\mathbb{Z},|\frac{k}{T}|\le\Omega}\widehat{f}(\frac{k}{T})\,e^{2\pi i t\frac{k}{T}}\qquad(t\in\mathbb{R}).$$

Parseval's formula and Theorem 2.4(a) implies that

$$\int_{-T}^{T} |f(t) - f_{T,\Omega}(t)|^2 \,\mathrm{d}t = \sum_{k\in\mathbb{Z}, |\frac{k}{T}|>\Omega} \frac{1}{T}|\widehat{f}(\frac{k}{T})|^2 \leq \int_{|\omega|\geq\Omega} h(\omega)\,\mathrm{d}\omega + \frac{1}{T}\|h'\|_1.$$

Here, $h \equiv |\widehat{f}|^2$, and for the last estimate we used (20) in Exercise 1.12.
Since $\omega \rightsquigarrow \widehat{f}(\omega)e^{2\pi i t\omega}$ is uniformly continuous, we have that, for any $K > 0$,

$$\int_{-K}^{K} |f_\Omega(t) - f_{T,\Omega}(t)|^2 \,\mathrm{d}t \to 0 \quad \text{if } T \to 0.$$

Combining these two results shows that $\int_{-K}^{K} |f(t) - f_\Omega(t)|^2 \,\mathrm{d}t \leq \int_{|\omega|\geq\Omega} h(\omega)\,\mathrm{d}\omega$, whence $\|f - f_\Omega\|_2^2 \leq \int_{|\omega|\geq\Omega} h(\omega)\,\mathrm{d}\omega$. Since $h \in L^1(\mathbb{R})$, the lemma follows. $\square$

3.10 Lemma. *Let $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Then $\widehat{f} \in L^2(\mathbb{R})$ and $\|f\|_2 = \|\widehat{f}\|_2$.*

Proof. For $n \in \mathbb{N}$, consider $f_n \equiv f\Pi_n$. Note that $f_n \in L^2(\mathbb{R})$ is of bounded support. By Lemma 3.8, $\|f_n - f_m\|_2 = \|\widehat{(f_n - f_m)}\|_2 = \|\widehat{f_n} - \widehat{f_m}\|_2$: $(\widehat{f_n})$ is a Cauchy sequence in $L^2(\mathbb{R})$. Hence, $\|\widehat{f_n} - g\|_2 \to 0$ $(n \to \infty)$ for some $g \in L^2(\mathbb{R})$. Since $\widehat{f_n}(\omega) \to \widehat{f}(\omega)$, we have that $\|\widehat{f} - g\|_2 = 0$. Apparently, $\widehat{f} \in L^2(\mathbb{R})$ and $\|\widehat{f_n} - \widehat{f}\|_2 \to 0$ $(n \to \infty)$.

Since also $\|f_n - f\|_2 \to 0$ and $\|\widehat{f_n}\|_2 = \|f_n\|_2$, we find that $\|\widehat{f}\|_2 = \|f\|_2$. $\square$

We can define the Fourier transform for functions in $L^2(\mathbb{R})$.

3.11 Fourier integrals for L^2-functions. If $f \in L^2(\mathbb{R})$, then there is a sequence (f_n) in $L^2(\mathbb{R})\cap L^1(\mathbb{R})$ for which $\|f - f_n\|_2 \to 0$ (e.g., $f_n = f\Pi_n$). According Lemma 3.10, $(\widehat{f_n})$ is a Cauchy sequence in $L^2(\mathbb{R})$. Hence, there is a $g \in L^2(\mathbb{R})$ for which $\|\widehat{f_n} - g\|_2 \to 0$. The limit function g is in some sense unique:
(i) If $\widetilde{g} \in L^2(\mathbb{R})$ shows up in a similar way as limit function of another sequence of functions in $L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ that approximate f, then $\|g - \widetilde{g}\|_2 = 0$.
(ii) If $f \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$, then, by Lemma 3.10, $\|g - \widehat{f}\|_2 = 0$.
Since there is no real confusion concerning the selection of g, we put $\widehat{f}$ instead of g.
Usually, we have that $\widehat{f}(\omega) = \lim_{T\to\infty} \int_T^T f(t)e^{-2\pi i t\omega}\,\mathrm{d}t$ for almost all $\omega \in \mathbb{R}$.
Therefore, we simply use the formula $\widehat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-2\pi i t\omega}\,\mathrm{d}t$.

If we respect the conventions from 3.11, then Lemma 3.9 leads to the following analogue of (a) of Theorem 2.4 and of Parseval's formula ((34) and (35) in §2) can be proved. The analogue of Parseval's formula is know as *Plancherel's formula*.

3.12 Theorem. *Let $f \in L^2(\mathbb{R})$. Then,*

$$\widehat{f} \in L^2(\mathbb{R}), \qquad \|f\|_2 = \|\widehat{f}\|_2, \qquad (f,g) = (\widehat{f},\widehat{g}) \quad (g \in L^2(\mathbb{R})), \tag{73}$$

$$\widehat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-2\pi i t\omega}\,\mathrm{d}t, \quad \textit{and} \quad f(t) = \int_{-\infty}^{+\infty} \widehat{f}(\omega)e^{+2\pi i t\omega}\,\mathrm{d}\omega. \quad \square \tag{74}$$

As in (c) and (d) of Theorem 2.4, there are onditions on f that guarantee that the sequence $(I_n(f))$, where $I_n(f)(t) \equiv \int_{-n}^{+n} \widehat{f}(\omega)e^{+2\pi i t\omega}\,\mathrm{d}\omega$, converges uniformly or point-wise to f. We refer to literature for more details.

3.13 Note. The factor 2π in the definition of the Fourier integrals often shows up in literature on other places.

(a) For sound reasons, the following relations are often used in physics

$$F(\omega)=\int_{-\infty}^{+\infty} f(t)e^{-it\omega}\,\mathrm{d}t \quad\text{and}\quad f(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)e^{+it\omega}\,\mathrm{d}\omega.$$

(b) For mathematical reasons, many mathematical textbooks employ

$$F(\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(t)e^{-it\omega}\,\mathrm{d}t \quad\text{and}\quad f(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} F(\omega)e^{+it\omega}\,\mathrm{d}\omega.$$

In applications, the functions f should often be real-valued. It would be useful if this could be concluded from $\widehat{f}$. The relations in the theorem lead to the following result.

3.14 Property. *Let f be in $L^2(\mathbb{R})$ or in $L^1(\mathbb{R})$.*
Then, f is real-valued $\Leftrightarrow$ $\widehat{f}$ is even, f is even $\Leftrightarrow$ $\widehat{f}$ is real-valued. □

With Fourier integrals, electric networks (cf. 2.6) can easily be analysed.

3.15 Application (continued). Again, consider the electric network in the last paragraph of Application 2.6.

As before, V is the voltage difference between two end points in the network and I is the resulting current (V is not periodic now). Then, assuming that V and I are L^2-functions (which is physically reasonable), we have that $\widehat{V}(\omega)=Z(\omega)\widehat{I}(\omega)$ where $Z(\omega)=\frac{1}{2\pi i\omega C}+2\pi i\omega L+R$.

3.16 Interpretation. In many fields in Physics and Technology, t is a point in time and $\mathbb{R}$ represents *time*. The value $f(t)$ may describe the voltage drop at time t between two points in an electric network or the changes in pressure in the air at time t at a certain place in space (such changes in pressure may be caused by sound waves). The function f is called a *signal* then and $\int_{-\infty}^{+\infty}|f(t)|^2\,\mathrm{d}t$ can be interpreted as the *energy of the signal*. All signals have finite energy. Therefore, the space $L^2(\mathbb{R})$ is also called *the space of all signals*. The quantity $\Delta\omega|\widehat{f}(\omega)|^2$ can be viewed as the energy contained in the part of the signal contained in the frequency band $(\omega-\frac{1}{2}\Delta\omega,\omega+\frac{1}{2}\Delta\omega)$: $|\widehat{f}(\omega)|^2$ is the *power spectrum* or *spectral power density* of the signal.

Periodic functions in $L^2_T(\mathbb{R})$ are sometimes called stationairy signals, herewith referring to the fact that the function is the same after T-seconds. Non-periodic signals are said to time-dependent.

Since $f(t)=\int\widehat{f}(\omega)e^{2\pi it\omega}\,\mathrm{d}\omega$, the function f can be viewed as a superposition of the harmonic oscillations $t\rightsquigarrow\widehat{f}(\omega)e^{2\pi it\omega}$ with frequency ω.[11] The function f has *amplitude* $|\widehat{f}(\omega)|$ and *phase* $\phi(\omega)$ at frequency ω. Here, $\phi(\omega)\in[0,2\pi)$ is such that $\widehat{f}(\omega)=|\widehat{f}(\omega)|\,e^{-i\phi(\omega)}$.

The functions f and $\widehat{f}$ can be identified (via the Fourier transforms). In discussions, f is often referred to as the function in *time domain* and $\widehat{f}$ as the function in *frequency domain.*

[11] In 1 second, $t\rightsquigarrow e^{2\pi it\omega}$ runs through ω periods, that is, the number of $t\in[T,T+1]$ for which $e^{2\pi it\omega}=1$, is, on average, averaging over all $T\in\mathbb{R}$, equal to ω.

3.17 Extensions. The function $\phi_\nu(t) \equiv e^{2\pi i\nu t}$ $(t \in \mathbb{R})$ is not in $L^1(\mathbb{R})$, nor in $L^2(\mathbb{R})$ and its Fourier transform has not been formally defined. The inversion formula $f(t) = \int \widehat{f}(\omega)\exp(2\pi i t\omega)\,\mathrm{d}\omega$ for functions in $L^2(\mathbb{R})$ tells us that f can be viewed as a superposition of harmonic oscillations. The function ϕ_ν is an harmonic oscillation itself (with frequency ν) and only one oscillation contributes to ϕ_ν. Thefore, if such an inversion formula makes sense for ϕ_ν, then $\widehat{\phi_\nu}(\omega) = 0$ for all $\omega \neq \nu$. It is tempting to postulate that $\widehat{\phi_\nu}(\nu) = 1$. However, such a function is equivalent to the zero function and the integral $\int \widehat{\phi_\nu}(\omega)\exp(2\pi i\omega t)\,\mathrm{d}\omega$ is equal to zero rather than to the desired value $\exp(2\pi i\nu t)$; we are looking for a 'function' or construction δ_ν with the following property

$$\int_{-\infty}^{\infty} \delta_\nu(x)\,g(x)\,\mathrm{d}x = g(\nu) \quad \text{for all } g \in C(\mathbb{R}).$$

Of course, δ_ν is not a function —it would be zero at $x \neq \nu$ and ∞ at ν. But, it does form a bounded linear map on the space of continuous functions that vanish at infinity.[12] Therefore, δ_ν can be identified with a so-called 'measure'. Nevertheless, physicists and technicians like to view δ_ν as a 'function' and call it the *Dirac delta function* at ν. For mathematicians, δ_ν is the *point-measure* at ν.

For δ_ν, we have that $\widehat{\delta_\nu}(t) \equiv \int \delta_\nu(\omega)\,e^{2\pi i t\omega}\,\mathrm{d}\omega = e^{2\pi i t\nu}$. Conversely, if an inversion formula can be defined for ϕ_ν (and it can, as we will argue below), then it should be $\widehat{\phi_\nu} = \delta_\nu$.

Recall that we had to employ limit processes in order to get the Fourier transform of $L^2(\mathbb{R})$ functions rigorously defined. The Dirac delta function can also be obtained as the result of a limit process. To see this and to simplify notation, take $\nu = 0$. Now, note that

$$\lim_{\varepsilon\to 0}\int_{-\infty}^{\infty}\frac{1}{2\varepsilon}\Pi_\varepsilon(x)\,g(x)\,\mathrm{d}x = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_{-\varepsilon}^{+\varepsilon} g(x)\,\mathrm{d}x = \lim_{\varepsilon\to 0}\frac{G(\varepsilon)-G(-\varepsilon)}{2\varepsilon} = g(x).$$

Here, g is continuous and G is its primitive ($G' = g$). Hence, in some weak sense, the Dirac delta function δ_0 at 0 appears to be the limit of the functions $\frac{1}{2\varepsilon}\Pi_\varepsilon$ for $\varepsilon \to 0$. The function values $\frac{1}{2\varepsilon}\Pi_\varepsilon(x)$ converge to 0 if $x \neq 0$, and to ∞ if $x = 0$ $(\varepsilon \to 0)$. However, there are other sequences of continuous functions that converge weakly to δ_0 and that do not converge point-wise.[13] For instance, there is no x for which $\frac{2}{\varepsilon}\mathrm{sinc}(x\frac{2}{\varepsilon})$ converges for $\varepsilon \to 0$, while $\int \frac{2}{\varepsilon}\mathrm{sinc}(x\frac{2}{\varepsilon})\,g(x)\,\mathrm{d}x \to g(0)$ $(\varepsilon \to 0)$ for each smooth function g with bounded support; see Exercise 3.17. Hence, $(x \rightsquigarrow \frac{2}{\varepsilon}\mathrm{sinc}(x\frac{2}{\varepsilon}))$ converges weakly to δ_0, or, using a more sloppy notation, $(\frac{2}{\varepsilon}\mathrm{sinc}(x\frac{2}{\varepsilon}))$ converges weakly to $\delta_0(x)$ $(\varepsilon \to 0)$. In particular, δ_0 can be viewed as the Fourier transform of ϕ_0:

$$\int_{-\frac{1}{\varepsilon}}^{\frac{1}{\varepsilon}} \phi_0(t)\,e^{-2\pi i t\omega}\,\mathrm{d}t = \int_{-\frac{1}{\varepsilon}}^{\frac{1}{\varepsilon}} e^{-2\pi i t\omega}\,\mathrm{d}t = \tfrac{2}{\varepsilon}\mathrm{sinc}(\omega\tfrac{2}{\varepsilon}) \to \delta_0(\omega) \quad (\varepsilon \to 0).$$

This approach may help to understand some puzzling 'consequences' of Fourier's inversion formula. Routine manipulation with integrals would lead to

$$\begin{aligned} f(t) &= \int \widehat{f}(\omega)\,e^{2\pi i t\omega}\,\mathrm{d}\omega = \int \left(\int f(s)e^{-2\pi i s\omega}\,\mathrm{d}s\right)e^{2\pi i t\omega}\,\mathrm{d}\omega \\ &= \int\int f(s)e^{2\pi i\omega(t-s)}\,\mathrm{d}\omega\,\mathrm{d}s = \int f(s)\left(\int e^{2\pi i\omega(t-s)}\,\mathrm{d}\omega\right)\,\mathrm{d}s. \end{aligned}$$

[12] If $C_\infty(\mathbb{R})$ is the space of continuous functions g that vanish at infinity (that is, $g(x) \to 0$ if $|x| \to \infty$), then a linear map μ from $C_\infty(\mathbb{R})$ to $\mathbb{C}$ is *bounded* if, for some $\kappa > 0$, $|\mu(g)| \leq \kappa\|g\|_\infty$ $(g \in C_\infty(\mathbb{R}))$. Such a map can be identified with a *measure* and the notation $\int g(x)\,\mathrm{d}\mu(x)$ is used instead of $\mu(g)$.

[13] A sequence (f_ε) of functions converge *wealkly* to a measure μ if $\int f_\varepsilon(x)\,g(x)\,\mathrm{d}x \to \int g(x)\,\mathrm{d}\mu(x)$ for each function $g \in C^{\infty)}(\mathbb{R})$ with bounded support. A function g has a *bounded support* if it vanishes outside some bounded interval, that is, if there is an $L > 0$ such that $g(x) = 0$ whenever $|x| > L$.

All integrals are from $-\infty$ to ∞. The expression $\int e^{2\pi i\omega(t-s)}\,\mathrm{d}\omega$ is hard to interpret, since $|e^{2\pi\omega(t-s)}|$, nor its square, is integrable. However, is we saw above, this expression can be viewed as the Dirac delta function and that leads to the familiar conclusion:

$$\int f(s)\left(\int e^{2\pi i\omega(t-s)}\,\mathrm{d}\omega\right)\,\mathrm{d}s=\int f(s)\,\delta_0(s-t)\,\mathrm{d}s=f(t).$$

The discussions appear to be consistent.

For a rigorous and extended discussion on Fourier transforms for more general type of functions and measures, we refer to a course on *'Distribution Theory'*.

Exercises

Exercise 3.1.
(a) Prove that $f \rightsquigarrow \widehat{f}$ is linear.

Exercise 3.2. Let $f\in L^1(\mathbb{R})$ and let $a,\alpha,\nu\in\mathbb{R}$.
Prove the following, where t and ω range over $\mathbb{R}$.
(a) If $f_a(t)\equiv f(t-a)$, then $\widehat{f_a}(\omega)=e^{-2\pi i a\omega}\widehat{f}(\omega)$.
(b) If $g(t)\equiv e^{2\pi i t\nu}f(t)$, then $\widehat{g}(\omega)=\widehat{f}(\omega-\nu)=(\widehat{f})_\nu(\omega)$.
(c) If $g(t)\equiv\cos(2\pi t\nu)\,f(t)$, then $\widehat{g}(\omega)=\frac{1}{2}(\widehat{f}(\omega-\nu)+\widehat{f}(\omega+\nu))$.
(d) If $g(t)\equiv\sqrt{\alpha}\,f(\alpha\,t)$, then $\widehat{g}(\omega)=\frac{1}{\sqrt{\alpha}}\widehat{f}(\frac{1}{\alpha}\omega)$.

Exercise 3.3.
(a) For $\lambda\in\mathbb{C}$, consider the function f defined by

$$f(t)\equiv e^{\lambda t}\quad\text{for } t\geq 0\quad\text{and}\quad f(t)\equiv 0\quad\text{for } t<0.$$

Show that $f\in L^1(\mathbb{R})\cap L^2(\mathbb{R})$ if $\mathrm{Re}(\lambda)<0$ and that then

$$h(\omega)\equiv\widehat{f}(\omega)=\frac{1}{2\pi i\omega-\lambda}\qquad(\omega\in\mathbb{R}).$$

(b) Compute the inverse Fourier transform of the function h in case $\mathrm{Re}(\lambda)>0$.
(Hint: consider $h(-\omega)$.)
(c) For $n\in\mathbb{N}$, consider the function $h_n(\omega)\equiv(\lambda-2\pi i\omega)^{-n}$ $(\omega\in\mathbb{R})$. For which $n\in\mathbb{N}$ does h_n belong to $L^2(\mathbb{R})$? When is h_n in $L^1(\mathbb{R})$? Show that the inverse Fourier transform of h_n is continuous for $n\geq 2$ without explicitly computing the inverse Fourier transform.
(d) Compute the inverse Fourier transform of $h_n(\omega)\equiv(\lambda-2\pi i\omega)^{-n}$.
(Hint: Consider the nth derivative $h^{(n)}$ of h).

Exercise 3.4.
(a) Compute the Fourier transform of the function $t\rightsquigarrow\sqrt{\alpha}\,\cos(2\pi t\nu)\,e^{-\pi(\alpha t)^2}$.
A function of the form $t\rightsquigarrow\cos(2\pi t\nu)f_0(t)$ is a so-called *wavepacket*, with (frequency ν and) *envelope* f_0.
(b) Compute the Fourier transform of the function $t\rightsquigarrow\sqrt{\alpha}\,\cos^2(2\pi t\nu)\,e^{-\pi(\alpha t)^2}$.
(c) Compute the Fourier transform of the function f defined by $f(t)=1$ if $||t|-d|\leq\frac{1}{2}a$ and $f(t)=0$ elsewhere. Here, $a,d>0$ such that $a<2d$.
(d) Let $a,d>0$ such that $a<d$. Let $d_k\equiv kd$ for $k\in\mathbb{Z}$, $|k|\leq K$. Compute the Fourier transform of f, where the function f is defined by $f(t)=1$ if $|t-d_k|\leq\frac{1}{2}a$ for some $k\in\mathbb{Z}$, $|k|\leq K$, and $f(t)=0$ elsewhere.

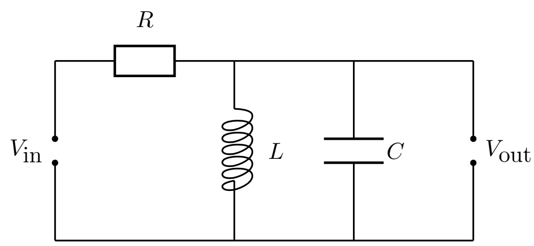


FIGURE 8. A simple electronic circuit.

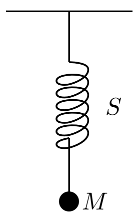


FIGURE 9. Harmonic oscillator.

Hint: to get a manageable expression, use the fact that

$$\sum_{k=-K}^{K} \zeta^k = \zeta^{-K}\,\frac{1-\zeta^{2K+1}}{1-\zeta} = \frac{\zeta^{-K-\frac{1}{2}} - \zeta^{K+\frac{1}{2}}}{\zeta^{-\frac{1}{2}} - \zeta^{\frac{1}{2}}}.$$

Exercise 3.5. Prove the following statements.

(a) f is real $\Leftrightarrow$ $\widehat{f}$ is even. f is purely imaginary $\Leftrightarrow$ $\widehat{f}$ is odd.

(b) f is even $\Leftrightarrow$ $\widehat{f}$ is real. f is odd $\Leftrightarrow$ $\widehat{f}$ is purely imaginary.

Exercise 3.6. Let f be a continuous function in $L^1(\mathbb{R})$.
Consider the function $F(t) \equiv \sum_{k\in\mathbb{Z}} f(t-k)$.

(a) Show that $\int_0^1 |F(t)|\,\mathrm{d}t \leq \|f\|_1$.

(b) Show that F is 1-periodic. Hence $F \in L^1_1(\mathbb{R})$.

(c) Prove that F is constant if and only if $\widehat{f}(k) = 0$ for all $k \in \mathbb{Z}$, $k \neq 0$.

Exercise 3.7. **Complex Gaussian function.**

For $\alpha \in \mathbb{C}$, let f be given by $f(t) \equiv \exp(-\alpha\,\pi\,t^2)$ $(t \in \mathbb{R})$.

(a) Show that $f \in L^1(\mathbb{R})$ if $\mathrm{Re}(\alpha) > 0$. Show that $f \in L^\infty(\mathbb{R})$ if $\alpha \in i\mathbb{R}$. Does f belong to $L^1(\mathbb{R})$ or to $L^2(\mathbb{R})$ for these purely imaginary α?

(b) We want to compute $\widehat{f}$ in case $\mathrm{Re}(\alpha) > 0$. It is tempting to try and to combine (69) and (d) of Exercise 3.2. Unfortunately, (d) of Exercise 3.2 is not applicable if $\alpha \notin \mathbb{R}$. Why not? Nevertheless, use (d) of Exercise 3.2 to compute $\widehat{f}$.

(c) Consider the case where $\mathrm{Re}(\alpha) > 0$. Prove that

$$\widehat{f}(\omega) = \sqrt{\frac{1}{\alpha}}\,e^{-\pi\frac{\omega^2}{\alpha}} \qquad (\omega \in \mathbb{R}). \tag{75}$$

If $\mathrm{Re}(\alpha) = 0$, then it seems reasonable to define $\widehat{f}(\omega) = \sqrt{\frac{1}{\alpha}}\,e^{-\pi\frac{\omega^2}{\alpha}}$ $(\omega \in \mathbb{R})$. Why?

Exercise 3.8. Compute the Fourier transform of $\frac{1}{2}(\delta_\nu + \delta_{-\nu})$.

Exercise 3.9. **Electric circuit.**

Consider the electric circuit in Fig. 8. We put a potential $V_{\rm in}$ at the two points of the left hand side of the circuit. We are interested in the resulting voltage difference $V_{\rm out}$ at the points at the right hand side of the circuit. $V_{\rm in}$ is sufficiently smooth ($V_{\rm in} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$).

(a) Derive an expression of the form $\widehat{V_{\rm out}}(\omega) = H(\omega)\widehat{V_{\rm in}}(\omega)$ and determine the '*transfer function*' H.

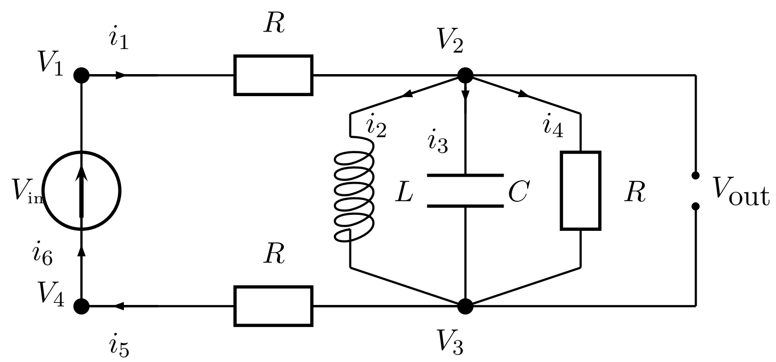


FIGURE 10. An electric network can be viewed as a directed graph. Volages are measured on the vertices of the graph. Each edge contains exactly one electrical component. See Exercise 3.11.

(b) Sketch, for $R = 1$, $L = 0.14$ and $C = 10$, the graph of the absolute value $|H|$ of H.

Exercise 3.10. **Harmonic oscillator.**

Consider the equation for a simple *harmonic oscillator*

$$x''(t) + \gamma\, x'(t) + \omega_0^2\, x(t) = f(t). \tag{76}$$

This equation describes the movement of a body M with mass m on a spring S (see Fig. 9). The body moves along the x-axis (vertically in the figure). $x(t)$ is the displacement distance of the body at time t from its position at rest, ω_0 is the natural oscillation frequency, γ is a damping coefficient, $f(t) = F(t)/m$, where $F(t)$ is a force on the body. The force $F(t)$, and therefore $f(t)$ is known (the input). We are interested in the position $x(t)$ (the output). We assume that the input as well as the output is in $L^2(\mathbb{R})$

(a) Assume that $f(t) = f_0 \exp(-2\pi i t\omega)$. Solve (76). Determine the transfer function $H(\omega)$ defined by $x(t)/f(t)$.

(b) For general $f \in L^2(\mathbb{R})$, we have $f(t) = \int \widehat{f}(\omega) \exp(2\pi i t\omega)\, \mathrm{d}\omega \quad (t \in \mathbb{R})$. Show that the associated solution x satisfies $x(t) = \int \widehat{f}(\omega)\, H(\omega)\, \exp(2\pi i t\omega)\, \mathrm{d}\omega \quad (t \in \mathbb{R})$.

Exercise 3.11. **Electric circuits 2.**

An electrical network can be represented by a directed graph: each edge represents a wire with exactly one electrical component. So, some edges contain resistors, others capacitors, and so on. The electrical currency in a wire is positive when flowing according to the direction of the edge. Otherwise, it is negative (or 0). Voltages are measured at the vertices of the graph. A directed graph can be described by a so-called *incidence matrix*. This is an M by N matrix, if M is the number of vertices and N is the number of edges. The M vertices, as well as the N edges are numbered. The jth column corresponds with the jth edge. It takes the value $+1$ at the kth coordinate and the value -1 at the ℓth coordinate if the ith edge starts at the kth vertix and ends at the ℓth one. The other values in the column are 0.

Let $\mathbf{V}$ be the M-vector of voltages, that is, the kth coordinate V_k is the voltage at the kth vertix. Let $\mathbf{I}$ be the N-vector of currencies, i.e., $i_j = I_j$ is the currency in the wire correspoding with the jth edge (positive if in the direction of the edge). Let $\mathbf{A}$ denote the incidence matrix.

(a) Give the incidence matrix for the network in Fig. 10.

(b) Give an interpretion of the value of the kth coordinate of $\mathbf{AI}$. What is the value of this coordinate?

(c) Give an interpretation of the jth coordinate of $\mathbf{A}^{\mathrm{T}}\mathbf{V}$.

Let $\mathbf{C}$ be an N by N diagonal matrix, C_{jj} is the value of the capicitor at the jth wire if this wire contains a capicitor, $C_{jj} = 0$ otherwise. The N by N diagonal mattices $\mathbf{R}$ and $\mathbf{L}$ are defines similarly for resistors and inductors, respectively.

(d) Show that

$$\begin{cases} \mathbf{A}\mathbf{I} = \mathbf{0} \\ \mathbf{A}^{\mathrm{T}}\dot{\mathbf{V}} = \mathbf{R}\dot{\mathbf{I}} + \mathbf{C}^{-1}\mathbf{I} + \mathbf{L}\ddot{\mathbf{I}} + \dot{\mathbf{f}} = 0. \end{cases} \tag{77}$$

What does $\mathbf{f}$ describe?

(e) Has Kirchoff's law of voltages been modelled? How?

(f) Put $\mathbf{J} \equiv \dot{\mathbf{I}}$. Assemble the vectors $\mathbf{I}$, $\mathbf{V}$ and $\mathbf{J}$ in one large vector $\mathbf{X} \equiv (\mathbf{I}^{\mathrm{T}}, \mathbf{V}^{\mathrm{T}}, \mathbf{J}^{\mathrm{T}})^{\mathrm{T}}$. The coupled system (77) can be written as

$$\mathbb{E}\dot{\mathbf{X}} = \mathbb{A}\mathbf{X} + \mathbf{U}.$$

Give a describtion of $\mathbb{E}$ and $\mathbb{A}$ in terms of $\mathbf{A}$, $\mathbf{A}^{\mathrm{T}}$, $\mathbf{R}$, $\mathbf{C}$, $\mathbf{L}$, and the N by N indentity matrix $\mathbf{Id}_N$. Describe $\mathbf{U}$ in terms of $\mathbf{f}$.

(g) Suppose that, in the situation of Fig. 10, we are interested in the voltage difference $V_3 - V_2$. Show that this difference can be expressed as $\mathbf{c}^{\mathrm{T}}\mathbf{X}$ for some vector $\mathbf{c}$. Give an expression for $\mathbf{c}$. This leads to a so-called *control system*

$$\begin{cases} \mathbb{E}\dot{\mathbf{X}} = \mathbb{A}\mathbf{X} + \mathbf{U} \\ Y \equiv \mathbf{c}^{\mathrm{T}}\mathbf{X}. \end{cases}$$

$\mathbf{U}$ is the *control* parameter, Y is the *observable.* Explain these terms.

(h) Now, suppose that $\mathbf{U}$ is of the form $\mathbf{U}(t) = \mathbf{u}\exp(2\pi i t\omega)$ for all $t \in \mathbb{R}$, where $\mathbf{u}$ is a constant vector (of appropriate dimension). Suppose $\mathbf{X}$ and Y can be written as $\mathbf{X}(t) = \mathbf{x}\exp(2\pi i t\omega)$ and $Y(t) = y\exp(2\pi i t\omega)$ with $\mathbf{x}$ and y time independent. Show that

$$y = H(\omega) \equiv \mathbf{c}^{\mathrm{T}}(2\pi i\omega\mathbb{E} - \mathbb{A})^{-1}\mathbf{u} \qquad (\omega \in \mathbb{R}).$$

Exercise 3.12. The wave equation.

We are interested in solutions u of the the wave equation

$$\frac{\partial^2 u}{\partial t^2}(x,t) = c^2\,\frac{\partial^2 u}{\partial x^2}(x,t) \quad \text{for} \quad t \geq 0,\, x \in \mathbb{R}$$

for which $x \rightsquigarrow u(x,t)$ is in $L^2(\mathbb{R})$ for each $t \geq 0$. Note that the requirement $u(\cdot,t) \in L^2(\mathbb{R})$ can be viewed as a boundary condition. The solution u also satisfies initial conditions

$$u(x,0) = \phi_1(x) \quad \text{and} \quad \frac{\partial u}{\partial x}(x,0) = \phi_2(x),$$

where ϕ_1 and ϕ_2 are real-valued functions in $L^2(\mathbb{R})$ and ϕ_1 is continuously differentiable.

Consider a function u of the form $u(x,t) = \frac{1}{2}[\psi_1(ct+x) + \psi_2(x-ct)]$ where ψ_1 and ψ_2 are sufficiently smooth function in $L^2(\mathbb{R})$.

(a) Prove that u satisfies the wave equation and the 'boundary condition'.

(b) Show that

$$\widehat{\psi}_1(\omega) = \widehat{\phi}_1(\omega) + \frac{1}{2\pi i c\omega}\widehat{\phi}_2(\omega) \quad \text{and} \quad \widehat{\psi}_2(\omega) = \widehat{\phi}_1(\omega) - \frac{1}{2\pi i c\omega}\widehat{\phi}_2(\omega).$$

leads to a solution that also satisfies the initial conditions.

(c) Prove that

$$u(x,t) = \int_{-\infty}^{\infty} \widehat{\phi}_1(\omega)\,\cos(2\pi\omega ct)\,e^{2\pi i\omega x}\,\mathrm{d}\omega + \int_{-\infty}^{\infty} \widehat{\phi}_2(\omega)\,\frac{\sin(2\pi\omega ct)}{2\pi\omega c}\,e^{2\pi i\omega x}\,\mathrm{d}\omega.$$

(d) Suppose that both ϕ_1 and ϕ_2 are odd around 0 (i.e., $\phi_i(-y) = -\phi_i(y)$) and odd around 1 (i.e., $\phi(1-y) = -\phi_i(1+y)$). Show that $\psi_1(-y) = -\psi_2(y)$ and $\psi_1(1-y) = -\psi_1(1+y)$ $(y \in \mathbb{R})$.

Show that $u(0,t) = u(1,t) = 0$ for all $t > 0$ (Note that these results are consistent with the ones in §2.12; see also Exercise 2.19).

Exercise 3.13. **The heat equation.**
We are interested in solutions u of the the heat equation

$$\frac{\partial u}{\partial t}(x,t) = \gamma\,\frac{\partial^2 u}{\partial x^2}(x,t) \quad \text{for} \quad t \geq 0,\, x \in \mathbb{R}$$

for which $x \rightsquigarrow u(x,t)$ is in $L^2(\mathbb{R})$ for each $t \geq 0$. The solution u also satisfies the initial condition

$$u(x,0) = \phi(x),$$

where ϕ is a real-valued function in $L^2(\mathbb{R})$ and $\gamma > 0$.

(a) Prove that

$$u(x,t) = \int_{-\infty}^{\infty} \widehat{\phi}(\omega)\, e^{-4\pi^2\omega^2\gamma t}\, e^{2\pi i\omega x}\, \mathrm{d}\omega.$$

Exercise 3.14. **Local mode analysis.**
To asses the stability of a time dependent partial differential equation one often proceed as follows. The differential equation is linearized around the exact solution. Then it is analyzed how the linearized differential equation responds 'locally' to a perturbation by a Fourier 'mode', i.e., by a perturbation of the form $\varepsilon\, e^{2\pi i\omega x}$. The influence of the boundary conditions is discarded.

To illustrate this approach, consider the heat equation

$$\frac{\partial u}{\partial t} = \gamma\frac{\partial^2 u}{\partial x^2}. \tag{78}$$

Note that the heat equation is already linear.

(a) Try to find a solution of the form $e^{\lambda t}\, e^{2\pi i\omega x}$. Show that $\lambda = \lambda(\omega) = -4\pi^2\omega^2\gamma$. Note that $\lambda < 0$ and that $e^{\lambda t} \to 0$ for $t \to \infty$ only if $\gamma > 0$.

Generally (for other PDEs), we would like to see that $\mathrm{Re}(\lambda(\omega)) < 0$ for all $\omega \in \mathbb{R}$.

(b) What is the effect of perturbing the solution u at time $t = t_0$ by $\varepsilon e^{2\pi i\omega x}$?

(c) What is the effect of perturbing the solution at time $t = t_0$ by an $f \in L^2(\mathbb{R})$?

Note that any perturbation in $L^2(\mathbb{R})$ at $t = t_0$ can be written as a superposition of perturbations of the form $\varepsilon e^{2\pi i\omega x}$. If one of the components increases if $t \to \infty$, then the perturbation amplifies and we may fear that the differential equation is unstable.

In this approach, boundary conditions have not been considered. Perturbations may get controlled by imposing boundary conditions. However, the idea is here that, if a perturbations grows strongly, then the perturbed solution already may become completely spoiled even before the perturbation 'hits the boundaries'.

If $\gamma = \gamma(x,t)$ depends on time and on place, then the analysis in (a) is applied to the variant of (78) with '*frozen coefficients*', i.e., to the equation where $\gamma(x,t)$ is replaced by the constant $\gamma(x_0,t_0)$. Here, again the idea is that perturbations introduced in the neighborhood of (x_0,t_0), may get out of control, before they enter a more stable region (i.e., a region with nicer values for $\gamma(x,t)$). Now, λ depends also on x_0 and t_0: $\lambda = \lambda(\omega, x_0, t_0)$, and for stability we would like to see that $\mathrm{Re}(\lambda(\omega,x,t)) < 0$ for all ω, x, t.

The procedure is: linearize around the exact solution, discard boundary conditions, freeze the coefficients and analyze the growth behavior of solutions of the form $e^{\lambda t}\, e^{2\pi i\omega x}$. This '*local mode*' *analysis* is not a thorough stability analysis, but it is an easy way to obtain some insight in the stability of the differential equation.

Local mode analysis may also be applied to discretized partial differential equations.

Discretization of the heat equation (with Euler's method) leads to

$$u(x_j, t_{n+1}) = u(x_k, t_n) + \gamma\frac{\Delta t}{\Delta x^2} D_2 u(x_j, t_n)$$

Here, $x_j \equiv j\Delta x\ \ (j \in \mathbb{Z})$, $t_n \equiv n\Delta t\ \ (n \in \mathbb{N})$, and $D_2 w(x_j) \equiv w(x_j+\Delta x) - 2w(x_j) + w(x_j - \Delta x)$ for functions w defined on the grid $\{x_j\}$.

(d) Use local mode analysis to derive the Courant–Friedrich–Lewy (CFL) stability condition: for stability we need

$$0 < \gamma \frac{\Delta t}{\Delta x^2} \leq \tfrac{1}{2}.$$

Exercise 3.15. **Sampling signals of bounded bandwidth.**

Let $\Omega > 0$. Consider a function $f \in L^2(\mathbb{R})$ with non-trivial amplitudes only at frequencies ω with $|\omega| \leq \Omega$, i.e., $\widehat{f}(\omega) = 0$ if $|\omega| > \Omega$: f *is of bounded bandwidth*.

In this exercise it will be useful to consider the function F on $\mathbb{R}$ that is 2Ω-periodic and that coincides with $\widehat{f}$ for $\omega \in \mathbb{R}$, $|\omega| \leq \Omega$: $F(\omega) = \widehat{f}(\omega)$ if $|\omega| \leq \Omega$.

(a) Prove that $\widehat{f} \in L^1(\mathbb{R})$. Show that f is continuous (or, to be more precise, the function $g(t) \equiv \int \widehat{f}(\omega) \exp(2\pi i\omega t)\ \mathrm{d}\omega\ \ (t \in \mathbb{R})$ is continuous and $\|g - f\|_2 = 0$. Hence, it is no restriction to assume that $f = g$ and that is what we will do).

(b) With $\gamma_k \equiv \frac{1}{2\Omega}\int_{-\Omega}^{\Omega} F(\omega)\, e^{2\pi i\omega\frac{k}{2\Omega}}\ \mathrm{d}\omega$, we have that $F(\omega) = \sum \gamma_k \exp(-2\pi i\omega\frac{k}{2\Omega})$ for all $\omega \in \mathbb{R}$ (why?). Now, with $\Delta t \equiv 1/(2\Omega)$, show that $\gamma_k = \Delta t\, f(k\Delta t)$ and conclude that

$$\widehat{F}(\omega) = \sum_{k=-\infty}^{\infty} \Delta t\, f(k\Delta t)\, e^{-2\pi i\omega k\Delta t} \quad (\omega \in \mathbb{R}), \quad \text{and} \quad \widehat{f} = F\,\Pi_\Omega.$$

In particular, we have that $\widehat{f}(\omega) = \sum \Delta t\, f(k\Delta t)\, \exp(-2\pi i\omega k\Delta t)$ if $|\omega| \leq \Omega$.

Show that $(f(k\Delta t))$ is in $\ell^2(\mathbb{Z})$.

(c) Now, suppose that, for an $\ell \in \mathbb{N}$, $f \in L^2(\mathbb{R})$ is such that $f(\omega) = 0$ if $\omega \notin \mathcal{O}$, where $\mathcal{O} \equiv \{\omega \,\big|\, \ell\,\Omega \leq |\omega| < (\ell+1)\Omega\}$.

Show that there is a 2Ω-periodic function F such that $\widehat{f} = F\,\chi_{\mathcal{O}}$. Here, $\chi_{\mathcal{O}}(\omega) \equiv 1$ if $\omega \in \mathcal{O}$ and $\chi_{\mathcal{O}} \equiv 0$ elsewhere. Prove that $F(\omega) = \sum \Delta t f(k\Delta t)\, \exp(-2\pi i\omega k\Delta t)\ \ (\omega \in \mathbb{R})$.

Exercise 3.16. **Fourier transform of discrete measures.**

Let δ_t be the point measure or Dirac delta function at t: $\int \delta_t(s)\, g(s)\ \mathrm{d}s = g(t)$ for continuous functions g. We have that $\widehat{\delta_t}(\omega) = \exp(-2\pi i t\omega)$.

Let (γ_k) be a sequence in $\ell^1(\mathbb{Z})$ and let (t_k) be a sequence in $\mathbb{R}$. Consider the *discrete measure* μ defined by $\mu \equiv \sum \gamma_k\, \delta_{t_k}$: $\int g(x)\ \mathrm{d}\mu(x) = \sum \gamma_k g(t_k)$ ($g \in C(\mathbb{R})$ bounded).

(a) Show that $\widehat{\mu}(\omega) \equiv \sum \gamma_k \exp(-2\pi i t_k\omega)\ \ (\omega \in \mathbb{R})$ is bounded and continuous.

(b) Show that $\widehat{\mu}$ is 2Ω-periodic if $t_k \equiv k\Delta t$ for all $k \in \mathbb{Z}$ and $\Delta t \equiv 1/(2\Omega)$. Here $\Omega > 0$.

Note that this allows an extension to the case where $(\gamma_k) \in \ell^2(\mathbb{Z})$.

(c) Consider a function $f \in L^2(\mathbb{R})$ for which $\widehat{f}(\omega) = 0$ if $|\omega| > \Omega$ and let F be 2Ω-periodic such that $F(\omega) = \widehat{f}(\omega)$ for all ω, $|\omega| \leq \Omega$.

Show that $\widehat{\mu} = F$ and $\widehat{f} = \widehat{\mu}\,\Pi_\Omega$ if $\mu \equiv \sum \Delta t f(k\Delta t)\delta_{k\Delta t}$.

Exercise 3.17. **Weak convergence to the Dirac delta function.**

For $T > 0$, let

$$f_T(t) \equiv 2\,T\,\mathrm{sinc}\,(2\,Tt) = \frac{\sin(2\pi Tt)}{\pi t} \qquad (t \in \mathbb{R}).$$

Consider a $g \in C(\mathbb{R})$ with bounded support (i.e., $g(x) = 0$ for $|x|$ large).

(a) In this part, we only need that $g \in L^1(\mathbb{R})$. Prove that, for $\delta > 0$,

$$\lim_{T\to\infty} \int_{|t|\geq\delta} f_T(t)\, g(t)\ \mathrm{d}t = 0 \quad \text{for all} \quad g \in L^1(\mathbb{R}).$$

(Hint: Consider the function G defined by $G(t) \equiv g(t)/(\pi t)$ if $|t| \geq \delta$ and $G(t) \equiv 0$ if $|t| < \delta$. Then $G \in L^1(\mathbb{R})$ and $\int_{|t|\geq\delta} f_T(t)\, g(t)\; \mathrm{d}t = \int_{-\infty}^{\infty} G(t)\, \sin(2\pi T t)\; \mathrm{d}t = \frac{1}{2i}\left(\widehat{G}(-T) - \widehat{G}(T)\right) \to 0$ $(T \to \infty)$. Why?)

(b) Use the fact that $\int_0^\infty \frac{\sin(t)}{t}\; \mathrm{d}t = \frac{1}{2}\pi$ to show that for each $\delta > 0$,

$$\int_{-\delta}^{+\delta} f_T(t)\; \mathrm{d}t \to 1 \quad \text{for} \quad T \to \infty.$$

(c) Prove that f_T is even and

$$\left|\int_0^a f_T(t)\; \mathrm{d}t\right| \leq 1 \qquad \text{for all} \;\; a > 0.$$

(Hint: Consider $F_T(t) \equiv \int_0^t f_T(s)\; \mathrm{d}s \;\; (s \in \mathbb{R})$, inspect the graph of f_T, and show that $\|F_T\|_\infty \leq F_T(\frac{1}{2T}) \leq f_T(0)\frac{1}{2T} = 1$.) Show that $f \notin L^1(\mathbb{R})$.

(d) Suppose that, in addition, g' exists on $(-a, a)$ for some $a > \delta > 0$ and is in $L^1(-a, +a)$. Then, $|g(t) - g(0)| \leq \int_0^\delta |g'(s)|\; \mathrm{d}s$ for $t \geq \delta$. Prove that

$$\left|\int_{-\delta}^{+\delta} f_T(t)\, g(t)\; \mathrm{d}t - g(0)\right| \leq 2\int_{-\delta}^{\delta} |g'(s)|\; \mathrm{d}s \qquad \text{for large } T.$$

(e) Prove that

$$\lim_{T\to\infty} \int_{-\infty}^{+\infty} f_T(t)\, g(t)\; \mathrm{d}t = g(0) \qquad \text{if} \qquad g \in C^{(1)}(\mathbb{R}) \cap L^1(\mathbb{R}). \tag{79}$$

Note. In (d) we used integration by part for functions g that are absolutely continuous (locally at 0). The second mean value theorem for integral calculus gives a similar result: $\left|\int_{-\delta}^{+\delta} f_T(t)\, g(t)\; \mathrm{d}t - g(0)\right| \leq 2|g(\delta) - g(-\delta)|$ for continuous functions g that are locally non-decreasing. Therefore, for (79) it suffices to require that the continuous function g with bounded support is of bounded variation on some neighborhood of 0.

Exercise 3.18. **The Dirac comb.**

In the previous exercise, we defined the Fourier transform for discrete measure, 'summable' linear combinations of Dirac delta functions, even though these objects are not in $L^1(\mathbb{R})$ or $L^2(\mathbb{R})$ (even worse; they are no functions). In this exercise, we push the idea even further and define the Fourier transform for

$$Ш_{\Delta t} \equiv \sum_{k\in\mathbb{Z}} \delta_{t_k}, \tag{80}$$

where $t_k \equiv k\Delta t$ for some step size $\Delta t > 0 \;\; k \in \mathbb{Z}$). This 'operator' is sometimes called the *Dirac comb* by physicists (or *shah function*. 'Shah' is the name of the Cyrillic symbol ш that is used to denote the function): if the Dirac delta function can graphically be represented as an infinitely large spike, then the Dirac comb can graphically be represented as infinite sequence of infinitely large spikes. The Dirac comb simplifies certain formulas in the theory of sampling signals (see, for instance, Exercise 6.9).

(a) For $N \in \mathbb{N}$, consider the discrete measure $\mu_N \equiv \sum_{|k|\leq N} \delta_{t_k}$. Note that $\widehat{\mu}_N$ is defined in Exercise 3.16. Show that $\widehat{\mu}_N$ is a scaled version of the Nth Dirichlet kernel (see Exercise 2.20, Eq. (52) and (54)).

(b) With $\Omega \equiv \frac{1}{2\Delta t}$, consider the functions $h_N(\omega) \equiv \widehat{\mu}_N \Pi_\Omega \;\; (N \in \mathbb{N})$. Argue that, for each continuously differentiable function g, the sequence $(\int_{-\infty}^{\infty} g(\omega) h_N(\omega)\; \mathrm{d}\omega)$ converges to $g(0)$ $(N \to \infty)$. (Hint: adapt the arguments in Exercise 3.17.)

(c) Argue that the Fourier transform of the Dirac comb ш with step size Δt is the Dirac comb $Ш_{1/\Delta t}$:

$$\widehat{Ш}_{\Delta t} = Ш_{\frac{1}{\Delta t}}. \tag{81}$$

4 Fourier integrals in more dimensions

In the preceding sections, we interpreted the variable t as time. However, in applications, it can also be a space variable. In those situations, more than one variable will usually play a role. It is easy to generalize the theory for one dimension to a theory for more dimensions. First, we introduce some notation.

We put $\vec{x} = (x_1, x_2, \ldots, x_d)^{\mathrm{T}}$ for the space variable in a d-dimensional space $\mathbb{R}^d$. Here, $d = 2$ or $d = 3$, but d can also be larger. The frequency vector $\vec{\omega}$ will also be d-dimensional. We will write $\vec{\omega} = (\omega_1, \omega_2, \ldots, \omega_d)^{\mathrm{T}} \in \mathbb{R}^d$.[14] The multiplication between a space vector and a frequency vector is the inner product $(\vec{x}, \vec{\omega})$ between these vectors defined by

$$(\vec{x}, \vec{\omega}) \equiv \vec{\omega}^{\mathrm{T}} \vec{x} = x_1\, \omega_1 + x_2\, \omega_2 + \ldots + x_d\, \omega_d.$$

If f is real-valued on $\mathbb{R}^d$ and square integrable, $f \in L^2(\mathbb{R}^d)$, i.e.,

$$\|f\|_2^2 \equiv \int \!\ldots\! \int |f(x_1, \ldots, x_d)|^2 \;\mathrm{d}x_1 \;\mathrm{d}x_2 \ldots \;\mathrm{d}x_d = \int \!\ldots\! \int |f(\vec{x})|^2 \;\mathrm{d}\vec{x} < \infty,$$

then the Fourier integral $\widehat{f}(\vec{\omega})$ can be defined:

$$\widehat{f}(\vec{\omega}) \equiv \int \!\ldots\! \int f(\vec{x})\, e^{-2\pi i (\vec{x}, \vec{\omega})} \;\mathrm{d}\vec{x} \qquad (\vec{\omega} \in \mathbb{R}^d). \tag{82}$$

The integrals range here, as elsewhere in this section (unless stated differently), from $-\infty$ to $+\infty$.

As in §3, we first should define Fourier integrals for functions in $L^1(\mathbb{R}^d)$ and then introduce the Fourier integral for a function in $L^2(\mathbb{R}^d)$ as the 2-norm limit of a sequence of functions that are both in L^1 as well as in L^2. We leave the details of such a formal introduction to the interested reader and concentrate on the situation where $f \in L^2(\mathbb{R}^d)$.

It can be shown that

$$f(\vec{x}) = \int \!\ldots\! \int \widehat{f}(\vec{\omega})\, e^{2\pi i (\vec{x}, \vec{\omega})} \;\mathrm{d}\vec{\omega} \qquad (\vec{x} \in \mathbb{R}^d). \tag{83}$$

4.A Application: diffraction

4.1 Huygens' principle. In physics, *diffraction* is a wave phenomenon: the apparent bending and spreading of waves when they meet an obstruction. Diffraction occurs with electromagnetic waves, such as light and radio waves, and also in sound waves and water waves. In diffraction theory, *Huygens' principle* plays a central role: *every point on a wavefront which comes from a source can itself be regarded as a (secondary) source*. All the wavefronts from all these secondary sources combine and interfere to form a new wavefront.

Here, we consider the case of far-field or *Fraunhofer diffraction*, where the diffracting obstruction is many wavelengths distant from the point at which the wave is measured. The more general case is known as near-field or *Fresnel diffraction*, and involves more complex mathematics. Fraunhofer diffraction is commonly observed in nature.

[14] Higher dimensional Fourier techniques play a crucial role in analyzing the behavior of waves. In these application, $\vec{\omega}$ is the *wave* vector and is usually denoted by $\vec{k} = (k_1, \ldots, k_d)$ instead of $\vec{\omega}$.

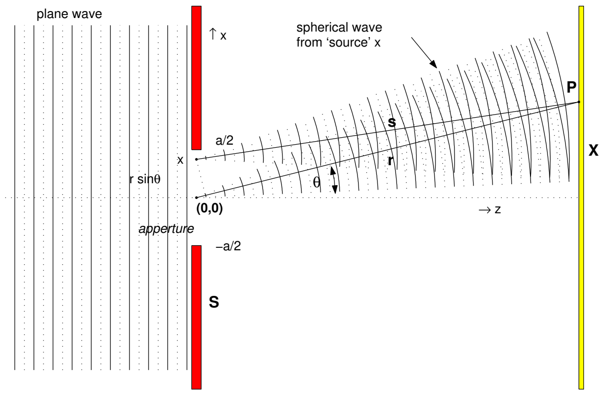


FIGURE 11. The picture shows a two-dimensional intersection of a screen S with a long narrow rectangular slit of width a. The intersection is perpendicular to the long direction of the slit. Plane waves emitted from a mono-chromatic light source located at $-\infty$ along the horizontal axis approach the screen S. The waves are diffracted at the aperture. According to Huygens' principle, points in the aperture can be viewed as sources that emit (spherical) waves. These 'secondary' waves form an interference pattern at the screen through P parallel to S: the distance s from P at (X, Z) to $(x, 0)$ differs from the distance r from P to the origin $(0, 0)$. Depending on the wavelength λ of the light and the difference $r - s$ in distance, the waves emitted from $(0, 0)$ and from $(x, 0)$ may amplify each other or may cancel out at P.

4.2 Fraunhofer diffraction. Our obstruction is a screen S with a small transparent part on which we shine light. For ease of exposition we restrict ourselves here to a two-dimensional situation, but we urge the reader to adapt the formulas below for three dimensions. We consider a situation in which the y-coordinate can be discarded. We assume that the aperture that bounds the transparent part of S is unbounded, rectangular, and perpendicular to the plane of the diagram in Fig. 11. In the plane of the diagram we use Cartesian coordinates: x for the vertical direction, z for the horizontal one (here, we follow traditional notation). The surface S is along the x- and the y-axis at $z = 0$. The y-coordinate can be discarded here (Why?). Furthermore, we assume that the width a of the aperture is small compared to the distance r from the origin $(0, 0)$ to a point of interest P. We also assume that the light source at L is located on the negative z-axis with a distance from $(0, 0)$ that is so large that a mono-chromatic electric field E (light is composed of electromagnetic waves)[15] emitted from the source at L has the same magnitude and a constant phase at all points on the transparent part of S (i.e., we assume that L is at $-\infty$ along the z axis, E is a plane wave).[16] The field E at time t and location (x, z) with $z \le 0$ is given by $E = E_0 \exp(2\pi i(kz - \omega t))$.

4.3 For ease of notation, we will assume in the exposition below that $t = 0$.

[15] An electromagnetic wave is a combination of oscillating electric and magnetic fields in perpendicular orientation to each other, moving through space. The direction of oscillation is perpendicular to the wave's direction of travel.

[16] The electric field E is a *plane wave* if it is of the form $\vec{E}_0 \exp(2\pi i[(\vec{k}, \vec{x}) - \omega t])$, where, with $\vec{k} \equiv (k_1, k_2, k_3)$ and $\vec{x} \equiv (x, y, z)$, $(\vec{k}, \vec{x}) \equiv k_1 x + k_2 y + k_3 z$. $E_0 \equiv \|\vec{E}_0\|_2$ is the *amplitude*. The wave travels in the direction of the *wave vector* $\vec{k}$. $k \equiv \|\vec{k}\|_2$, or actually $2\pi k$, is the *wave number*, $\lambda \equiv 1/k$ is the *wavelength*, and with c the speed of light, $\omega = \frac{c}{\lambda}$ is the *frequency*. As mentioned above, the vector $\vec{E}_0$ is perpendicular to $\vec{k} = (k_1, k_2, k_3)$: electromagnetic waves are *transverse*, in contrast to, for instance, sound waves that vibrate in the direction of travel. They are *longitudinal*: $\vec{E}_0$ is a multiple of $\vec{k}$.

Here, we restrict ourselves to the (x, z) plane (we omit y), and we assume that the wave travels along the z-axis, i.e., $\vec{k} = (0, k_3)$. We write k instead of k_3 and we write E_0 instead of $\vec{E}_0$, thus omitting the direction of E_0.

The field at $(x,0)$ is E_0. Then, according to Huygens' principle, the field at $P=(X,Z)$ induced from a strip of width $\mathrm{d}x$ at $(x,0)$ will be

$$\Gamma\, A(x)\, E_0\, e^{2\pi i k s}\ \mathrm{d}x,$$

where s is the distance from $(x,0)$ to P, Γ is some proportionality factor that depends on the wavelength and Z,[17] and $A(x)$ is the aperture function, which describes the transparent and opaque parts of the screen S. For instance, A is the top-hat function with width a in case of one slit and A is the sum of two shifted top-hat functions in case of two slits.

Huygens' principle is applicable if both X and the width a of the aperture are small relative to Z and that is what we will assume here.

If r is the distance from $(0,0)$ to P, and ϑ is the angle between the line from P to $(0,0)$ and the z-axis, then $P=(X,Z)=(r\sin\vartheta, r\cos\vartheta)$. Note that $\sin\vartheta\approx X/Z$. Hence, if we use that $|x|\ll r$, then $s\approx r-x\sin(\vartheta)$ and we find that the field at P is given by

$$\Gamma\, E_0\, e^{2\pi i k r}\int A(x)\, e^{2\pi i k x\sin(\vartheta)}\ \mathrm{d}x=\Gamma\, E_0\, e^{2\pi i k r}\,\widehat{A}(k\sin(\vartheta)). \tag{84}$$

Therefore, the reduction at P of the strength of the original field is proportional to the Fourier transform of the aperture function. The *intensity* $I(P)$ of the diffracted waves at P is the square of the absolute value of the expression in (84). Hence,

$$I(P)=I_0\,(\widehat{A}(k\sin(\vartheta))^2,\quad \text{where}\quad I_0\equiv|\Gamma\, E_0|^2.$$

The intensity of the wave is what we measure/observe, or, on other words, the absolute value of the Fourier transform can be measured. There is no information on the phase. This fact is known as *the missing phase problem* in crystallography.

To obtain the formula's for the more dimensional situation, note that, if we put $\vec{x}\equiv(x,0)$ and $\vec{p}\equiv\frac{1}{r}(X,Z)$, then $(\vec{x},\vec{p})=x\sin(\vartheta)$.

4.4 Single-slit diffraction. In case of a single slit, A is the top-hat function $\Pi_{a/2}$ and

$$I(P)=I_0\,a^2\mathrm{sinc}^2(a\,\omega),\quad \text{where}\quad \omega\equiv\frac{\sin\vartheta}{\lambda}.$$

If the aperture is circular (in the (x,y)-plane), the pattern is similar to a radially symmetric version of this equation, representing a series of concentric rings surrounding a so-called central *Airy* disc.

4.5 Diffraction gradings. The aperture is a *diffraction grading* if it consists of N parallel rectangular slits of width a at equal distance d from each other, with $N\in\mathbb{N}$ large. Then (see Exercise 4.5) A is a sum of N shifted top-hat functions and

$$I(P)=I_0\left(a\,\mathrm{sinc}(a\,\omega)\,\frac{\sin(N\pi\, d\,\omega)}{\sin(\pi\, d\,\omega)}\right)^2,\quad \text{where}\quad \omega\equiv\frac{\sin\vartheta}{\lambda}. \tag{85}$$

If $a\ll d$, then the light intensity takes locally maximal values at the angles ϑ that satisfy

$$d\,\sin\vartheta=\ell\,\lambda\qquad(\ell\in\mathbb{Z}). \tag{86}$$

[17] To be precise, $\Gamma=-i/(\lambda Z)$ in three dimensions and $\Gamma=\sqrt{1/(i\lambda Z)}$ in two dimensions. This precise expression for Γ is not part of Huygens' principle. The secondary sources in the Huygens' principle produce spherical waves $\Gamma'\frac{1}{r}\exp(2\pi i r)$ in 3-d and cylindrical waves $\Gamma''\frac{1}{\sqrt{r}}\exp(2\pi i r)$ in 2-d; r is the distance to the secondary source, Γ' and Γ'' are appropriate constants.

This relation is known in crystallography as *Bragg's law*.[18]

The most common demonstration of Bragg diffraction is the spectrum of colors seen reflected from a compact disc: the closely-spaced tracks on the surface of the disc form a diffraction grading, and the individual wavelengths of white light are diffracted at different angles from it, in accordance with Bragg's law.

Exercises

Exercise 4.1. Suppose $f \in L^2(\mathbb{R}^2)$ is of the form $f(x,y) = f_1(x)f_2(y) \quad (x, y \in \mathbb{R})$. Derive an expression for $\widehat{f}$ in terms of $\widehat{f_1}$ and $\widehat{f_2}$.

Exercise 4.2.

(a) Formulate the Fourier transform (82) in polar coordinates for $d = 2$.

(b) Compute the Fourier transform $\widehat{f}$ of the the 2-dimensional top-hat function:

$$f(x,y) \equiv 1 \text{ if } |x|^2 + |y|^2 \leq 1 \quad \text{and} \quad f(x,y) = 0 \text{ elsewhere.}$$

Exercise 4.3. **The wave equation.**

Consider the three dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2}(\vec{x}, t) = c\,\Delta u(\vec{x}, t) \quad \text{for} \quad t \geq 0,\, \vec{x} \in \mathbb{R}^3,$$

where the *Laplace operator* Δ acts on the space variables $\vec{x}$: $\Delta f \equiv \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \frac{\partial^2 f}{\partial x_3^2}$ for sufficiently smooth real-valued functions f with domain in $\mathbb{R}^3$. c is a positive constant.

(a) Let $\vec{k} \in \mathbb{R}^3$. Show that

$$u(\vec{x}, t) = e^{2\pi i[(\vec{k},\vec{x}) - \omega t]} \tag{87}$$

is a solution if $\omega \in \mathbb{R}$ is such that $|\omega| = c\kappa$, where, $\kappa \equiv \|\vec{k}\|_2 = \sqrt{k_1^2 + k_2^2 + k_3^2}$.

(b) The real part of u in (87) has maxima at the $(\vec{x}, t)$ for which $(\vec{k}, \vec{x}) - \omega t = \ell$ for some $\ell \in \mathbb{Z}$. For a fixed time t, the ℓth maximum is at $\mathcal{P}_\ell \equiv \{\vec{x} \,\big|\, (\vec{k}, \vec{x}) = \ell + \omega t\}$. Note that $\mathcal{P}_\ell$ is a plane orthogonal to $\vec{k}$ (if $\vec{x} \in \mathcal{P}_\ell$ and $\vec{y} \perp \vec{k}$ then $\vec{x} + \vec{y} \in \mathcal{P}_\ell$). For this reason, waves of the form (87) are called *plane waves*; $\vec{k}$ is the *wave vector* and κ (or actually $2\pi\kappa$) is the *wave number* (in many texts, waves of the form $\exp((\vec{k}, \vec{x}) - \omega t)$ are considered: then the factor 2π is included in the wave vector and in the frequency). Show that $\mathcal{P}_0 = \{ct\vec{\zeta} + \vec{y} \,\big|\, \vec{y} \perp \vec{\zeta}\}$, where $\vec{\zeta}$ is the direction of $\vec{k}$: $\vec{\zeta} \equiv \frac{1}{\kappa}\vec{k}$, $\|\vec{\zeta}\|_2 = 1$. Observe that these planes move in the direction $\vec{k}$ at speed c. Show that the distance between two neighboring planes is $\lambda \equiv 1/\kappa$: λ is the *wave length.*

Now, fix an $\vec{x}$. Show that time between two consecutive tops at $\vec{x}$ is $1/|\omega|$: ω is the *frequency* of the wave. Note that $\omega = c/\lambda$.

(c) Prove that

$$u(\vec{x}, t) = \iiint \widehat{\phi}_1(\vec{k})\, \cos(2\pi\omega t)\, e^{2\pi i(\vec{k},\vec{x})}\, \mathrm{d}\vec{k} + \iiint \widehat{\phi}_2(\vec{k})\, \frac{\sin(2\pi\omega t)}{2\pi\omega}\, e^{2\pi i(\vec{k},\vec{x})}\, \mathrm{d}\vec{k} \tag{88}$$

[18] From `en.wikipedia.org`: Diffraction from multiple slits, as described above, is similar to what occurs when waves are scattered from a periodic structure, such as atoms in a crystal. Each scattering center (e.g., each atom) acts as a point source of spherical wavefronts; these wavefronts undergo constructive interference to form a number of diffracted beams. The direction of these beams is described by Bragg's law. Now, d is the distance between scattering centers, ϑ is the angle of diffraction and ℓ is an integer known as the order of the diffracted beam. Bragg diffraction is used in X-ray crystallography to deduce the structure of a crystal from the angles at which X-rays are diffracted from it.

is a solution in $L^2(\mathbb{R}^3)$ of the wave equation that satisfies the initial conditions $u(x,0) = \phi_1$ and $\frac{\partial u}{\partial t}(x,0) = \phi_2(x)$, where ϕ_i are sufficiently smooth functions in $L^2(\mathbb{R}^3)$. Here, ω depends on $\vec{k}$ according to $\omega = c\kappa$, with $\kappa = \|\vec{k}\|_2$.

(d) Show that there are $L^2(\mathbb{R}^3)$ functions ψ_1 and ψ_2 such that

$$u(\vec{x},t) = \iiint \widehat{\psi}_1(\vec{k})\, e^{2\pi i[(\vec{k},\vec{x})+\omega t]}\, \mathrm{d}\vec{k} + \iiint \widehat{\psi}_2(\vec{k})\, e^{2\pi i[(\vec{k},\vec{x})-\omega t]}\, \mathrm{d}\vec{k}.$$

Again, $\omega = c\kappa = c\|\vec{k}\|_2$. Give an expression for $\widehat{\psi}_1$ and $\widehat{\psi}_2$ in terms of $\widehat{\phi}_i$.

Apparently, u can be viewed as a superposition of harmonic plane waves.

(e) In the one dimensional case, we could represent the solution as a sum of two waves, of which one travels to the left and the other travels to the right (ψ_1, ψ_2, respectively in Exercise 3.12). Do we have a similar representation in this three dimensional case?

For each direction vector $\vec{\zeta}$ (i.e., $\vec{\zeta} \in \mathbb{R}^3$ such that $\|\vec{\zeta}\|_2 = 1$), define

$$\Psi_i(\vec{\zeta}, s) \equiv \int_{-\infty}^{\infty} \kappa^2\, \widehat{\psi}_i(\kappa\,\vec{\zeta}\,)\, e^{2\pi i\kappa s}\, \mathrm{d}\kappa \quad (s \in \mathbb{R},\ i = 1, 2).$$

Show that

$$u(\vec{x},t) = \iint_B \left[\Psi_1(\vec{\zeta}, (\vec{\zeta},\vec{x}) + ct) + \Psi_2(\vec{\zeta}, (\vec{\zeta},\vec{x}) - ct)\right] \mathrm{d}\vec{\zeta},$$

where we integrate over the semi-ball $B \equiv \{\vec{\zeta} \,\big|\, \|\vec{\zeta}\|_2 = 1, \zeta_1 \geq 0\}$. Where does the term κ^2 come from in the definition of Ψ_i?

It appears that, for each direction $\vec{\zeta}$, a wave that moves to the right and one that moves to the left play a role. To obtain the solution u, we have to integrate over all directions $\vec{\zeta}$.

NOTE. *In reality, waves are vector valued. For instance, they may describe the displacement of molecules in air by a change in air pressure (acoustic wave). Then $\vec{u}(\vec{x},t)$ is the displacement vector: at time t, it describes the dispacement of the molecule that originally was located at position $\vec{x}$. For vector-valued waves, harmonic plane waves are of the form $\vec{a}\exp(2\pi i[(\vec{k},\vec{x})-\omega t])$, where $\vec{a}$ is a constant vector in $\mathbb{R}^3$. In case of an acoustic wave, $\vec{a}$ is a multiple of $\vec{k}$ (the pressure changes in the direction in which the wave travels). In case of an electromagnetic wave, both the $\vec{a}_e$ for the electric component and the $\vec{a}_b$ for the magnetic component are orthogonal to $\vec{k}$; in addition, $\vec{a}_e$ is orthogonal to $\vec{a}_b$.*

For vector-valued waves, the above analysis applies coordinate wise.

Exercise 4.4. Wavepacket, group velocity.

The function

$$(\tau, x) \rightsquigarrow e^{2\pi i(\mu\tau - kx)}$$

is a(n harmonic) wave with *frequency* μ and *wavenumber* k that travels with *speed* $c \equiv \mu/k$. The quantity $1/k$ is the *wavelength*.

(a) Explain the notions ‘frequency’, ‘speed’, and ‘wavelength’.

Consider a function f in $L^2(\mathbb{R})$ with frequencies concentrated around some frequency Ω, i.e., $\widehat{f}(\omega) = \widehat{f}_0(\omega - \Omega)$ with $\widehat{f}_0(\rho) = 0$ (or ≈ 0) if $|\rho| > \varepsilon$ for some small positive ε.

If x is a space variable, and τ the time variable, then $(\tau, x) \rightsquigarrow f(c\tau - x)$, with c some positive scalar, represents a *group of waves* or *wavepacket* that travels with speed c. Why can c be viewed as speed?

(b) Prove that $f(c\tau - x)$ can be written as

$$g(\tau, x) \equiv f(c\tau - x) = \int_{-\infty}^{\infty} \widehat{f}_0(\rho)\, e^{2\pi i(\Omega+\rho)(c\tau - x)}\, \mathrm{d}\rho = f_0(c\tau - x)e^{2\pi i(\Omega c\tau - \Omega x)}.$$

This expression reveals that $f_0(c\tau - x)$ can be viewed as a modulation of the wave with frequency $\mu = \Omega c$ and wavenumber $k = \Omega$ and explains the term ‘wavepacket’. The function f_0 is the *enveloppe* of the wavepacket g.

Now, suppose that the speed depends on the frequency (as is often the case in practical situation), $c = c(\omega)$. We are interested in the development in time of the group of waves that, at x, is equal to $f(-x)$ at time $\tau = 0$.

(c) Explain why this group of waves is described by

$$g(\tau, x) = \int_{-\infty}^{\infty} \widehat{f}_0(\rho)\, e^{2\pi i(\Omega+\rho)(c(\Omega+\rho)\tau - x)}\ \mathrm{d}\rho$$

(d) Put $c_0 \equiv c(\Omega)$ and $c_0' \equiv c'(\Omega)$. Show that

$$(\Omega+\rho)(c(\Omega+\rho)\tau - x) = \Omega(c_0\tau - x) + \rho([c_0 + \Omega c_0']\tau - x) + \mathcal{O}(|\rho|^2) \quad (|\rho| \to 0)$$

and conclude that

$$g(\tau, x) \approx f_0([c_0 + \Omega c_0']\tau - x)\, e^{2\pi i(\Omega c_0 \tau - \Omega x)}.$$

Note that the enveloppe f_0 of the wavepacket travels at a different speed as the wave $e^{2\pi i(c_0\tau - x)}$: $c(\Omega)$ is the *phase velocity* and $c(\Omega) + \Omega c'(\Omega)$ is the *group velocity*. Nevertheless, the wavepacket hardly desintegrates in time: the shape of the wavepacket, the enveloppe, is (approximately) preserved.

Suppose we travel with the wavepacket. How will we experience the differences in speed?

(e) Note that the frequency $\mu \equiv \omega c$ depends on the wavenumber according to $\mu(\omega) = \omega c(\omega)$. The function that expresses μ in terms of the wavenumber is called the *dispersion relation*. Show that the group velocity is given by $\mu'(\Omega)$ and the phase velocity by $\mu/\Omega \quad (= \frac{\mu(\Omega)-\mu(0)}{\Omega})$.

If the spacial domain is d-dimensional, then the wavenumber is a d-vector $\vec{k} = (k_1, \ldots, k_d)^{\mathrm{T}}$ and if the frequency μ depends on $\vec{k}$ then the group velocity is the vector

$$(\frac{\partial\mu}{\partial k_1}, \ldots, \frac{\partial\mu}{\partial k_d})^{\mathrm{T}}$$

(f) Describe the notion of wavepacket on a d-dimensional spatial domain and interpret the notion of group velocity.

Exercise 4.5. **Diffraction gradings.**

Suppose the aperture in §4.A is a diffraction grading consisting of N parallel rectangular slits of width a at equal distance d from each other.

(a) Use the notation of §4.A and prove that the intensity $I(P)$ at P of mono-chromatic light of wavelength λ is as in (85).
(Hint: use the result of (c) of Exercise 3.4: $N = 2K + 1$.)

(b) Assume that $a \ll d$. Prove (86), Bragg's law.

(c) For wavelength $\lambda_1 \equiv \lambda$, there is a locally maximal light intensity at the angle ϑ_1 for which $\vartheta_1 \approx \sin\vartheta_1 = \ell\frac{\lambda_1}{d}$. Determine the smallest $\Delta\lambda$ for which $\lambda_2 \equiv \lambda + \Delta\lambda$ has zero light intensity at angle ϑ_1: $\Delta\lambda$ is the theoretical resolution of the grading (the difference between wavelengths that can produce separate images).

Exercise 4.6. Consider the situation of §4.A where the screen S is again in the (x, y)-plane, but where the aperture is bounded now (with a diameter that is small compared to the distance from S to P). The aperture function $A(x, y)$ is two-dimensional now.

(a) To describe the field E and the intensity $I(P)$ at P we need expressions involving $A(x, y)$. Give these expressions.

(b) Compute $I(P)$ in case the aperture is circular: A is the two-dimensional top-hat function.

5 Discrete Fourier transforms

5.1 If f is T-periodic and sufficiently smooth, then we have that (cf. §§2.2 and 2.4)

$$f(t)=\sum_{k=-\infty}^{\infty}\gamma_k(f)\,e^{2\pi i t\frac{k}{T}},\quad \text{where}\quad \gamma_k(f)\equiv\frac{1}{T}\int_0^T f(t)\,e^{-2\pi i t\frac{k}{T}}\,\mathrm{d}t. \tag{89}$$

Suppose that the function values of f are sampled at $t=n\Delta t$ with $\Delta t=\frac{T}{N}$ and that only the sample values $f_n\equiv f(n\Delta t)$, for $n=0,\ldots,N-1$, are available. Then it seems reasonable to approximate $\gamma_k(f)$ by a Riemann sum, that is, by $\widetilde{\gamma}_k$, where

$$\widetilde{\gamma}_k\equiv\frac{1}{T}\sum_{n=0}^{N-1}\Delta t\,f(n\Delta t)\,e^{-2\pi i\,n\Delta t\,\frac{k}{T}}=\frac{1}{N}\sum_{n=0}^{N-1}f_n\,e^{-2\pi i\frac{nk}{N}}. \tag{90}$$

We would like to know how accurate the approximations are.

Note that the oscillations $t\rightsquigarrow\exp(2\pi i t\frac{k}{T})$ and $t\rightsquigarrow\exp(2\pi i t\frac{k+N}{T})$ coincide on the sample points $n\Delta t$. This phenomenon is known as *aliasing*. Two implications are important for our discussion here. The first one is slightly disappointing: $\widetilde{\gamma}_{k+N}=\widetilde{\gamma}_k$ for all $k\in\mathbb{Z}$. Hence, $\widetilde{\gamma}_k$ can not form a good approximation to $\gamma_k(f)$ for all $k\in\mathbb{Z}$. The relation that follows from the first equality in (89),

$$f(n\Delta t)=\sum_{k=0}^{N-1}\mu_k(f)e^{2\pi i\frac{nk}{N}},\quad \text{where}\quad \mu_k(f)\equiv\sum_{j=-\infty}^{\infty}\gamma_{k+jN}(f), \tag{91}$$

is another implication. In combination with the theorem below, this implies that

$$\widetilde{\gamma}_k=\mu_k(f),\qquad (k\in\mathbb{Z}) \tag{92}$$

and estimates for the size of the $|\gamma_{k+jN}(f)|$ lead to estimates of the error in the approximation $\widetilde{\gamma}_k$ of $\gamma_k(f)$, because

$$|\widetilde{\gamma}_k-\gamma_k(f)|\leq\sum_{j\neq 0}|\gamma_{k+jN}(f)|\qquad (k\in\mathbb{Z}). \tag{93}$$

For instance, for $|k|\leq\frac{1}{2}N$, the upper bound can be small, particularly if f is smooth (for an illustration, see Fig. 12).

The following theorem is not only important for proving our claim (92), but it plays a central role in many computations (in 'digital Fourier' techniques)

5.2 Inversion theorem for discrete Fourier transform (DFT).

$$\widetilde{\gamma}_k\equiv\frac{1}{N}\sum_{n=0}^{N-1}f_n\,e^{-2\pi i\frac{nk}{N}}\quad\Rightarrow\quad f_n=\sum_{k=0}^{N-1}\widetilde{\gamma}_k\,e^{2\pi i\frac{nk}{N}}. \tag{94}$$

Proof. As in 2.7, our proof relies on orthogonal bases.

Let ℓ_N be the space of sequences $f=(f(0),\ldots,f(N-1))$ of N complex numbers, or, equivalently, of complex-valued functions f on $\{0,1,\ldots,N-1\}$. Note that, for ease of notation, we write $f(n)$ in this proof instead of f_n.

Note that $<f,g>\equiv\sum f(k)\,\overline{g(k)}$ defines an inner product on ℓ_N. Here and in the rest of the proof, we sum over $k=0,1,\ldots,N-1$.

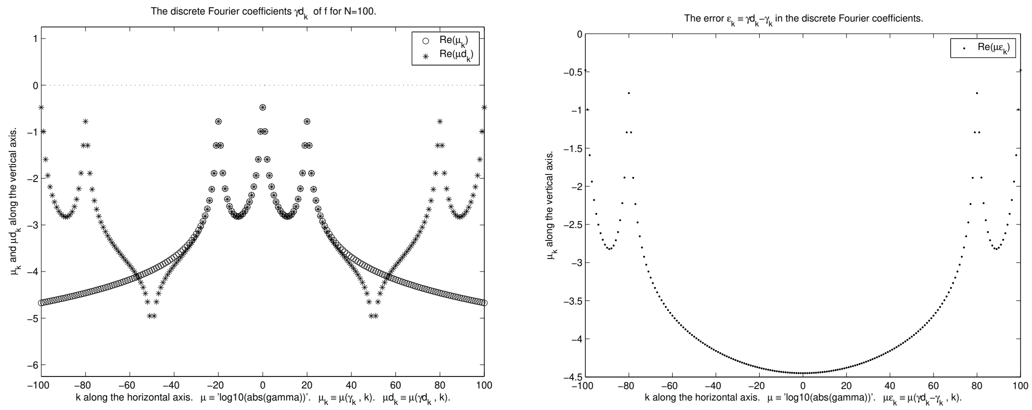


FIGURE 12. The left picture shows the $\log_{10}$ of the absolute value of the Fourier coefficients $\gamma_k(f)$ (marked with $\circ\circ\circ$) of a real-valued smooth T-periodic function f and of the discrete Fourier coefficients $\widetilde{\gamma}_k(f)$ (marked with $***$) for $N=100$, the right picture shows the $\log_{10}$ of the absolute value of the error $\gamma_k(f)-\widetilde{\gamma}_k(f)$. For these pictures, $f(t)\equiv(t^2-1)\cos^2(10\pi t)$ on $[-1,+1]$, $T=2$.

With $\phi_n(k)\equiv\exp(2\pi i\frac{kn}{N})$ and $\zeta\equiv\exp(2\pi i\frac{n-m}{N})$, we have that

$$<\phi_n,\phi_m>=\sum e^{2\pi i k\frac{n-m}{N}}=\sum\zeta^k=\begin{cases}\dfrac{\zeta^N-1}{\zeta-1}=0 & \text{if } n\neq m\\ N \quad \text{if } n=m.\end{cases}$$

This shows that the ϕ_k, $(k=0,\ldots,N-1)$, form an orthogonal system. In particular, they are linearly independent. Since their number is precisely the dimension of ℓ_N, they form a basis. Therefore, f is of the form $f=\sum\alpha_k\phi_k$ for certain scalars α_k. Taking the inner product of this expression with ϕ_m and using the orthogonality shows that $N\alpha_k=<f,\phi_k>$. Hence,

$$f=\sum\alpha_k\,\phi_k,\quad\text{where}\quad\alpha_k=\frac{<f,\phi_k>}{N}=\frac{1}{N}\sum f(k)\,e^{-2\pi i\frac{kn}{N}}=\widetilde{\gamma}_n.$$

The last expression follows from the definition of the inner product and of ϕ_k. $\square$

The definition of the $\widetilde{\gamma}_k$ and the statement for f_n in the theorem is correct for all k and n in $\mathbb{Z}$, although it suffices to let k and n run from 0 to $N-1$. This observation is of interest, if we want to use $\widetilde{\gamma}_k$ as an approximation to $\gamma_k(f)$: the approximation will be useful for $|k|<\frac{1}{2}N$ and most accurate for $|k|\ll N$ (cf., Fig. 12).

5.3 Discrete cosine transform (DCT). If the T-periodic function f is real-valued and even, then the sequence of Fourier coefficients $\gamma_k(f)$ is even and real and f can be expressed as a linear combination of cosines. Discrete variants lead to *discrete cosine transforms* (DCTs). The DCTs, and in particular the DCT-II, are often used in signal and image processing, especially for lossy data compression, because they have a strong "energy compaction" property: most of the signal information tends to be concentrated in a few low-frequency components of the DCT.

If $f_0,\ldots,f_N$ is a sequence of real numbers, then there are several ways of expanding this sequence to an even one. This leads to several variants of the DCT. For instance, with $g_n\equiv f_n$ and $g_{N+n}\equiv f_{N-n}$ for $n=0,\ldots,N-1$, i.e.,

$$(g_n)=(f_0,f_1,f_2,\ldots,f_{N-1},f_N,f_{N-1},\ldots,f_2,f_1),$$

g_n is defined for $n = 0, \ldots, 2N-1$ and the DFT in (94) leads to DCT-I:

$$\begin{aligned}\gamma_k &= \frac{1}{2N}\sum_{n=0}^{2N-1} g_n e^{-2\pi i \frac{nk}{2N}} = \frac{1}{2N}\left(\sum_{n=0}^{N-1} f_n e^{-\pi i \frac{nk}{N}} + \sum_{n=1}^{N} f_n e^{-\pi i \frac{(2N-n)k}{N}}\right) \\ &= \frac{1}{2N}\left(\sum_{n=0}^{N-1} f_n e^{-\pi i \frac{nk}{N}} + \sum_{n=1}^{N} f_n e^{\pi i \frac{nk}{N}}\right) \\ &= \frac{1}{2N}\left(f_0 + (-1)^k f_N\right) + \frac{1}{N}\sum_{n=1}^{N-1} f_n \cos(\pi\frac{nk}{N}).\end{aligned}$$

Note that the periodic extension $g_n \equiv g_{n+2jN}$ $(n \in \mathbb{Z})$ is even both around $n = 0$ and around $n = N$. The expression for γ_k is correct for all $k \in \mathbb{Z}$. Note that γ_k is also even around $k = 0$ and $k = N$. Therefore, the expression for γ_k has to be computed only for $k = 0, \ldots, N$. Since γ_k and g_n have the same symmetry properties, we can easily conclude from the second relation in (94) that the inverse of DCT-I is given by

$$f_n = (\gamma_0 + (-1)^n \gamma_N) + 2\sum_{k=1}^{N-1} \gamma_k \cos(\pi\frac{nk}{N}).$$

The extension

$$(g_n) = (f_0, f_1, \ldots, f_{N-1}, f_{N-1}, \ldots, f_1, f_0) = (f, f^{\mathrm{T}}),$$

where $f \equiv (f_0, \ldots, f_{N-1})$ and f^{T} is f in reverse order, leads to DCT-II:

$$\gamma_k \equiv \frac{1}{N}\sum_{n=0}^{N-1} f_n \cos(\pi\frac{(n+\frac{1}{2})k}{N}) \quad \text{and} \quad f_n = \gamma_0 + 2\sum_{k=1}^{N-1} \gamma_k \cos(\pi\frac{(n+\frac{1}{2})k}{N}).$$

The periodic extension of (g_n) is even around $n = -\frac{1}{2}$ and even around $n = N - \frac{1}{2}$. The sequence of coefficients γ_k is even around $k = 0$, odd around $k = N$, and $\gamma_N = 0$. Actually, application of (94) leads to a coefficient $\widetilde{\gamma}_k$ that is the above γ_k times ζ^k, where $\zeta \equiv \exp(\pi i \frac{1}{2N})$. Coordinate-wise multiplication of the sequence of γ_k by the sequence of ζ^k results in a sequence that is even around $k = N$. These factors ζ^k have been moved to the formula for f_n to keep the formulas real.

DCT-III arises from the sequence

$$(g_n) = (f_0, f_1\zeta^1, \ldots, f_{N-1}\zeta^{N-1}, 0, -f_{N-1}\zeta^{N+1}, \ldots, -f_1\zeta^{2N-1}). \tag{95}$$

The periodic extension of (g_n) is equal to $(\ldots, f_0, \widetilde{f}, 0, -\widetilde{f}^{\mathrm{T}}, -f_0, -\widetilde{f}, 0, \ldots)$ times z, where $\widetilde{f} \equiv (f_1, \ldots, f_{N-1})$ and $z \equiv (\zeta^n)$. Note that $\zeta^{2N} = 1$, $\zeta^{N-n} = -\overline{\zeta^{N+n}}$, and $\zeta^{-n} = \overline{\zeta^n}$. Hence, z is $2N$-periodic, odd around $n = N$, and even around $n = 0$. This shows that the periodic extension of (g_n) is not only even around $n = N$, but also around $n = 0$. Except for the position of the scaling factor $2N$, DCT-III is the inverse of DCT-II. Without the factors ζ^n, the sequence in (95) is odd around $n = N$.

The sequence $(\ldots, f, -f^{\mathrm{T}}, -f, f^{\mathrm{T}}, \ldots)$ is odd around $n = N - \frac{1}{2}$ and even around $n = -\frac{1}{2}$. Multiplication by $\widetilde{z} \equiv (\zeta^{n+\frac{1}{2}})$ leads to DCT-IV:

$$\gamma_k = \frac{1}{N}\sum_{n=0}^{N-1} f_n \cos(\pi\frac{(n+\frac{1}{2})(k+\frac{1}{2})}{N}) \quad \text{and} \quad f_n = 2\sum_{k=0}^{N-1} \gamma_k \cos(\pi\frac{(n+\frac{1}{2})(k+\frac{1}{2})}{N}).$$

The sequence $\widetilde{z}$ is also odd around $N-\frac{1}{2}$ ($z_{N-1}=-\overline{z_N},\ldots$) and even around $n=-\frac{1}{2}$ ($z_{-1}=\overline{z_0},\ldots$).

DCT of type V–VIII follow from making sequences (g_n) of length $2N-1$ by odd or even extensions (before the periodic extension). However, these variants seem to be rarely used in practice.

DCT-II seems to be the most commonly used form of discrete cosine transform and is often referred to as "the DCT". Matlab uses DCT-IV.

From `en.wikipedia.org`: The DCT is used in JPEG image compression, MJPEG video compression, and MPEG video compression. There, the two-dimensional DCT-II of 8x8 blocks is computed and the results are filtered to discard small (difficult-to-see) components. That is, n is 8 and the DCT-II formula is applied to each row and column of the block. The result is an array in which the top left corner is the DC (zero-frequency) component and lower and rightmore entries represent larger vertical and horizontal spatial frequencies.

A related transform, the modified discrete cosine transform, or MDCT (see Exercise 5.9), is used in AAC, Vorbis, and MP3 audio compression.

DCTs are also widely employed in solving partial differential equations by spectral methods, where the different variants of the DCT correspond to slightly different even/odd boundary conditions at the two ends of the array; cf., Exercise 5.8.

The discrete Fourier transform (and the real variants, the DCTs) plays a central role in *all* numerical computations of Fourier transform. This is the case for periodic functions, as we saw above, but also for non-periodic ones, as we will explain now.

5.4 If f is in $L^2(\mathbb{R})$ then (see Theorem 3.12).

$$\widehat{f}(\omega)=\int_{-\infty}^{\infty} f(t)\, e^{-2\pi i t\omega}\, \mathrm{d}t \quad \text{and} \quad f(t)=\int_{-\infty}^{\infty} \widehat{f}(\omega)\, e^{2\pi i t\omega}\, \mathrm{d}\omega.$$

For numerical computations, we have to discretize.

First, we can select an interval $[a,b]$ such that f is small outside $[a,b]$, or, to be more precise, $\int_{-\infty}^{\infty}|f(t)|\,\mathrm{d}t\approx\int_a^b|f(t)|\,\mathrm{d}t$. Then,

$$\widehat{f}(\omega)\approx\int_a^b f(t)\, e^{-2\pi i t\omega}\, \mathrm{d}t \qquad (\omega\in\mathbb{R}).$$

Now, with $T\equiv b-a$, we can follow the same discretization strategy as for T-periodic functions. For $\Delta t\equiv\frac{T}{N}$, select a sequence of equally space base points $t_n=t_0+n\Delta t$, $n=0,\ldots,N-1$ in $[a,b)$. For ω_0, let $\omega_k\equiv\omega_0+\frac{k}{T}$ for $k\in\mathbb{Z}$. Then, a Riemann sum approach leads to $\widehat{f}(\omega_k)\approx\int_a^b f(t)\,e^{-2\pi i t\omega_k}\,\mathrm{d}t\approx\sum_{n=0}^{N-1}\Delta t f(t_n)\,e^{-2\pi i t_n\omega_k}$. Hence,

$$\widehat{f}(\omega_k)\approx\left(\frac{1}{N}\sum_{n=0}^{N-1} f_n\, e^{-2\pi i\frac{nk}{N}}\right) T\, e^{-2\pi i t_0(\omega_k-\omega_0)}, \quad \text{where } f_n\equiv f(t_n)\, e^{-2\pi i t_n\omega_0}. \tag{96}$$

Similarly,

$$f(t_k)\approx\left(\sum_{k=0}^{N-1} g_k\, e^{2\pi i\frac{nk}{N}}\right) e^{2\pi i t_n\omega_0}, \quad \text{where } g_k\equiv\widehat{f}(\omega_k)\,\frac{e^{2\pi i t_0(\omega_k-\omega_0)}}{T}. \tag{97}$$

Due to aliasing, our approximations for $\widehat{f}(\omega_k)$ will only be of interest for N values of k. Therefore, it is convenient to select the ω_0 such that the ω_k are in the range of interest, for $k = 0, \dots, N-1$.

Note that the expression in brackets are discrete Fourier transforms (cf., Theorem 5.2).

The step size $\frac{1}{T}$ in the frequency domain is determined by the length of the interval $[a, b]$ in time domain that carries the interesting part of f. The number N of interesting sample points in both time and frequency domain is determined by $\Delta t = T/N$, that is, by the sample rate in time domain. Since the discrete Fourier transform only leads to a useful approximation for at most N frequencies spanning an interval $[\omega_0, \omega_N)$ of length $\omega_N - \omega_0 = N/T = 1/\Delta t$, one can also argue that N is determined by the length of the interval in frequency domain that carries the interesting part of $\widehat{f}$.

Often large values are required for both T as well as for $1/\Delta t$. This implies that N will be huge. For instance, with $T = 1000$ and $1/\Delta t = 1000$, N will be 10^6. But larger values for N are not unexceptional. Since discrete Fourier transforms are performed in all numerical computations involving Fourier transforms (for instance, in digital signal processing, in date compression as in MP3 and JPEG, etc.) it will be clear that it is important to have an efficient algorithm for this transform. This is the subject of the next section.

5.5 Fast Fourier Transform. Let $\alpha_0, \dots, \alpha_{N-1}$ be scalars and suppose we want to compute

$$f_n \equiv \sum_{k=0}^{N-1} \alpha_k \, e^{2\pi i \frac{kn}{N}} \qquad \text{for} \quad n = 0, \dots, N-1. \tag{98}$$

Obviously, an efficient algorithm for computing the f_n, is an efficient algorithm for the discrete Fourier transform.

If we assume that the values for $\exp(2\pi i \frac{kn}{N})$ are available, then a naive implementation of the formula in (98) would still require N multiplications and $N-1$ additions to compute one f_n: the N values f_n can be computed in $(2N-1)N \approx 2N^2$ flops (floating point operations). The *Fast Fourier Transform* (FFT) algorithm can it do in $2N\ell$ flop where $\ell = \log_2(N)$. If, for instance, $N = 2^{20} \approx 10^6$, then $\ell = 20$ and there is a gain by a factor $N/\ell = 5\,10^4$ which is the difference between approximately 14 hours and 1 second!

Radix 2. To explain the idea of FFT, assume that $N = N_\ell \equiv 2^\ell$ for some positive integer ℓ. Put $M \equiv N_{\ell-1} = N/2$. Now, write f_n as

$$f_n = \left(\sum_{2k<N} \alpha_{2k} \, e^{2\pi i \frac{2kn}{N}} \right) + \left(\sum_{2k+1<N} \alpha_{2k+1} \, e^{2\pi i \frac{2kn}{N}} \right) e^{2\pi i \frac{n}{N}}. \tag{99}$$

For ease of discussion, let us denote the term in the first pair of brackets by $f_{e,n}$ and the one in the second pair by $f_{o,n}$ ('e' for 'even', 'o' for 'odd'):

$$f_n = f_{e,n} + f_{o,n} \, e^{2\pi i \frac{n}{N}} = f_{e,n} + f_{o,n} \, e^{\pi i \frac{n}{M}}. \tag{100}$$

We now concentrate on $f_{e,n}$. Note that

$$f_{e,n} = \sum_{k=0}^{M-1} \alpha_{2k} \, e^{2\pi i \frac{kn}{M}},$$

which is an expression as in (99), but with N reduced by a factor 2 to M. We are only summing over half of the number of terms (M instead of N). But, of course, that does not lead to savings in the computational costs, since we also have to compute the $f_{o,n}$. However, $\exp(2\pi i\frac{k(n+M)}{M}) = \exp(2\pi i\frac{kn}{M})$ and, consequently, $f_{e,n+M} = f_{e,n}$ for all n. The aliasing phenomenon is helpful here: we only have to compute $f_{e,n}$ for the values $n = 0,\ldots,M-1$ and we get the $f_{e,n}$ for n from M to $N-1$ for free. The same observation applies to $f_{o,n}$. Since $\exp(\pi i\frac{n+M}{M}) = -\exp(\pi i\frac{n}{M})$, we now have that

$$f_{n+M} = f_{e,n} - f_{o,n}\, e^{\pi i \frac{n}{M}}. \tag{101}$$

We apply (100) and (101) for $n = 0,\ldots M-1$ to obtain the f_n for $n = 0,\ldots,N-1$: we have to compute only half of the $f_{e,n}$ and $f_{o,n}$ values and this is where the savings come from. Of course, this idea can be applied recursively (to the f_e and f_o, etc.), and that is what FFT does.

We explained two recursive step at the 'highest' level (from $N_{\ell-1}$ to N_ℓ). FFT starts the computation at the lowest level (from N_0 to N_1) and then works recursively towards the highest. Note that the 'Fourier coefficients' α_{2k} and α_{2k+1} of $f_{e,n}$ and $f_{o,n}$, respectively, coincide with the Fourier coefficients of f_n; there is only a difference in numbering. Similarly, the Fourier coefficients at the lowest level are equal to the α_k's.

To analyze the computational costs of FFT, let $\kappa_{\ell-1}$ denote the number of flops to compute the values $f_{e,n}$ for $n = 0,\ldots,M-1$. Then we have that

$$\kappa_\ell = 2\kappa_{\ell-1} + 2N. \tag{102}$$

The first factor 2 expresses the fact that computing the $f_{o,n}$ is as expensive as computing the $f_{e,n}$. The factor $2N$ comes from the multiplication by $\exp(2\pi i n/N)$ and summing the 'e' and 'o' components for each of the N values of n (see (100) and (101)). Recursive application of (102) leads to $\kappa_\ell = 2(\kappa_{\ell-1} + N) = 2(2\kappa_{\ell-2} + 2M) + 2N = 2^2\kappa_{\ell-2} + 2N + 2N = \ldots = 2^\ell\kappa_0 + 2N\ell$. The number of flops for computing (98) at the lowest level is 0: $\kappa_0 = 0$. Therefore,

$$\kappa_\ell = 2N\ell.$$

In our discussion above, we assumed that the values of $\exp(2\pi i\frac{kn}{N})$ are available. They can be computed in an implicit way: with $\omega_1 \equiv \exp(\pi i\frac{1}{M})$, we can compute the $\widetilde{f}_{o,n} \equiv f_{o,n}\,\exp(\pi i\frac{n}{M})$ recursively as

$$\begin{array}{l} \omega = 1, \\ \texttt{for } n = 0 \texttt{ to } M-1 \\ \quad \widetilde{f}_{o,n} = f_{o,n}\,\omega \\ \quad \omega = \omega\,\omega_1 \end{array}$$

This scheme requires two multiplications for each n, but since $\widetilde{f}_{o,n}$ can be used twice (cf. (100) and (101)), the cost formula (102) is still correct. Apparently, all the f_n can be computed in $2N\ell$ flops plus ℓ exponentials $\exp(\pi i 2^{-j})$ $(j < \ell)$.

Although it is relatively simple to write a code for the FFT algorithm, we do not encourage to do it yourself. There are so many details to take care of if you really want your code to be efficient. There are excellent codes available, codes that have been optimized for specific computer architectures and processors.

Radix d. In the discussion above we assumed N to be a power of 2. What if that is not the case?

Of course, we can adopt the above saving strategy when N is, say, a power of 3. Then, we split the sum for f_n in three parts involving α_{3k}, α_{3k+1}, and α_{3k+2}, respectively, (cf. (99)), etc.. The resulting FFT algorithm is called FFT with *radix 3*. The FFT with radix 4 appears to be the most efficient variant, even more efficient than the above 'basic' variant of radix 2.

General. More generally, we can decompose N as a product of powers of primes and we can formulate FFT variants accordingly. However, such a strategy will not be highly efficient: on each level, we have to deal with another prime number. The following observations are more useful.
Suppose we have a choice of applying FFT with $N = 2^\ell$ or the naive approach with a much smaller N, say, $N = \frac{1}{2}2^\ell + 1$. Then, it is interesting to note that the FFT on the longer sequence is still faster than the naive approach. It is faster by approximately a factor $N/(4\ell)$; this is still the difference between 3h30 and 1 sec. in case $N = 2^{20} \approx 10^6$. Therefore, if we are in the situation of §§5.1 or 5.4 and we can select our sample ratio Δt or the sample length T freely, then it is more efficient to take a larger N in order to get a power of 2. In a number of other applications, the Fourier transforms are only used for fast computations and not for, say, filtering: filtering affects the functions in frequency domain (see, for instance, Exercise 5.4). In such applications, the sequence of (α_k) can safely be trailered with zeros in order to get a sequence of the desirable length.
Finally, we mention that a discrete Fourier transform can be viewed as a convolution product between two sequences of complex numbers (see Exercise 5.5). Since convolution products can be efficiently computed with FFT with radix 2 (and 4), this offers also a possibility for fast computation of the discrete Fourier; for more details, see Exercise 5.5.

The DCTs also have fast variants. This is not surprising, since the DCTs can be obtained from the complex DFT by simple rearranging terms.

NOTE. *The FFT algorithm has been introduced by Cooley and Tukey in 1965 [3]. Their paper is one of the most cited mathematical papers. It was published at a time when computers firstly made large scale computations possible. Earlier publications of the fast algorithm (by Gauss in the first half of the 19th century (published in his collected works, 1866) and by Runge (1903)) appear to have been forgotten in 1965.*

Exercises

Exercise 5.1. Accuracy of the DFT.

Let f be sufficiently smooth and T-periodic. We sample f at $t_n \equiv n\Delta t$ with sample ratio $\Delta t = T/N$ and approximate γ_k by $\widetilde{\gamma}_k(f) = \frac{1}{N}\sum_{n=0}^{N-1} f(t_n)e^{-2\pi i \frac{nk}{N}}$.

(a) Prove that $\|f - S_n(f)\|_\infty \leq \sum_{|k|>n} |\gamma_k(f)|$.

(b) Prove that $|\gamma_k(f) - \widetilde{\gamma}_k(f)| \leq \sum_{j\neq 0} |\gamma_{k+jN}(f)|$.

(c) Prove that $\|f - \widetilde{S}_n(f)\|_\infty \leq 2\sum_{|k|>n} |\gamma_k(f)|$ if $|k| < \frac{1}{2}N$.

Here, $\widetilde{S}_n(f)$ is the nth partial Fourier series of f with $\gamma_k(f)$ replaced by $\widetilde{\gamma}_k(f)$.
Prove that $\widetilde{S}_n(f)$ is the inverse discrete Fourier transform of the $(\widetilde{\gamma}_k(f))$ as defined in Theorem 5.2.

(d) Suppose that $f \in C^{(i)}(\mathbb{R})$ for some $i \geq 2$. Put $\kappa \equiv \left(\frac{T}{2\pi}\right)^i \|f^{(i)}\|_1$.
Show that, if $|k| < \frac{1}{2}N$,

$$|\gamma_k(f) - \widetilde{\gamma}_k(f)| \leq 2\kappa \sum_{j=1}^{\infty} (jN - |k|)^{-i} \leq 3\kappa\,(N - |k|)^{-i}.$$

(Hint: use the integral criterion to estimate $\sum_{j=2}^{\infty}(jN - |k|)^{-i}$.)

(e) If N is even and $N = 2M$ then $f(t_n) = \widetilde{S}_M(f)(t_n)$ for all $n \in \mathbb{Z}$. Prove this.
$\widetilde{S}_M(f)$ interpolates f at the equally space grid-points $t_n = n\Delta t$: $\widetilde{S}_M(f)$ is the interpolating trigonometric polynomial of f at t_n.

Exercise 5.2. Consider the construction of §5.4. Put

$$\widehat{f}_d(\omega_k) \equiv \left(\frac{1}{N} \sum_{n=0}^{N-1} f_n\, e^{-2\pi i \frac{nk}{N}}\right) T e^{-2\pi i t_0 (\omega_k - \omega_0)}.$$

Estimate the computational costs to compute $\widehat{f}_d(\omega_k)$ for $k = 0, \ldots, N-1$.

Exercise 5.3. **Discrete convolution product.**
Let ℓ_N be the space of sequences of N complex number.
If $\alpha = (\alpha_0, \ldots, \alpha_{N-1})$ the we will also write $\alpha(n)$ instead of α_n. We denote the discrete Fourier transform of α by $\mathcal{F}(\alpha)$:

$$\mathcal{F}(\alpha)(k) \equiv \sum_{n=0}^{N-1} \alpha_n\, e^{-2\pi i \frac{kn}{N}} \qquad (k = 0, \ldots, N-1).$$

We define the *convolution product* $\alpha * \beta$ on ℓ_N as follows

$$\alpha * \beta(n) \equiv \sum_{k=0}^{N-1} \alpha(n-k)\, \beta(k) = \sum_{k=0}^{N-1} \alpha(k)\, \beta(n-k) \qquad (n = 0, \ldots, N-1). \tag{103}$$

Here, we assumed $\alpha(k)$ to be defined for values of k outside $\{0, \ldots, N-1\}$ by periodic extension: $\alpha(k+N) = \alpha(k)$ $(k \in \mathbb{Z})$.

(a) Prove the equality in (103).

(b) Prove that $\mathcal{F}$ is a bijection from ℓ_N onto ℓ_N.

(c) Prove that

$$\mathcal{F}(\alpha * \beta) = \mathcal{F}(\alpha) \cdot \mathcal{F}(\beta) \qquad (\alpha, \beta \in \ell_N).$$

Here, $\cdot$ is the coordinate-wise product: $\mu \cdot \nu(n) \equiv \mu(n)\, \nu(n)$ for $\mu, \nu \in \ell_N$.

(d) Show that $\mathcal{F}^{-1}(\mu) * \mathcal{F}^{-1}(\nu) = \mathcal{F}^{-1}(\mu \cdot \nu)$ for all $\mu, \nu \in \ell_N$.

Exercise 5.4. **Multiplying polynomials.**
Let p and q be polynomials of degree M and L, respectively, $p(x) = \sum_{n=0}^{M} \alpha_n\, x^n$ and $q(x) = \sum_{k=0}^{L} \beta_k\, x^k$.

Let $N \geq M + L$. By defining $\alpha_n \equiv 0$ if $n > M$, and $\beta_k \equiv 0$ if $k > L$, the sequences (α_n) and (β_k) belong to ℓ_N. The convolution is defined as in (103).

(a) Prove that the product polynomial pq is given by

$$pq(x) \equiv p(x)\, q(x) = \sum_{m=0}^{N} \alpha * \beta(m)\, x^m.$$

(b) Compute the convolution product of the sequence $\alpha = (1, 2, 1)$ and $\beta = (1, 4, 6, 4, 1)$.

(c) Prove that $\alpha * \beta = \mathcal{F}^{-1}(\mathcal{F}(\alpha) \cdot \mathcal{F}(\beta))$ also if we increase N and trailer the sequences of αs and βs with zeros.

(d) The coefficients of the product polynomial can be computed i) directly from the definition of the convolution product, but also ii) via the discrete Fourier transform and its inverse. The discrete Fourier transform and its inverse can be computed in ii.a) the naive way with $N = M + L$ or ii.b) with FFT and $N = 2^\ell$, ℓ such that $2^{\ell-1} < M + L \leq 2^\ell$.
Analyze and compare the computational costs of these three approaches i), ii.a), and ii.b). What is the most efficient one? Does it depend on the size of M and L?

Exercise 5.5. ZCT and DFT as a convolution product.

Here, we will see that the DFT can be written as a convolution product.

Let $N \in \mathbb{N}$ and let $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{N-1})$ be a sequence in $\mathbb{C}$.

(a) For $\zeta_0, \zeta_1 \in \mathbb{C}$, non-zero, let $z_k \equiv \zeta_0 \, \zeta_1^k \quad (k \in \mathbb{Z})$. Consider the transform

$$\gamma_k \equiv \sum_{n=0}^{N-1} \alpha_n z_k^{-n} \qquad (k = 0, \ldots, N-1). \tag{104}$$

A transform of this form is called a *z-chirp transform* (ZCT) .
Show that the DFT is a special kind of z-chirp transform ($\zeta_0 = 1$ and $\zeta_1 = \exp(2\pi i/N)$).

(b) Consider formula (104). Note that $2kn = n^2 + k^2 - (n-k)^2$. Let $\zeta \in \mathbb{C}$ be such that $\zeta^2 = \zeta_1$. Prove that

$$\gamma = (\widetilde{\alpha} * \beta)/\beta, \quad \text{where } \widetilde{\alpha}(n) \equiv \alpha_n \, \zeta_0^{-n} \, \zeta^{-n^2}, \ \beta(n) \equiv \zeta^{n^2}.$$

What is the range of n for $\gamma(n) = \gamma_n$, $\widetilde{\alpha}(n)$ and $\beta(n)$? Has the range to be extended?

(c) Describe a procedure using FFT with radix 2 that computes the z-chirp transform (104) also in case N is not a power of 2. Pay special attention to the way you extend the range of β.

(d) Analyze the computational costs of this approach. What are the costs for perfoming the DFT (98) with this approach?

(e) Write a MATLAB code to perform DFT for any N based on FFT with radix 2.

Exercise 5.6. Convolution and circulant matrices.

We can identify sequences $(f_0, \ldots, f_{N-1})$ of N complex numbers with vectors f in the N-dimensional vector space ℓ_N. Consequently, the discrete Fourier transform $\mathcal{F}$ and its inverse can be identified with N by N matrices. Let the N by N matrix F represent the discrete Fourier transform (i.e., the transform from the f_n to the $\widetilde{\gamma}_k$ in Theorem 5.2. See also Exercise 5.3).

(a) Describe the matrix F.

(b) Prove that the inverse Fourier transform is represented by F^{-1} and that $F^{-1} = NF^{\mathrm{H}}$, where F^{H} is the conjugated transpose of F.

(c) If $\alpha \in \ell_N$, then the convolution product $\alpha * \beta$ defines a linear map from ℓ_N to ℓ_N.
Describe the matrix A that represents this map: $A\,\beta = \alpha * \beta$.
Show that $A = (a_{ij})$ is a *circulant*, that is $a_{ij} = a_{i+1\,j+1}$ if $j < N$ and $a_{ij} = a_{i+1\,1}$ if $j = N$.
Show that the 'convolution map' identifies the circulant matrices with vectors in ℓ_N.

(d) Show that the columns of F are eigenvectors of A and that the associated eigenvalues are the discrete Fourier coefficients $\mathcal{F}(\alpha)(k)$ of α.

Exercise 5.7. DCT.

Prove the formulas for DCT-II, DCT-III, and DCT-IV.

Exercise 5.8. DCT and differential equations.

The Fourier transform and the discrete Fourier transform can be viewed as transformations from one orthogonal basis system to another. The same is true for discrete cosine transforms, although this is a bit hard to see. The following questions offer an alternative approach for proving orthogonality of the cosines that are involved in the DCTs: the cosines show up in the numerical solution of the symmetric eigenvalue problem $f'' = \lambda f$ with appropriate boundary conditions as $f'(0) = f'(1) = 0$ ($f \in C^{(2)}([0,1])$).

(a) Consider the N by N matrix A with all 1 on the two first co-diagonals ($A_{i,i+1} = A_{i+1,i} = 1$) and -2 on the diagonal, except for $A_{1,1}$ and $A_{N,N}$ which are equal to -1. Prove that the vectors $\vec{\phi}_k$ with nth coordinate equal to $\cos(\pi k(n+\frac{1}{2})/N)$ are eigenvectors of A. Conclude from the fact that A is symmetric that the vectors $\vec{\phi}_k$ form an orthogonal system. Show that the cosines involved in DCT-II form an orthogonal system.

Explanation: A can be viewed as the discretization of the second derivative $f \rightsquigarrow f''$, where f has been discretized on $[0,1]$ in the points $t_n \equiv (n-\frac{1}{2})\Delta t$ with $\Delta t \equiv 1/N$: $f_n \approx f(t_n)$. If we require the derivative of f to be zero at $t=0$, then $f(\frac{1}{2}\Delta t) - f(-\frac{1}{2}\Delta t) \approx f_0 - f_{-1} = 0$ combined with $f(-\frac{1}{2}\Delta t) - 2f(\frac{1}{2}\Delta t) + f(\frac{3}{2}\Delta t) \approx f_0 - 2f_1 + f_2 = \lambda f_1$ yields $-f_1 + f_2 = \lambda f_1$. This explains the choice $A_{1,1} = -1$. Similarly, the boundary condition $f'(1) = 0$ leadsto $A_{N,N} = -1$.

(b) For DCT-I, consider the $N+1$ by $N+1$ matrix A with ones on the two first co-diagonals and -2 on the diagonal except for $A_{1,2} = A_{N+1,N} = 2$. Show that that the vectors $\vec{\phi}_k$ with nth coordinate $\cos(\pi\frac{kn}{N})$ are eigenvectors of A. A is not symmetric but $J^{-1}AJ$ is, if J the identity matrix except for $J_{1,1} = J_{N+1,N+1} \equiv \sqrt{2}$.

Explanation: Here f has been discretized in the points $t_n = (n-1)\Delta t$, $\Delta t = 1/N$. Again the derivative of f is taken to be zero at 0 and at 1: $f(\Delta t) - f(-\Delta t) \approx f_2 - f_0 = 0$, and $f_0 - 2f_1 + f_2 = \lambda f_1$ transforms into $-2f_1 + 2f_2 = \lambda f_1$, etc..

(c) DCT-IV can be treated with the N by N matrix A with 1 one the first two co-diagonals, -2 on the diagonal, except for $A_{1,2} = 2$ and $A_{N,N} = -3$.

Explanation: the discretization is again in $t_n = (n-\frac{1}{2})/N$. The derivative of f at 0 is again 0, but at 1 we require $f(1) = 0$. $f(1) = 0$ discretizes as $f((N-\frac{1}{2})\Delta t) = -f((N+\frac{1}{2})\Delta t)$, or, $f_N = -f_{N+1}$, which leads to $A_{N,N} = -3$.

Exercise 5.9. **Modified discrete cosine transform (MDCT).**

The MDCT is unusual as compared to the discrete Fourier-related transforms in that it has half as many outputs as inputs.

For $N = 2M$ even and a sequence $f = (f_0, \dots, f_{2N-1})$ of $2N$ real numbers the N coefficients $\gamma_0, \dots, \gamma_{N-1}$ are given by

$$\gamma_k \equiv \frac{1}{N}\sum_{n=0}^{2N-1} f_n \cos(\pi\frac{(k+\frac{1}{2})(n+\frac{1}{2}+M)}{N}) \tag{105}$$

with 'inverse' the sequence $\widetilde{f} = (\widetilde{f}_0, \dots, \widetilde{f}_{2N-1})$ of $2N$ real numbers given by

$$\widetilde{f}_n \equiv \sum_{k=0}^{N-1} \gamma_k \cos(\pi\frac{(k+\frac{1}{2})(n+\frac{1}{2}+M)}{N}). \tag{106}$$

Note that MDCT as defined in (105) and the 'inverse' IMDCT from (106) are shifted variants of DCT-IV and its inverse. Of course, MDCT does not have an inverse (why not?). But if we add the results of the IMDCT of subsequent overlapping blocks, we have perfect invertibility as we will explain in (c) below.

(a) Let (F_1, F_2, F_3, F_4) be the sequence f partioned in blocks F_i of length M. Show that MDCT applied to the sequence (F_1, F_2, F_3, F_4) of $2N$ reals leads to the same coefficients γ_k as DCT-IV applied to the sequence $(-F_3^{\mathrm{T}} - F_4, F_1 - F_2^{\mathrm{T}})$ of N reals. Here, F^{T} is F in reverse order (the 'transpose' of F).

(b) Show that the inverse of DCT-IV applied to (γ_k) is equal to

$$(-F_3^{\mathrm{T}} - F_4, F_1 - F_2^{\mathrm{T}}, -F_1^{\mathrm{T}} + F_2, F_3 + F_4^{\mathrm{T}}, F_3^{\mathrm{T}} + F_4, -F_1 + F_2^{\mathrm{T}}).$$

Here, we considered the inverse DCT-IV to be defined for all $n \in \mathbb{Z}$; we displayed the inverse values only for $n = 0, \dots, 3N-1$.

Conclude that the inverse MDCT leads to

$$\begin{aligned}\widetilde{f} &= \tfrac{1}{2}(F_1 - F_2^{\mathrm{T}}, -F_1^{\mathrm{T}} + F_2, F_3 + F_4^{\mathrm{T}}, F_3^{\mathrm{T}} + F_4) \\ &= \tfrac{1}{2}((F_1, F_2) - (F_1, F_2)^{\mathrm{T}}, (F_3, F_4) + (F_3, F_4)^{\mathrm{T}}).\end{aligned}$$

The last line interprets the result when the sequence f is supposed to be partioned into two blocks each of length N, as $((F_1, F_2), (F_3, F_4))$. Note that time values from the second block (in F_2) get represented in the first block, etc.. This phenomenon is referred to as *time-domain aliasing.*

(c) **TDAC.** Now, consider the sequence $F \equiv (F_1, F_2, \ldots, F_\ell)$, where now F_i are blocks of length N (twice as long as the blocks above!) and $\ell \geq 3$. Apply MDCT to the subsequences (F_i, F_{i+1}) of length $2N$, to obtain the sequences Γ_i of length N. Note that the subsequences (F_j, F_{j+1}) overlap: for $j = i, i+1$, they share the block F_{i+1}.

Let the inverse MDCT applied to Γ_i be denoted by (G_i, H_{i+1}). Show that $H_{i+1} + G_{i+1} = F_{i+1}$: the reverse terms cancel. This property is called 'time-domain aliasing cancellation' (TDAC).

Conclude that we have perfect reconstruction by adding the overlapping IMDCTs of MDCTs applied to subsequent overlapping blocks if the first block F_1 and the last block F_ℓ are zero.

(d) **Windows.** Usually window functions are used to attenuate effects of applying the transform per block (to F_i rather than f with a Fourier transform). Let $W = (W_1, W_2)$ be a sequence of length $2N$ partioned into two blocks W_i each of length N. As in c), $F = (F_1, \ldots, F_\ell)$ is partioned into blocks F_i of size N. Apply MDCT to the sequence $(W_1 F_i, W_2 F_{i+1})$ to obtain Γ_i. The multiplication is point-wise. Here, W acts as a 'window'; W is a *window* function. Let the results of IMDCT applied to Γ_i be denoted by (G_i, H_{i+1}). Application of the same window leads to $(W_1 G_i, W_2 H_{i+1})$.

Prove perfect reconstruction, i.e., $W_2 H_{i+1} + W_1 G_{i+1} = F_{i+1}$, if

$$W_2 = W_1^{\mathrm{T}} \quad \text{and} \quad W_1^2 + W_2^2 = 1.$$

The first property makes the window *symmetric* (around $k = N + \frac{1}{2}$), the second property is the *Princen–Bradley condition.*

Show that both the *sine window* $W = (w_k)$ with $w_k \equiv \sin \frac{\pi}{2N}(k + \frac{1}{2})$ $(k = 0, \ldots, 2N-1)$ and the window $w_k \equiv \sin\left(\frac{\pi}{2} \sin^2\left[\frac{\pi}{2N}\left(k + \frac{1}{2}\right)\right]\right)$ are symmetric and satisfied the Princen–Bradley condition. Note that the w_k are small for $k \approx 0$ and $\approx 2N-1$. The sine window is used in MP3, MPEG2-AAC, the second one in Vorbis. The Kaiser windows (used in, e.g., MPEG4-AAC) are based on 0th order Bessel functions.

Here, we assumed that the *analysis* window (the MDCT transforms to frequency domains, thus, prepares for spectral analysis) and the *synthesis* window (with IMDCT, spectrum information is transformed back to time domain) are the same. However, perfect reconstruction is also possible with different (but carefully selected) windows.

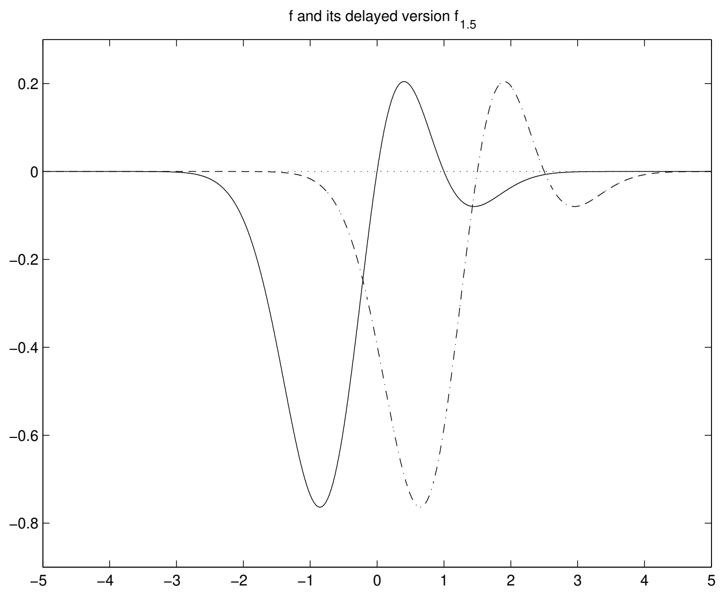


FIGURE 13. A signal f (solid line) and its delayed version f_s (dashed-dot).

6 Convolution products

We can multiply functions coordinate-wise. However, there is another type of multiplication, the convolution product (see §6.1), that is extremely important in practice. Although, this multiplication looks complicated, the Fourier transform transforms it to coordinate-wise multiplication (see Theorem 6.4).

All integrals in this section run from $-\infty$ to ∞.

If f is a function and $s \in \mathbb{R}$, then the *translated*, *shifted*, or *delayed* function f_s (see Fig. 13) is defined by

$$f_s(t) \equiv f(t-s) \qquad (t \in \mathbb{R}):$$

the graph of f_s is precisely the graph of f but translated (or shifted) over s, or, if f is a signal in time, then f_s is the same signal, but with a delay s.

6.1 Convolution products. For f and h from a large class of functions on $\mathbb{R}$, new functions $f{*}h$ on $\mathbb{R}$ can be defined by

$$f{*}h(t) = \int_{-\infty}^{+\infty} f(t-s)\,h(s)\,\mathrm{d}s = \int_{-\infty}^{+\infty} f(s)\,h(t-s)\,\mathrm{d}s \quad \text{for} \quad t \in \mathbb{R}. \tag{107}$$

$f{*}h$ is the *convolution product* of f and h.

6.2 Example. If $h(t) = \frac{1}{T}$ for $t \in [0,T]$ and $h(t) = 0$ elsewhere, then $f{*}h(t) = \frac{1}{T}\int_{t-T}^{t} f(s)\,ds$ for all $t \in \mathbb{R}$: $f{*}h(t)$ is an average over the f-values in the time interval $[t-T,t]$ of length T prior to t.

In general, the function value $f{*}h(t)$ can be viewed as the average of the delayed f-values, $f(t-s)$, where the f-values have been weighted with weights, $h(s)$, that depend on the delay, s: h is the weight function. Such a situation occurs in practice when, for instance, an acoustic or electromagnetic signal f is broadcast. The signal will be received with a delay, say s_0. The delay depends on the speed with which the signal travels and the distance from source to receiver. Of course, the received signal will be reduced in strength, so $f(t-s_0)\,h(s_0)$ will be received at time t, where $h(s_0)$ is the damping factor. In addition, there will be echoes. The echoes will have to travel a longer distance and will be less energetic than the 'primary' signal: one echo will contribute to the received signal at time t with $f(t-s_1)\,h(s_1)$ where $s_1 > s_0$ and (probably) $|h(s_1)| < |h(s_0)|$. If there is a range of echoes, then the signal that is received at time t is equal to $\int f(t-s)\,h(s)\,\mathrm{d}s$, where the integral represents the superposition of the echoed values. In practice, the received signal will, in addition, be polluted with noise. Therefore, the received signal will be of the form $f*h+n$, where n represents the noise.

Obviously, the convolution product (107) is well-defined in case $f \in L^1(\mathbb{R})$ and $h \in L^\infty(\mathbb{R})$, but also if both f and h are in $L^2(\mathbb{R})$. Then, we have the following estimates

$$\|f*h\|_\infty \le \|f\|_1\,\|h\|_\infty \quad \text{and} \quad \|f*h\|_\infty \le \|f\|_2\,\|h\|_2, \tag{108}$$

respectively. To prove the last estimate, apply the Cauchy–Schwartz inequality. The convolution product $f*h$ for $f \in L^1(\mathbb{R})$, $h \in L^\infty(\mathbb{R})$ or both f and h in $L^2(\mathbb{R})$ are not only bounded, but also continuous.

6.3 Theorem.
If $f \in L^1(\mathbb{R})$ and h is bounded, then $f*h$ is a uniformly continuous function on $\mathbb{R}$. The convolution $f*h$ is also uniformly continuous if $h, f \in L^2(\mathbb{R})$.

Sketch of a proof. We leave the details to the reader; see Exercise 6.11.

Suppose $f \in L^1(\mathbb{R})$. Then, for $\varepsilon > 0$, there is a smooth function $\widetilde{f}$ such that $\|f-\widetilde{f}\|_1 < \varepsilon$ (see §1.8). It is easy to see that $\widetilde{f}*h$ is smooth as well. The first estimate in (108) teaches us that $\|f*h-\widetilde{f}*h\|_\infty = \|(f-\widetilde{f})*h\|_\infty < \varepsilon\|h\|_\infty$. Apparently, there are smooth functions $(\widetilde{f}*h)$ that are arbitrarily close to $f*h$ in the sup-norm. This implies that $f*h$ itself is smooth.

The statement for f and h in $L^2(\mathbb{R})$ can be proved with similar arguments. □

If we do not insist on defining a function point-wise, but if we also accept definitions by means of converging sequence of functions (as in 3.11), then $f*h$ is also defined for, for instance, $f \in L^1(\mathbb{R})$ and $h \in L^1(\mathbb{R}) \cup L^2(\mathbb{R})$; then

$$f*h \in L^1(\mathbb{R}), \quad \|f*h\|_1 \le \|f\|_1\,\|h\|_1 \qquad \text{if } h \in L^1(\mathbb{R}) \tag{109}$$

and

$$f*h \in L^2(\mathbb{R}), \quad \|f*h\|_2 \le \|f\|_1\,\|h\|_2 \qquad \text{if } h \in L^2(\mathbb{R}). \tag{110}$$

The statements are quite surprising, because, if $f \in L^1(\mathbb{R})$, then f^2 need not be integrable (as is shown by $f(t) = 1/\sqrt{t}$ for $|t| < 1$ and $f(t) = 0$ elsewhere). So, products of L^1-functions need not be integrable: the integral need not be finite. However, according to the statement in (109), the infinite value in a convolution of two L^1-functions can only occur in a negligible set of points t.

We sketch a proof of (109) using the following fact: *if $f \in L^1(\mathbb{R})$ then*

$$\|f\|_1 = \sup|(f,g)| = \sup|\int f(t)\,\overline{g(t)}\,\mathrm{d}t|,$$

where the supremum is taken over all $g \in L^\infty(\mathbb{R})$ with $\|g\|_\infty \le 1$.

Let both f and h be in $L^1(\mathbb{R})$. For the moment, let us not worry about the existence of $f*h$, but simply estimate its 1-norm using the above fact:

$$\begin{aligned} |(f*h,g)| &\leq \int\int |f(s)|\,|h(t-s)|\,|g(t)|\ \mathrm{d}t\ \mathrm{d}s \\ &= \int |f(s)| \int\ |h(t-s)|\,|g(t)|\ \mathrm{d}t\ \mathrm{d}s \\ &\leq \left(\int |f(s)|\ \mathrm{d}s\right)\left(\sup_s \int\ |h(t-s)|\,|g(t)|\ \mathrm{d}t\right) \\ &= \|f\|_1\,\|h*g\|_\infty \leq \|f\|_1\,\|h\|_1\,\|g\|_\infty. \end{aligned}$$

Hence, $\|f*h\|_1 \leq \|f\|_1\,\|h\|_1$.
Now, define $h_n(t) = h(t)$, if $|h(t)| < n$, and $h_n(t) = 0$, elsewhere. Then h_n is bounded (by n) and $\|h-h_n\|_1 \to 0$ if $n \to \infty$. Since h_n is bounded, $f*h_n$ is well-defined, and since h_n is in $L^1(\mathbb{R})$, we have that $\|f*h_n - f*h_m\|_1 = \|f*(h_n-h_m)\|_1 \leq \|f\|_1\,\|h_n-h_m\|_1 \to 0$ if $n > m \to \infty$. Therefore, there is an L^1 function, which we denote by $f*h$, such that $\|f*h - f*h_n\|_1 \to 0$ if $n \to \infty$.

The proof of (110) is similar. It exploits the fact that

$$\|f\|_2 = \sup\left\{|(f,g)|\,\Big|\, g \in L^2(\mathbb{R}),\ \|g\|_2 \leq 1\right\} \quad (f \in L^2(\mathbb{R})).$$

We learned from Theorem 6.3 that convolution tends to smooth functions. However, we emphasize that $f*h$ need *not* be continuous if $f \in L^1(\mathbb{R})$ and h in $L^1(\mathbb{R})$ or in $L^2(\mathbb{R})$ see Exercise 6.12. The arguments in the proof of Theorem 6.3 are not applicable to these functions for the following reason. In (108), we are dealing with the sup-norm, whereas the estimates in (109) and (110) involve the 1- and the 2-norm. If a function is arbitrarily close in the sub-norm to continuous functions, then that function itself must be continuous, but this need not to be the case if closeness is measured in 1- or 2-norm.

The estimates in (108), (109), and (110), are instances of a more general result. We give this more general result now since it may be easier to memorize:
if $p, q, r \in [1,\infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$, then

$$f*h \in L^r(\mathbb{R}), \quad \|f*h\|_r \leq \|f\|_p\,\|h\|_q \quad \text{if } f \in L^p(\mathbb{R}),\ h \in L^q(\mathbb{R}) \tag{111}$$

Here, $f \in L^p(\mathbb{R})$ for $p \in [0,\infty)$ if $\|f\|_p \equiv \left(\int |f(t)|^p\ \mathrm{d}t\right)^{\frac{1}{p}}$ is finite.
If $r = \infty$, i.e., $\frac{1}{p} + \frac{1}{q} = 1$, then $f*h$ is continuous.

Fourier transform maps the convolution product to the standard coordinate-wise product.

6.4 Theorem. *If $f, h \in L^1(\mathbb{R}) \cup L^2(\mathbb{R})$, then*

$$\widehat{f*h} = \widehat{f}\,\widehat{h}. \tag{112}$$

Sketch of a proof.

$$\begin{aligned} \widehat{f*h}(\omega) &= \int f*h(t)\,e^{-2\pi i t\omega}\ \mathrm{d}t = \int\int f(s)\,h(t-s)\,e^{-2\pi i t\omega}\ \mathrm{d}t\ \mathrm{d}s \\ &= \int\int f(s)\,e^{-2\pi i s\omega}\,h(t-s)\,e^{-2\pi i(t-s)\omega}\ \mathrm{d}t\ \ \mathrm{d}s \\ &= \int \left(\int h(t-s)\,e^{-2\pi i(t-s)\omega}\ \mathrm{d}t\right) f(s)\,e^{-2\pi i s\omega}\ \mathrm{d}s \\ &= \int \widehat{h}(\omega)\,f(s)\,e^{-2\pi i s\omega}\ \mathrm{d}s = \widehat{h}(\omega)\,\widehat{f}(\omega). \qquad \square \end{aligned}$$

The statement (112) is also correct if $h \in L^\infty(\mathbb{R})$ and is the Fourier transform of a function in $L^1(\mathbb{R})$: $\widehat{h} \in L^1(\mathbb{R})$.

The convolution product can also be formulated for discrete functions (see Exercise 5.3) and for T-periodic functions (see Exercise 6.13). Here also, the Fourier transform leads to coordinate-wise multiplication.

6.5 Correlation product. The *correlation product* $f \odot h$, defined by

$$f \odot h(t) \equiv \int_{-\infty}^{\infty} \overline{f(s)}\, h(s+t)\, \mathrm{d}s = \int_{-\infty}^{\infty} \overline{f(s-t)}\, h(s)\, \mathrm{d}s = (h, f_t) \qquad (t \in \mathbb{R}),$$

is related to the convolution product (see Exercise 6.3).

This product plays an important role in stochastics. It measures the stochastical dependence, the correlation, of h and f or a translate f_t of f. The *autocorrelation* $f \odot f$ measures the correlation of f and translates f_t of f. If, for instance, f and h are independent in a stochastical sense (uncorrelated), then $(h, f) = \int \overline{f(s)}\, h(s)\, \mathrm{d}s = 0$. If n is a function that represents noise and f is a signal such that $\int f(t)\, \mathrm{d}t = 0$, then n and f, but also n and f_t will be uncorrelated. Therefore, $f \odot n = 0$, and $f \odot (\delta f + n) = \delta(f \odot f)$; see Fig. 14. here, δ is a scalar (scaling factor). This can conveniently be exploited specifically in situations where the noise is large relative to δf.

6.6 Application: radar. This last property is used in practice to measure distances. The distance between two objects can be measured by sending a signal (acoustic or electromagnetic) from one object to the other. The signal will be reflected. The time of arrival of the reflected signal at the source is proportional to the distance between the objects. Unfortunately, the reflected signal will be much weaker than the original one ($\delta \ll 1$) and is often heavily polluted by noise (see the two left pictures in Fig. 14). As a consequence, it will be hard to accurately determine the time of arrival of the reflected signal. By forming the correlation product of the received, reflected, signal with the original signal, the obscuring effect of most of the noise can be annihilated (see the right pictures in Fig. 14). Here we assume that the reflection $\widetilde{f}$ of the source signal f is of the form $\delta f_\tau + n$, where δ is a damping factor and τ is the delay, that is, the time that the reflected signal needs to return; n represents noise. So, we assume that there are no (significant) echoes in the reflected signal. We are interested in computing τ. Note that $f \odot (\delta f_\tau + n) = \delta(f \odot f)_\tau$. Therefore, if $f \odot f$ is largest at $t = 0$, then $f \odot \widetilde{f}$ will be largest at time τ.

The source signal in the pictures in Fig. 14, is a sine function. The pictures indicate that the autocorrelation can be used to determine time of arrival of the reflected signal. However, determining when the correlation function $f \odot \widetilde{f}$ is largest would be easier if this function were more localized. (This is particularly so if the reflected signal contains echoes from multiple objects.) This can be done by making the initial pulse shorter in duration, but then the energy transmitted —and more importantly, the energy received— would decrease (δ gets smaller), while the magnitude of the noise would not be affected. Another approach is to find a source signal that has more localized autocorrelation. The premier example of such a signal is the *chirped* pulse $\sin \omega t$, in which the frequency ω is itself time-dependent. That is, the frequency changes as a function of time, like a bird's chirp, see Fig. 15.

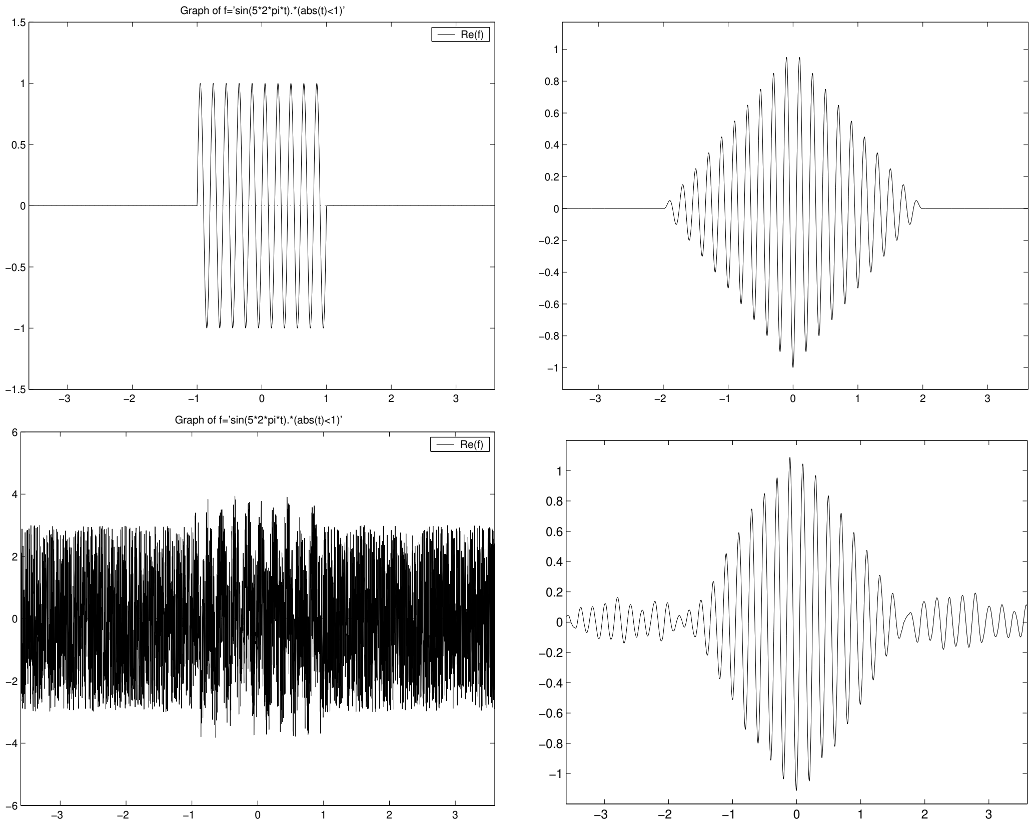


FIGURE 14. The source signal f (top-left) and its autocorrelation function $f \odot f$ (top-right). The received signal $\widetilde{f} = f_\tau + n$ is polluted with noise n. The noise n is large relative to f: $\|n\|_\infty = 3\|f\|_\infty$. The polluted signal is displayed in the bottom-left picture. Here, we took $\tau = 0$. The bottom-right picture shows the correlation product $\widetilde{f} \odot f$ of the received signal with the source signal f. The noise is not completely annihilated in the correlation product $\widetilde{f} \odot f$ (in our computation we used sampled values. This introduces an error, which partially explains why the noise did not completely vanish). Nevertheless, the time of arrival of the pure received signal f_τ, that is, the time where the 'noiseless' received signal f_τ starts to be non-zero, can be determined from $\widetilde{f} \odot f$, which is not possible from $\widetilde{f}$. Notice that the scale along the vertical axis is different in each picture.

6.7 The Wiener–Khintchini theorem. In the frequency domain we have that

$$\widehat{f \odot h} = \overline{\widehat{f}}\,\widehat{h}.$$

In particular, for the *autocorrelation function* $f \odot f$, we have that

$$\widehat{f \odot f} = |\widehat{f}\,|^2.$$

This result is the *Wiener–Khintchini theorem*: the Fourier transform of the autocorrelation function is the power spectrum.

These results allow efficient computation of the correlation product: the FFT can be exploited.

Exercises

Exercise 6.1.
(a) prove that $f * h = h * f$.

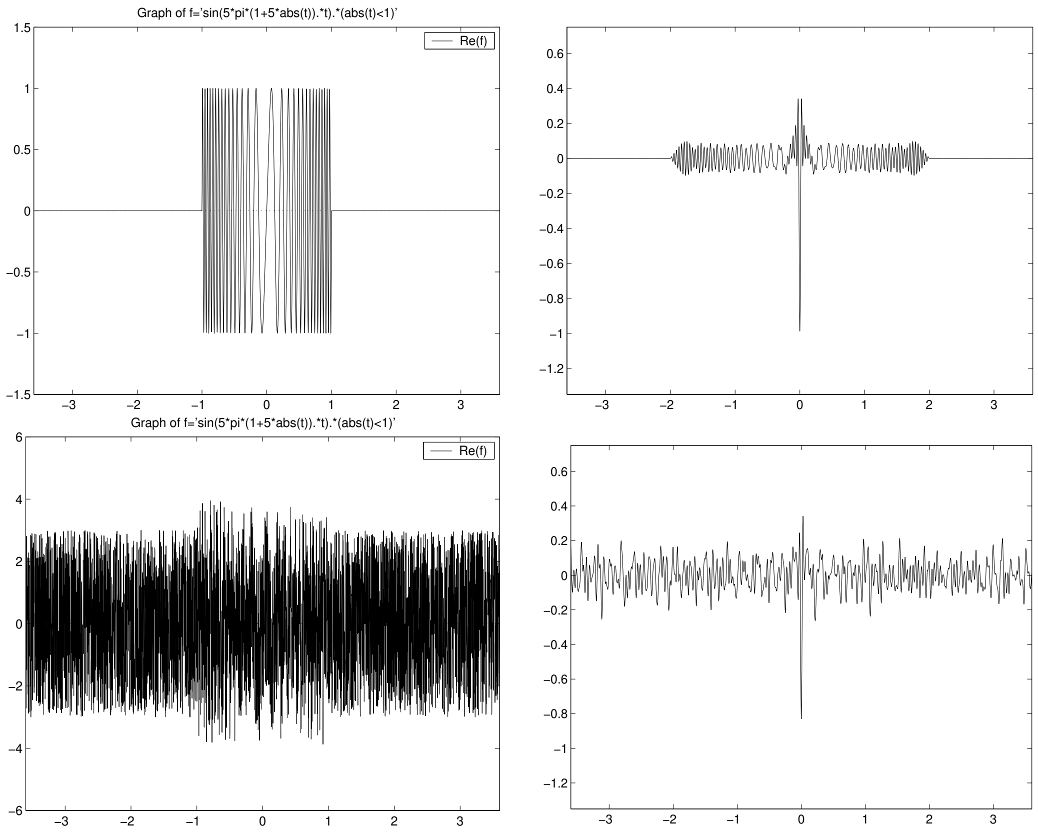


FIGURE 15. For an explanation, see Fig. 14. Here, we replaced the sine pulse from Fig. 14 by a chirped pulse. Note that this approach allows easy detection of the center of the broadcast signal as well as of the received signal.

(b) Prove that $f * h_\tau = f_\tau * h = (f * h)_\tau \quad (\tau \in \mathbb{R})$.

Exercise 6.2. Prove that $f \odot (\delta f_\tau + n) = \delta (f \odot f)_\tau$.

Exercise 6.3. Let $f, g \in L^2(\mathbb{R})$ and $h \in L^1(\mathbb{R})$. Define the *adjoint* h^{T} of h by

$$h^{\mathrm{T}}(t) \equiv \overline{h(-t)} \qquad (t \in \mathbb{R}).$$

Note that real even functions are *self-adjoint*, i.e., $h^{\mathrm{T}} = h$.

(a) Prove that $(f, h * g) = (h^{\mathrm{T}} * f, g)$.

For $h \in L^1(\mathbb{R})$, the convolution $f \rightsquigarrow h * f$ defines an operator from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$. This operator is linear and bounded, with operator norm (less than or equal to) $\|h\|_1$, because $\|h * f\|_2 \leq \|h\|_1 \, \|f\|_2$. In (a), we showed that the adjoint is given by $f \rightsquigarrow h^{\mathrm{T}} * f$. This is the reason, we called h^{T} the adjoint of h; formally adjoints are defined only for operators.

(b) Show that $(f * h)^{\mathrm{T}} = f^{\mathrm{T}} * h^{\mathrm{T}}$.

(c) Show that $\widehat{h^{\mathrm{T}}} = \overline{\widehat{h}}$.

(d) Show that $h \odot f = f * h^{\mathrm{T}} = (h * f^{\mathrm{T}})^{\mathrm{T}} = (f \odot h)^{\mathrm{T}}$.

(e) Combine (c) and (d) to show that $\widehat{h \odot f} = \widehat{f}\, \overline{\widehat{h}}$ (Wiener–Khintchini, see §6.7).

(f) Prove that $\|h \odot f\|_2^2 = (f \odot f, h \odot h)$ and $(h \odot f) \odot (h \odot f) = (f \odot f) * (h \odot h)$.

Exercise 6.4. **Dilation equations.**

Let f and g be in $L^2(\mathbb{R})$ such that

$$f(\tfrac{1}{2}x) = \sum_{j=-\infty}^{\infty} \alpha_j\, f(x-j) \quad \text{and} \quad g(\tfrac{1}{2}x) = \sum_{j=-\infty}^{\infty} \beta_j\, g(x-j) \qquad (x \in \mathbb{R}),$$

for some sequences $\alpha \equiv (\alpha_j)$ and $\beta \equiv (\beta_j)$ in $\ell^2(\mathbb{Z})$.

(a) Prove that

$$(f*g)(\tfrac{1}{2}x) = \sum_{j=-\infty}^{\infty} \tfrac{1}{2}(\alpha*\beta)(j)\,(f*g)(x-j) \qquad (x \in \mathbb{R}),$$

where the convolution product $\alpha*\beta$ of the sequence of scalars is defined by

$$\alpha*\beta(k) \equiv \sum_{j=-\infty}^{\infty} \alpha_j\, \beta_{k-j} = \sum_{j=-\infty}^{\infty} \alpha_{k-j}\, \beta_j \qquad (k \in \mathbb{Z})$$

(see also Exercise 5.4).

Exercise 6.5. **Splines.**

Let B_0 be the function defined by

$$B_0(t) \equiv 1 \quad \text{if} \quad t \in [0,1) \qquad \text{and} \qquad B_0(t) \equiv 0 \quad \text{elsewhere.}$$

(a) Compute $B_1 \equiv B_0 * B_0$ and $B_2 \equiv B_0 * B_0 * B_0$.

(b) Prove that the convolution B_n of $n+1$ copies of B_0 is a *spline of degree* n, where a function f is a *spline of degree* n if its restriction to each of the intervals $[k, k+1)$ is polynomial of degree $\leq n$ $(k \in \mathbb{Z})$ and $f \in C^{(n-1)}(\mathbb{R})$ (here $C^{(0)}(\mathbb{R})$ is the space $C(\mathbb{R})$ of continuous functions and $C^{(-1)}(\mathbb{R})$ is the space of all functions on $\mathbb{R}$ that are left continuous).

(c) Consider an $n \in \mathbb{N}_0$. Show that any finite collection of functions $x \rightsquigarrow B_n(x-k)$, i.e., k belongs to some finite subset of $\mathbb{Z}$, forms a linearly independent set.

(d) Check that $B_0(\frac{1}{2}x) = B_0(x) + B_0(x-1) \quad (x \in \mathbb{R})$.

Use this relation and the results in Exercise 6.4 to show that, for each $x \in \mathbb{R}$,

$$B_1(\tfrac{1}{2}x) = \tfrac{1}{2}B_1(x) + B_1(x-1) + \tfrac{1}{2}B_1(x-2)$$

and

$$B_3(\tfrac{1}{2}x) = \tfrac{1}{8}\left(B_3(x) + 4B_3(x-1) + 6B_3(x-2) + 4B_3(x-3) + B_3(x-4)\right).$$

Exercise 6.6. **Convolutions with elementary 'functions'.**

Let $f \in L^1(\mathbb{R})$.

(a) Show that the convolution product $f * \phi_\omega$ of f and the harmonic oscillation $\phi_\omega(t)$, $\phi_\omega(t) \equiv \exp(2\pi i t\omega)$, is equal to $\widehat{f}(\omega)$:

$$f * \phi_\omega = \widehat{f}(\omega).$$

(b) The convolution product can also be defined for Dirac delta functions or point-measures:

$$f * \delta_\nu(t) \equiv \int_{-\infty}^{\infty} f(t-s)\,\delta_\nu(s)\ \mathrm{d}s = f(t-\nu) = f_\nu(t) \qquad (t \in \mathbb{R}).$$

Show that this definition is consistent with the interpretation of δ_ν as a limit, that is, for $\nu = 0$,

$$f * \delta_0(t) = \lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} f(t-s) \frac{1}{2\varepsilon} \Pi_\varepsilon(s)\ \mathrm{d}s.$$

Exercise 6.7. **Convolution with discrete measures.**

Let $f \in L^1(\mathbb{R})$ and $\mu = \sum \gamma_k \delta_k$ for some sequence $(\gamma_k) \in \ell^1(\mathbb{Z})$ and some sequence (t_k) in $\mathbb{R}$ (see Exercise 3.16).

(a) Show that $f * \mu(t) = \sum \gamma_k \, f * \delta_{t_k}(t) = \sum \gamma_k \, f(t - t_k)$.

(b) Show that $\widehat{f * \mu} = \widehat{f}\,\widehat{\mu}$.

Exercise 6.8. **Shannon–Whittaker.**

Let $\Omega > 0$ and let $f \in L^2(\mathbb{R})$ be such that $\widehat{f}(\omega) = 0$ if $|\omega| > \Omega$. In §7, we will see that the bandwidth restriction on f implies that f is smooth, and, in particular, the function values $f(k\Delta t)$ are well-defined.

For $\Delta t \equiv 1/(2\Omega)$, put $\mu \equiv \sum \Delta t \, f(k\Delta t)\, \delta_{k\Delta t}$ (see Exercise 3.16). Put $\Psi \equiv \widehat{\Pi}_\Omega$.

(a) Show that $\widehat{\Psi * \mu} = \Pi_\Omega\, \widehat{\mu} = \widehat{f}$. Conclude that

$$f(t) = \Psi * \mu(t) = \sum \Delta t \, f(k\Delta t)\, \Psi(t - k\Delta t) = \sum f(k\Delta t)\, \mathrm{sinc}(2t\Omega - k) \quad (t \in \mathbb{R}).$$

Exercise 6.9. **Shannon–Whittaker II.**

Here is a 'quick proof' of Shannon–Whitaker's theorem. Proof and formulation use the Dirac comb $Ш_{\Delta t}$ of Exercise 3.18. f, Ω, Δt are as in Exercise 6.8.

(a) Show that $(f Ш_{\Delta t})\widehat{\ } = \widehat{f} * Ш_{1/\Delta t}$.

Interpret the spectrum of the sampled version of f as the 2Ω periodic extension of the spectrum of f (with spectrum of f being resticted to $[-\Omega, +\Omega]$).

(b) Show that $\widehat{f} = (\widehat{f} * Ш_{1/\Delta t})\, \Pi_\Omega$.

(c) Show that $f = (f Ш_{\Delta t}) * \widehat{\Pi}_\Omega$.

Interpret this result as the Shannon–Whittaker theorem.

Exercise 6.10. **The smoothing effect of convolutions I.**

Let h be the scaled top-hat function $\frac{1}{2\delta}\Pi_\delta$.

(a) Prove that $f*h$ is in $C^{(n+1)}(\mathbb{R})$ if f is bounded and in $C^{(n)}(\mathbb{R})$. Prove that $f*h$ is uniformly continuous if f is bounded.

Exercise 6.11. **The smoothing effect of convolutions II.**

(a) Prove (108).

(b) Prove that $f * h \in C^{(n)}(\mathbb{R})$ if $f \in C^{(n)}(\mathbb{R}) \cap L^1(\mathbb{R})$ and $h \in L^\infty(\mathbb{R})$.

(c) Use the Density Theorem 1.8 to prove that $f * h$ is uniformly continuous if $f \in L^1(\mathbb{R})$ and $h \in L^\infty(\mathbb{R})$.

(d) Use the Density Theorem 1.8 to prove that $f * h$ is uniformly continuous if $f \in L^2(\mathbb{R})$ and $h \in L^2(\mathbb{R})$.

Exercise 6.12. Compute the convolution product of f and h with h and f given by

$$f(t) = 0,\ h(t) \equiv \frac{1}{|t|} \ \text{ if } \ |t| \geq 1 \qquad \text{and} \qquad f(t) = \frac{1}{\sqrt{|t|}},\ h(t) = 0 \ \text{ if } \ |t| < 1.$$

Exercise 6.13.

(a) Formulate a convolution product for T-periodic functions. Derive an expression for the Fourier transform of this product.

(b) For $\alpha = (\alpha_k) \in \ell^2(\mathbb{Z})$, let $\widehat{\alpha}(\omega) \equiv \sum_{k=-\infty}^{\infty} \alpha_k \, e^{-2\pi i \omega k}$. Then $\alpha_k = \int_0^1 \widehat{\alpha}(\omega)\, e^{2\pi i \omega k}\, \mathrm{d}\omega$.

Formulate a convolution product for sequences in $\ell^2(\mathbb{Z})$ and derive an expression for the Fourier transform of this product.

Exercise 6.14. **Heat equation.**

We use the notations and results from Exercise 3.13:

$$u(x,t) = \int_{-\infty}^{\infty} \widehat{\phi}(\omega)\, e^{-4\pi^2\omega^2\gamma t}\, e^{2\pi i\omega x}\ \mathrm{d}\omega$$

solves the heat equation with initial condition $u(x,0) = \phi(x)$.

(a) Prove that $H_t = \widehat{h}_t$, where

$$H_t(\omega) \equiv e^{-4\pi^2\omega^2\gamma t} \quad \text{and} \quad h_t(x) \equiv \tfrac{1}{2\sqrt{\gamma t}}\, e^{-\pi(\frac{x}{2\sqrt{\gamma t}})^2}.$$

(b) Show that $u(x,t) = \phi * h_t(x)$.

(c) Prove that $(x,t) \rightsquigarrow u(x,t)$ is a smooth function on $\mathbb{R} \times (0,\infty)$.

(d) Note that the limit h_t for $t \to 0$ does not exist. Nevertheless, $\phi * h_t$ defines the solution that matches with the initial condition. To understand this, show that

$$\int_{-\infty}^{\infty} g(x)\, h_t(x)\ \mathrm{d}x \to g(0) = \int_{-\infty}^{\infty} g(x)\, \delta_0(x)\ \mathrm{d}x \quad (t \to 0),$$

for each continuous function $g \in L^1(\mathbb{R})$. Here, δ_0 is the Dirac delta function. Hence, in some weak sense, h_t converges to δ_0 for $t \to 0$. Show that $\phi * h_t \to \phi$ if, in addition, ϕ is continuous.

INTERPRETATION. *The solution of the heat equation can be represented as a convolution of the initial heat distribution with a Gaussian function h_t. This clearly represents the spreading of the heat in time. The Dirac delta function δ_0 at $t = 0$ spreads to a Gaussian distribution h_t for $t > 0$ that 'gets wider' with increasing t.*

Exercise 6.15. Wave equation.

We use the notations and results from Exercise 3.12:

$$u(x,t) = \int_{-\infty}^{\infty} \widehat{\phi}_1(\omega)\, \cos(2\pi\omega ct)\, e^{2\pi i\omega x}\ \mathrm{d}\omega + \int_{-\infty}^{\infty} \widehat{\phi}_2(\omega)\, \frac{\sin(2\pi\omega ct)}{2\pi\omega c}\, e^{2\pi i\omega x}\ \mathrm{d}\omega$$

solves the wave equation with initial conditions $u(x,0) = \phi_1(x)$ and $\frac{\partial u}{\partial t}(x,0) = \phi_2(x)$.

(a) Prove that, for $t > 0$,

$$u(x,t) = \phi_1 * (\tfrac{1}{2}[\delta_{ct} + \delta_{-ct}]) + \phi_2 * (\tfrac{1}{2c}\Pi_{ct}).$$

(b) Show that $\phi_1 * (\frac{1}{2}[\delta_{ct} + \delta_{-ct}]) \to \phi_1 * \delta_0 = \phi_1$ $(t \to 0)$ of ϕ_1 is also continuous.

(c) For a continuous function g, consider $\Phi(t) \equiv \int \frac{1}{2c}\Pi_{ct}(x)\, g(x)\ \mathrm{d}x$. Prove that $\Phi'(t) = \frac{1}{2}[g(ct) + g(-ct)]$ for $t > 0$ and conclude that $\Phi'(0) = \lim_{t\to 0} \Phi'(t) = g(0)$.

Show that the partial derivative with respect to t of $\phi_2 * (\frac{1}{2c}\Pi_{ct})$ is equal to $\phi_2 * (\frac{1}{2}[\delta_{ct} + \delta_{-ct}])$ and conclude that this derivative converges to ϕ_2 if $t \to 0$ and ϕ_2 is continuous.

Exercise 6.16. Huygens' principle.

Consider an electromagnetic wave, with electric field E in two space dimensions: $E(x,z)$ is the field at (x,z) at $t = 0$.

Suppose $A(x) \equiv E(x,0)$ $(x \in \mathbb{R})$ is known and suppose the wavenumber k, or, equivalently, the wavelength $\lambda \equiv \frac{1}{k}$ is known. We want to determine $E(x,z)$ from A.

If $A(x) = E_0 \exp(2\pi i k_1 x)$ then

$$E(x,z) = E_0 \exp(2\pi i(k_1 x + k_3 z)), \quad \text{where} \quad k^2 \equiv k_1^2 + k_3^2 = \frac{\omega^2}{c^2}.$$

Since the wavenumber k is know, the frequency ω, and k_3 can be computed: $E(x,z)$ can be determined.

(a) Show that, in the general situation,

$$E(x,z) = \int \widehat{A}(k_1)\, e^{2\pi i(k_1x+k_3z)}\ \mathrm{d}k_1 = \int \widehat{A}(k_1)\, H(k_1)\, e^{2\pi i k_1 x}\ \mathrm{d}k_1,$$

where the transfer function H is given by $H(k_1) = \exp(2\pi i k_3 z)$ with $k_1 = \sqrt{k^2 - k_1^2}$. Hence,

$$E(X,z) = A * h(X) = \int A(x)\, h(X-x)\ \mathrm{d}x, \tag{113}$$

where $h(x) = \int H(k_1)\exp(2\pi i k_1 x)\ \mathrm{d}k_1$. Note that H and, therefore, h, depend on z.
(b) Assume that $k_1 \ll k$ (i.e., the wave $E_0\, e^{2\pi i(k_1x+k_3z)}$ moves almost parallel to the z-axis). Then, show that $k_3 = k\sqrt{1-\frac{k_1^2}{k^2}} \approx k - \frac{k_1^2}{2k}$. Hence,

$$H(k_1) \approx e^{2\pi i k z}\ e^{-\pi i \frac{z}{k} k_1^2}.$$

Show that, if, in addition, $x \ll z$, then

$$h(x) = h_z(x) \approx \sqrt{\frac{k}{iz}}\, e^{2\pi i k z}\ e^{\pi i \frac{k}{z} x^2} \approx \sqrt{\frac{k}{iz}}\, e^{2\pi i k\sqrt{z^2+x^2}} \approx \sqrt{\frac{1}{i\lambda}}\, \frac{e^{2\pi i k r}}{\sqrt{r}}, \tag{114}$$

where $r \equiv \sqrt{x^2+z^2}$ is the distance from (x,z) to the origin $(0,0)$.
(c) The function $(x,z) \rightsquigarrow \exp(2\pi i k r)/\sqrt{r}$ in (114) represents a cylindrical wave emitted from the origin $(0,0)$. Apparently, h_z approximately represents a cylindrical wave, in case $\vec{k} = (k_1,k_3) \approx (0,k_3)$ and $x \ll z$.

Similarly, if in three dimensional space $A(x,y) = E(x,y,0)$ is known, then $E(X,Y,z)$ can be computed from the two-dimensional variant of (113), where

$$h(x,y) = h_z(x,y) \approx \frac{k}{iz}\, e^{2\pi i k z}\ e^{\pi i \frac{k}{z}(x^2+y^2)} \approx \frac{1}{i\lambda}\, \frac{e^{2\pi i k r}}{r}.$$

Here $r \equiv \sqrt{x^2+y^2+z^2}$: h approximately represents a spherical wave (here also assuming that the direction of the wave has a small angle with the z-axis and $\sqrt{x^2+y^2}$ is small relative to z). Formulate the two-dimensional variant of (113).

Now, use (113) and (114) to interpret $E(X,z)$ (and $E(X,Y,z)$) as a superposition of approximate cylindrical (spherical) waves (Huygens' principle).

T-periodic functions.

A convolution product can also be defined for T-periodic functions f and h

$$f * h(t) = \frac{1}{T}\int_0^T f(t-s)\, h(s)\ \mathrm{d}s = \frac{1}{T}\int_0^T f(s)\, h(t-s)\ \mathrm{d}s \quad \text{for} \quad t \in \mathbb{R}.$$

Note that the integral here has been adapted to the inner-product. For this convolution product we have statements similar to ones in the Theorems 6.3 and 6.4.

We will consider T-periodic functions and this convolution product in the following three exercises, where we prove some of the results from §2 that we stated there without proof. The proofs exploit convolution products.

Exercise 6.17. **Approximate identity.**

Consider a sequence (K_n) of functions in $L^1_T(\mathbb{R})$ for which
(i) $\sup_n \|K_n\|_1 \le M < \infty$,
(ii) $\frac{1}{T}\int_0^T K_n(t)\ \mathrm{d}t = 1$ for all $n \in \mathbb{N}$,
(iii) for each $\delta > 0$ we have that $\lim_{n\to\infty} \int_{\delta \le |t| \le \frac{1}{2}T} |K_n(t)|\ \mathrm{d}t = 0$.

We will show that

$$\lim_{n\to\infty} \|K_n * f - f\|_\infty = 0 \quad \text{for each } T\text{-periodic } f \in C(\mathbb{R}) \tag{115}$$

For this reason, a sequence (K_n) in $L^1_T(\mathbb{R})$ with these three properties is called an *approximate identity.*

(a) Show that, for each $\delta > 0$, we have that

$$|K_n * f(t) - f(t)| \le I_1 + I_2,$$

where

$$I_1 \equiv \frac{1}{T}\int_{|s|\le\delta} |K_n(s)|\,|f(t-s) - f(t)|\ \mathrm{d}s, \quad I_2 \equiv \frac{1}{T}\int_{\delta<|s|\le\frac{1}{2}T} |K_n(s)|\,|f(t-s) - f(t)|\ \mathrm{d}s.$$

(b) Prove that, for each $\varepsilon > 0$, there is a $\delta > 0$ such that $I_1 \le \varepsilon M$.
(Hint: f is uniformly continous. Why?)

(c) Prove that, for any $\delta > 0$, $I_2 \to 0$ if $n \to \infty$.

(d) Prove (115).

(e) Show that (115) also holds if (ii) is replaced by the weaker condition

(ii′) $\lim_{n\to\infty} \frac{1}{T}\int_0^T K_n(t)\ \mathrm{d}t = 1$.

Exercise 6.18. Dirichlet kernels.
Put

$$D_n(t) = \sum_{|k|\le n} \exp(2\pi t\frac{k}{T}).$$

(a) Show that

$$S_n(f) = D_n * f \quad \text{for all } \ f \in L^1_T(\mathbb{R}). \tag{116}$$

D_n is the so-called *Dirichlet kernel* of order n (see Exercise 2.20).

(b) Show that (see (54))

$$D_n(t) = \frac{\sin(\pi\frac{t}{T}(2n+1))}{\sin(\pi\frac{t}{T})}.$$

(c) Show that (see (53))

$$\frac{1}{T}\int_0^T D_n(t)\ \mathrm{d}t = 1.$$

(d) Show that there are constants $\kappa_1 \ge \kappa_2 > 0$ such that (see (55))

$$\kappa_2 \log(n) \le \|D_n\|_1 \le \kappa_1 \log(n+1) \quad \text{for all } n.$$

Is (D_n) an approximate identity (see Exercise 6.17)?

The proportionality of $\|D_n\|_1$ with $\log(n)$ is the reason for a lot of hard work and 'nasty' estimates in the convergence theory for Fourier series for functions f that are not very smooth.

Exercise 6.19. Cesàro sums and Fejér kernels.
For $f \in L^1_T(\mathbb{R})$ consider the trigoniometric polynomial

$$\sigma_n(f) \equiv \frac{1}{n+1}\sum_{j=0}^{n} S_j(f).$$

The function $\sigma_n(f)$ is the average of the first $n+1$ partial Fourier series $S_0(f), \ldots, S_n(f)$: it is the so-called *nth Cesàro sum.*

In this exercise, we will see that

$$\lim_{n\to\infty} \|\sigma_n(f) - f\|_\infty = 0 \quad \text{for all continous } f \in L^1_T(\mathbb{R}). \tag{117}$$

Here, we have uniform convergence for any continous T-periodic function f, while for $(S_n(f))$ uniform convergence requires some additional smoothness for f (as we know from the examples of du Bois-Reymond, see the discussion on Theorem 2.4, and of Exercise 2.22).

Define

$$F_n(t) \equiv \frac{1}{n+1}\sum_{j=0}^{n} D_j(t).$$

We will see that these so-called *Fejér kernels* form an approximate identity (see Exercise 6.17).

(a) Show that

$$\sigma_n(f) = F_n * f.$$

(b) With $\zeta \equiv \exp(2\pi i\frac{t}{T})$, show that

$$(n+1)\,F_n(t) = \frac{1}{\zeta^{\frac12} - \bar{\zeta}^{\frac12}} 2\,\mathrm{Im}\left(\sum_{j=0}^{n} \zeta^{j+\frac12}\right) = \frac{2\,\mathrm{Re}(\zeta^{n+1}-1)}{(\zeta^{\frac12} - \bar{\zeta}^{\frac12})^2} = \frac{\sin^2(\pi\frac{t}{T}(n+1))}{\sin^2(\pi\frac{t}{T})} \geq 0.$$

(c) Show that

$$\|F_n\|_1 = \frac{1}{T}\int_0^T F_n(t)\;\mathrm{d}t = \frac{1}{n+1}\sum_{j=0}^{n}\frac{1}{T}\int_0^T D_j(t)\;\mathrm{d}t = 1.$$

(d) Show that (F_n) is an approximate identity and prove (117).

Exercise 6.20. **Jordan's test.**

In this exercise we will show that, for $f \in L^1_T(\mathbb{R})$ and $x \in \mathbb{R}$,

$$\lim_{n\to\infty} S_n(f)(t) = \tfrac12\left(f(t+) + f(t-)\right), \tag{118}$$

if $f(t+)$ and $f(t-)$ exist and, for some $\delta > 0$, f is of bounded variation on $(t-\delta, t+\delta)$ (Jordan's test; see (b) of Th. 2.4 and the discussion following this theorem).

(a) Take $T = 2\pi$. For $t \in (-\pi, \pi]$, let $g(s) \equiv \frac12(f(t+s) + f(t-s))$ $(s \in \mathbb{R})$. Show that D_n and g are even and that (see (116))

$$\begin{aligned} S_n(f)(t) &= \tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi} D_n(s)\,f(t-s)\;\mathrm{d}s = \tfrac{1}{2\pi}\int_{-\pi}^{\pi} D_n(s)\,f(t+s)\;\mathrm{d}s \\ &= \tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi} D_n(s)\,g(s)\;\mathrm{d}s = \tfrac{1}{\pi}\int_0^{\pi} D_n(s)\,g(s)\;\mathrm{d}s. \end{aligned}$$

To prove (118) it suffices to show that

$$\lim_{n\to\infty}\frac{1}{\pi}\int_0^{\pi} D_n(s)\,g(s)\;\mathrm{d}s = g(0+). \tag{119}$$

Since $\frac{1}{\pi}\int_0^{\pi} D_n(s)\;\mathrm{d}s = 1$ (see (53)), it suffices to show (119) in case $g(0+) = 0$ and g is non-decreasing.

Therefore, assume that g is non-decreasing, $g(0+) = 0$, and, for $\delta > 0$, consider

$$I_1 \equiv \frac{1}{\pi}\int_0^{\delta} D_n(s)\,g(s)\;\mathrm{d}s \quad \text{and} \quad I_2 \equiv \frac{1}{\pi}\int_{\delta}^{\pi} D_n(s)\,g(s)\;\mathrm{d}s.$$

(b) Show that

$$\lim_{n\to\infty} I_2 = 0.$$

(Hint: Consider the 2π-periodic function G that is 0 on $(-\pi,\pi]$ except for $2s\in[\delta,\pi]$, where $G(s)\equiv g(2s)/|\sin(s)|$. Then,
$I_2=\frac{1}{\pi}\int_{-2\pi}^{2\pi}G(\frac{1}{2}s)\,\sin(s(n+\frac{1}{2}))\;\mathrm{d}s=\frac{2}{\pi}\int_{-\pi}^{\pi}G(s)\,\sin(s(2n+1))\;\mathrm{d}s=2\beta_{2n+1}(G)$.
Now, use the Riemann–Lebesgue lemma.)

(c) To prove that $|I_1|\leq\varepsilon$, first consider the case where $g(s)=\int_0^s h(\tau)\;\mathrm{d}\tau$ for some $h\in L^1([0,\pi])$, $h\geq 0$. Now, prove that

$$|I_1|\leq 4\int_0^\delta |h(s)|\;\mathrm{d}s=4\,g(\delta).$$

(Hint: With $\mathcal{D}_n(s)\equiv\int_0^s D_n(\tau)\;\mathrm{d}\tau$ we have that $|\mathcal{D}_n(s)|<2\pi$ for all $s\in[0,\pi]$, see (56)), and conclude that (119) holds in this case.

The 2nd mean theorem of integral calculus leads to the same estimate for I_2 also in case g is only non-decreasing on $[0,\delta]$ and $g(0+)=0$: then, $|I_2|\leq 4\,g(\delta-)$.

(d) Adapt the above arguments to prove that

$$\lim_{n\to\infty}\|f-S_n(f)\|_\infty=0$$

or each T-periodic function f that is continuous and of bounded variation on $[0,T]$ (see (c) of Th. 2.4 and the discussion following the theorem).

7 Signals of bounded bandwidth

We use the terminology of Par. 3.16.

If f is a signal, then $\{|\omega| \,\big|\, \widehat{f}(\omega) \neq 0\}$ is the *frequency band* of f. The signal f has a *bounded bandwidth* if there exists an $\Omega > 0$ for which the band of f is a subset of $[0, \Omega]$. The 'smallest' Ω for which this is the case, is the *bandwidth* of f.

The signals that play a role in practice all appear to have bounded bandwidth:

- The male voice does not contain frequencies of 8000 Hertz or higher ($\Omega \leq 8000$).
- A symphony orchestra produces music with bandwidth less than 20000 Hertz.
- The frequencies in TV technology are below $2\,10^6$ Hertz.

Devices that manipulate or transfer signals tend to strongly damp high frequent oscillations in a signal. As a consequence, these devices change any signal into a signal of which the bandwidth is more or less bounded. Depending on the device and the bandwidth, signals of bounded bandwidth can be well manipulated.

Signals of bounded bandwith are of interest in technology since they are natural and can easily be manipulated. In the sequel of this paper, we consider some mathematical aspects of technical questions concerning signals of bounded bandwidth.

As the following theorem shows, signal of bounded bandwidth are very beautiful from a mathematical point of view.

7.1 Theorem. *A signal f of bounded bandwidth is an analytic function* (i.e. $f \in C^{(\infty)}(\mathbb{R})$ and f has a converging Taylor series in each $t_0 \in \mathbb{R}$).

Proof. Let f be a signal with bandwidth $\leq \Omega$.

We will show that $f \in C^{(\infty)}(\mathbb{R})$ and

$$|f^{(n)}(t)| \leq \|f\|_2\ (2\pi\Omega)^{n+\frac{1}{2}} \qquad (t \in \mathbb{R}, n \in \mathbb{N}).$$

Then, with

$$\begin{aligned} f(t) &= f(t_0) + \frac{(t-t_0)}{1!} f^{(1)}(t_0) + \ldots + \frac{(t-t_0)^{n-1}}{(n-1)!} f^{(n-1)}(t_0) + R_n, \\ \text{where}\quad R_n &= \frac{(t-t_0)^n}{n!} f^{(n)}(\xi) \ \ \text{for some } \xi \text{ between } t \text{ and } t_0, \end{aligned}$$

the nth order Taylor expansion of f in t around t_0, we have that

$$|R_n| \leq |t-t_0|^n \frac{(2\pi\Omega)^{n+\frac{1}{2}}}{n!}\ \|f\|_2.$$

The $n!$ in the denominator implies that, for any $t \in \mathbb{R}$, $\lim_{n\to\infty} R_n = 0$.

Apply Cauchy–Schwartz and Plancherel's formula to find that

$$\|\widehat{f}\|_1 = \int_{-\Omega}^{\Omega} |\widehat{f}(\omega)| \cdot 1\ \mathrm{d}\omega \leq \|\widehat{f}\|_2 \sqrt{2\Omega} = \|f\|_2 \sqrt{2\Omega}.$$

Hence, $\|\widehat{f}\|_1 < \infty$ and a combination of 3.12 and 3.4 (interchanging the roles of t and ω) tells us that f is uniformly continuous.

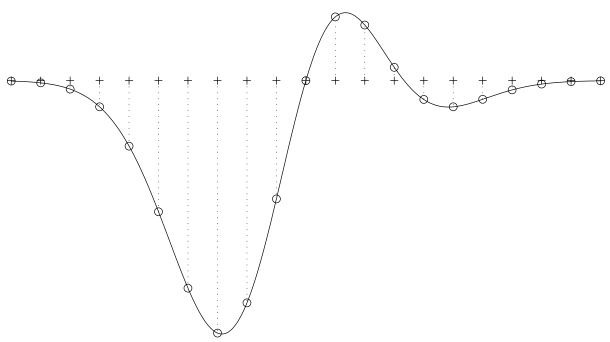


FIGURE 16. A function (solid line) sampled at $t_n = n\Delta t$ (marked with $+++$), with $\Delta t = 0.15$. The sampled function values are marked with $\circ\circ\circ$.

In addition, we have that

$$\frac{d^n f}{dt^n}(t) = \int_{-\Omega}^{+\Omega} \frac{\partial^n}{\partial t^n}(\widehat{f}(\omega)e^{2\pi i\omega t})\, d\omega = \int_{-\Omega}^{+\Omega} \widehat{f}(\omega)(2\pi i\omega)^n e^{2\pi i\omega t}\, \mathrm{d}\omega$$

Hence,

$$|f^{(n)}(t)| \leq \int_{-\Omega}^{+\Omega} |\widehat{f}(\omega)|\, |2\pi\omega|^n\, \mathrm{d}\omega \leq \int_{-\Omega}^{+\Omega} |\widehat{f}(\omega)|\, |2\pi\Omega|^n\, \mathrm{d}\omega \leq \|\widehat{f}\|_1 (2\pi\Omega)^n. \quad \square$$

Signals of bounded bandwidth are so smooth that they can not last in finite time.

7.2 Corollary. *Let f be a signal of bounded bandwidth.*
If $f = 0$ on an interval $[t_1, t_2]$ with $t_1 < t_2$, then $f = 0$ on $\mathbb{R}$.

Proof. Consider a $t_0 \in (t_1, t_2)$. Then $f^{(n)}(t_0) = 0$ for each $n \in \mathbb{N}$, whence the Taylor series of f around t_0 equals 0. Therefore, $f(t) = 0$ for all t. $\square$

Apparently, any spoken word is around until eternity (if it has bounded bandwidth). Common sense tells us that this is not the case. As an explanation, we can mention that each signal carries energy that is so low after a while that the signal is not detectable anymore for common humans nor for devices. (If f is a signal, i.e. $\int_{-\infty}^{+\infty} |f(t)|^2\, \mathrm{d}t < \infty$, and $\varepsilon > 0$, then there is a $T > 0$ such that $\int_{|t|>T} |f(t)|^2\, dt < \varepsilon$. If ε is the lowest energy level for which we can detect a signal, then we fail to detect the signal for $T > 0$).

7.A Sampled signals

In digital signal processing, one would like to reconstruct the signal from a discrete series of signal values (see Fig. 16). According to the following theorem, this is possible provided that the signal is of bounded bandwidth and the function values are '*sampled*' at a uniform speed with a frequency at least twice the bandwith.[19]

In general L^2 functions will not be continuous and the value of a function in a specific point may not be well-defined. However, this problem does not occur for signals of bounded bandwith, since they are very smooth.

[19] Devices that sample function values are called *analog to digital converters*. Since the point of digital signal processing is usually to measure or filter continuous real world analog signals, an analog to digital conversion is usually the first step. The target of the signal processing is often another analog output signal which requires a *digital to analog converter* for translation.

7.3 The sampling theorem of Shannon–Whittaker.
Let f be a signal of bandwidth $\leq \Omega$. Put $\Delta t \equiv \frac{1}{2\Omega}$. Then

$$f(t) = \sum_{k \in \mathbb{Z}} f(k\Delta t)\,\mathrm{sinc}(2t\Omega - k) = \sum_{k \in \mathbb{Z}} f(k\Delta t)\,\mathrm{sinc}\left(\frac{t - k\Delta t}{\Delta t}\right) \quad \text{for all} \quad t \in \mathbb{R}.$$

Convergence is in L^2 sense as well as uniform on bounded intervals.

Proof. The essential idea of the proof is simple. From (b) of 3.7 and Theorem 3.12, we see that $\widehat{\Psi_\Omega} = \frac{1}{2\Omega}\Pi_\Omega$, where $\Psi_\Omega \equiv \mathrm{sinc}(2t\Omega)$. Therefore,

$$\int_{-\infty}^{+\infty} \mathrm{sinc}(2t\Omega - k)\, e^{-2\pi i t\omega}\ \mathrm{d}t =< s = t - k\Delta t >= \exp\left(-2\pi i \frac{k}{2\Omega}\omega\right) \frac{1}{2\Omega}\Pi_\Omega(\omega). \quad (120)$$

If g is the 2Ω-periodic function for which $\widehat{f} = g\Pi_\Omega$, then the Fourier transform of $\widehat{f} = g\Pi_\Omega$ leads the claim in the theorem. The formal proof below uses Lemma 3.8 several times, but with the roles of $-t$ and ω interchanged.

For $n \in \mathbb{N}$, consider $f_n(t) \equiv \sum_{|k| \leq n} f(k\Delta t)\,\mathrm{sinc}(2t\Omega - k)$. By (120), we have that

$$\widehat{f_n}(\omega) = \sum_{|k| \leq n} \frac{1}{2\Omega} f(\frac{k}{2\Omega})\ \exp(-2\pi i \frac{k}{2\Omega}\omega)\,\Pi_\Omega(\omega).$$

By Lemma 3.8, $\lim_{n\to\infty} \|\widehat{f_n} - \widehat{f}\|_2 = 0$, and, with Theorem 3.12, this implies that $\lim_{n\to\infty} \|f_n - f\|_2 = 0$.

Proving convergence in a uniform sense requires an additional argument. Consider a $t \in [-\Delta t, +\Delta t]$. Then $|\mathrm{sinc}(2t\Omega - k)| \leq \min(1, \frac{1}{\pi|k-1|})$. Cauchy–Schwarz and Lemma 3.8 implies, for $m > n > 1$, that

$$\begin{aligned} \sum_{n \leq |k| \leq m} |f(k\Delta t)\,\mathrm{sinc}(2t\Omega - k)| \quad &\leq \quad \sqrt{\sum_k |f(k\Delta t)|^2}\sqrt{\frac{2}{\pi^2}\sum_{k=n}^{m} \frac{1}{(k-1)^2}} \\ &\leq \quad \frac{2}{\pi}\sqrt{\Omega}\|f\|_2\sqrt{\sum_{k=n}^{m} \frac{1}{(k-1)^2}}. \end{aligned}$$

Apparently, the sequence (f_n) converges uniformly on $[-\Delta t, +\Delta t]$. Uniform convergence on any other bounded interval can be proved similarly. Since (f_n) converges in L^2-norm to f and f is continuous, we may conclude that (f_n) converges to f uniformly on any bounded interval. □

The theorem is also correct if $\Delta t \leq 1/(2\Omega)$ (if $\Delta t < 1/(2\Omega)$, then consider $\Omega' \equiv 1/(2\Delta t)$ and note that the bandwidth of f is less than Ω'). The theorem states that the signal f with bandwith $\leq \Omega$ can be reconstructed from a sequence (f_k) of samplevalues $f_k \equiv f(k\Delta t)$, provided that the *sample frequency* $1/\Delta t$ is at least twice the largest frequency present in the signal (see also Fig. 17). The critical value $1/(2\Omega)$ is the so-called *Nyquist rate*.

7.4 Digital signals. If f is of bandwidth $\leq \Omega$, then, with $f_k \equiv f(\frac{k}{2\Omega})$, we have that

$$\widehat{f}(\omega) = \sum_{k=-\infty}^{\infty} \frac{1}{2\Omega}\, f_k\, e^{-2\pi i\omega\frac{k}{2\Omega}} \quad (|\omega| \leq \Omega) \quad \text{and} \quad f_k = \int_{-\Omega}^{\Omega} \widehat{f}(\omega)\, e^{2\pi i\omega\frac{k}{2\Omega}}\ \mathrm{d}\omega. \quad (121)$$

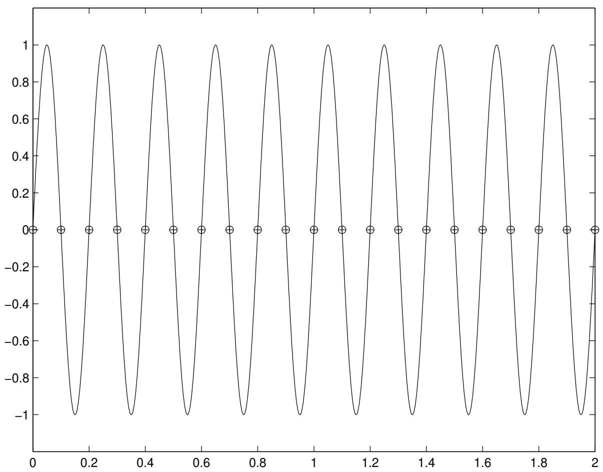


FIGURE 17. The harmonic oscillation $f(t) \equiv \sin(2\pi t\Omega)$ (solid line) with $\Omega = 10$ and the zero function coincide at $t_n \equiv n/(2\Omega)$ (marked with $+++$). The function f has bandwidth Ω, but can clearly not be reconstructed from its sampled values at sample frequency 2Ω. This example seems to contradict the Shannon–Whittaker Theorem. However, note that this function f is not square integrable ($f \notin L^2(\mathbb{R})$). Nevertheless, this example clearly shows what the critical value is of the sample frequency to allow reconstruction from the sampled function values (see also Exercise 7.2). It may help to memorize that the critical sample value is 2 times the bandwidth.

The first equality in (121) is the Fourier series of the 2Ω-periodic extension of the restriction of $\widehat{f}$ to $[-\Omega, +\Omega]$ (with the roles of $-t$ and ω interchanged; cf. Exercise 2.14).

A *digital signal* is a sequence $F = (f_k)$ in $\ell^2(\mathbb{Z})$, i.e., $\sum |f_k|^2 < \infty$. In view of the above arguments, we define the Fourier transform $\widehat{F}(\omega)$ by

$$\widehat{F}(\omega) \equiv \frac{1}{2\Omega} \sum_{k=-\infty}^{\infty} f_k \, e^{-2\pi i \omega \frac{k}{2\Omega}} \qquad (\omega \in \mathbb{R}).$$

Then $\widehat{F}$ is 2Ω-periodic and

$$f_k = \int_{-\Omega}^{\Omega} \widehat{F}(\omega) \, e^{2\pi i \omega \frac{k}{2\Omega}} \, \mathrm{d}\omega \qquad (k \in \mathbb{Z}).$$

Note that it is not clear what value for Ω should be taken if a digital signal F is given, but no information on the sample frequency is available. However, often the precise value for Ω does not play an essential role. Therefore, to simplify notation, one usually takes $\Omega = \pi$ (or $\Omega = \frac{1}{2}$). Then the ω with $\omega \approx 0$ are low frequencies and with $|\omega| \approx \pi$ are high frequencies.

The study of signals in a digital representation (or in discrete time) as above and the processing methods of these signals is called *digital signal processing*.

In practice, sampling is not straight-forward. We mention some of the problems that one may encounter.

7.5 Noise. The signal f will be polluted by some *noise*. Rather than f, we will have $\widetilde{f} \equiv f + \varepsilon$, where ε is some function representing the noise.

Noise does not lead to serious problems if the average energy $\frac{1}{\delta}\int_t^{t+\delta} |\varepsilon(s)|^2 \, ds$ of the noise per time unit $[t, t+\delta)$ is significantly less than the average energy in the signal f in that same time unit. If ε is continuous (unlikely), then $\lim_{\delta\to 0}\frac{1}{\delta}\int_t^{t+\delta}|\varepsilon(s)|^2\,ds=|\varepsilon(t)|^2$ and we will not be bothered by the noise if $|\varepsilon(t)|^2\ll|f(t)|^2$ for all t.

The noise in an audio-signal seems to be undetectable by ear if the *signal-noise ratio* is larger than 90 dB (*decibel*),[20] i.e., $10\,\log_{10}|f(t)|^2-10\,\log_{10}|\varepsilon(t)|^2\geq 90$.[21]

7.6 Approximation error. Signals of bounded bandwidth last forever. On, for instance, a CD only finitely many sample-values can be stored.[22] We have to stop the sample process. The following property can be used to estimate the resulting error.

Property. *For $T>0$, $n>2\Omega T$, and $|t|\leq\Delta t\equiv 1/(2\Omega)$, we have that*

$$\left|f(t)-\sum_{|k|\leq n}f(k\Delta t)\,\mathrm{sinc}(2t\Omega-k)\right|\lesssim\frac{2}{\pi}\sqrt{\frac{1}{2T}\int_{|s|\geq T}|f(s)|^2\,\mathrm{d}s}.$$

Proof. In step (α) of the proof of 7.3, we saw that
$|f(t)-\sum_{|k|\leq n}f(k\Delta t)\,\mathrm{sinc}(2t\Omega-k)|^2\leq(\sum_{|k|>2\Omega T}\Delta t|f(k\Delta t)|^2)\,(\sum_{|k|\geq n}\frac{4\Omega}{\pi^2k^2})$. $\square$

7.7 Aliasing. To determine the Nyquist ratio, an estimate of the bound Ω of the bandwith the signal should be available, which is often not the case. If the estimate for Ω is too small (and, consequently, the Nyquist ratio too large), then a harmonic oscillation $t\rightsquigarrow\widehat{f}(\omega)e^{2\pi it\omega}$ with frequency ω that is too high gets represented an a 'grid' $\{\frac{k}{2\Omega}\,\big|\,k\in\mathbb{Z}\}$ on which the oscillation can not be distinguish from its *alias*, i.e., an oscillation of lower frequency (see Fig. 18):[23]

$$e^{2\pi it\omega}=e^{2\pi it(\omega-2\Omega)}\quad\text{for each } t\in\{\frac{k}{2\Omega}\,\Big|\,k\in\mathbb{Z}\}.$$

This 'alias-phenomenon' can be observed in old western movies where the wheels of the post coach appear to turn backward (the spokes of the wheel have been samples with a frequency of 16 frames per second, while the number of times that a spoke passes a specific point is much larger than 16).

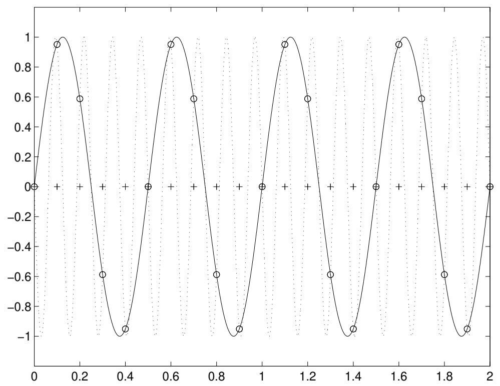
[20] The quality of a signal or of a device in audio-technology is given on the decibel scale. Roughly speaking, this scale corresponds to the sensitivity of our ear: a signal that is twice as strong on the decibel scale as another signal leads to a sound that appears to be twice as loud to us.

[21] The sample values on a CD are represented by 16 bits (sinks and non-sinks) numbers. These numbers will not be the exact function values. However, the 'noise' that is introduced by rounding function values to 16 bits numbers is below the threshold for our ear: the signal-noise ratio is ≥ 90 dB.

[22] The Nyquist ratio on a CD is $\frac{1}{44100}$ second: $\Omega=22500$. A CD can contain about $5\,10^9$ bits in total. A CD player reads 10^6 bits per second (the CD does not only contain the sampled values, but there are also many control bits. For instance, there are additional bits to allow the Reed-Solomon code to correct a few bits).

The digital to analog (DA) converter in a CD player transforms the discrete function of sampled values to a step function: if f is sampled at $t_n=n\Delta t$ $(n\in\mathbb{Z})$, then the DA converter produces $\widetilde{f}$ that is defined by $\widetilde{f}(t)\equiv f(t_n)$ if $t\in[t_n,t_{n+1})$. There is no accurate reconstruction of f! The discontinuities in this 'reconstruction' $\widetilde{f}$ of f seem to be audible. To mask the discontinuities, '*oversampling*' is applied before sending the sampled values to the DA converter, that is, the values of f are computed at $t_{n/2}$ (*one time oversampling*) or also at $t_{n/4}$ (*two times oversampling*). Here $t_r\equiv r\Delta t$. In case of two times oversampling, the DA converter produces $\widetilde{f}$ defined by $\widetilde{f}(t)\equiv f(t_{n/4})$ if $t\in[t_{n/4},t_{(n+1)/4})$. The Shannon–Whittaker Theorem can be exploited to do the oversampling.

[23] To handle this problem as gracefully as possible, most analog signals are filtered with an anti-aliasing filter (usually a low-pass filter; see §8 for definitions) at the Nyquist frequency before conversion to the digital representation.



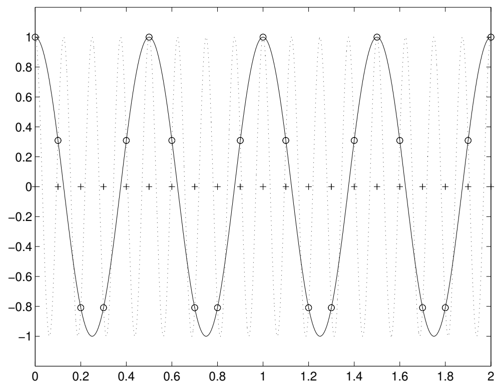


FIGURE 18. An harmonic oscillation and his alias. The sine functions $t \rightsquigarrow \sin(2\pi t\omega)$, with $\omega = \omega_1 = 2$ (solid line) and $\omega = \omega_2 = -8$ (dotted line) in the left picture coincide on $\{k/10 \,\big|\, k \in \mathbb{Z}\}$ (+'s with function values $\circ$). Note that $\omega_1 = \omega_2 + 10$. The cosine functions $\cos(2\pi t\omega)$, with $\omega = \omega_1 = 2$ (solid line) and $\omega = \omega_3 = 8$ (dotted line) in the right picture coincide on $\{k/10 \,\big|\, k \in \mathbb{Z}\}$. Now, $\omega_3 \neq \omega_1 + 10$. But $\cos(\phi) = \cos(-\phi)$ and $\omega_3 = -\omega_1 + 10$.

The following property allows us to estimate the error due to aliasing. We compare f with another nearby function that can be handled by the theory and we interpret the difference as noise.

Property. *Let f_Ω be the signal for which $\widehat{f_\Omega} = \widehat{f}\,\Pi_\Omega$. Then the Shannon–Whittaker theorem is applicable to f_Ω and*

$$|f(t) - f_\Omega(t)| \leq \int_{|\omega|>\Omega} |\widehat{f}(\omega)| \, \mathrm{d}\omega.$$

Proof. See (a) of theorem 3.4. □

7.8 Jitter. The times at which the sampled values are taken will in practice not be exactly equidistant: one will sample at $t_k = k\Delta t + \varepsilon$. The resulting error is called *jitter*.

7.9 Note. Since there is no $T > 0$ for which the sinc function $\mathrm{sinc}(\frac{t}{\Delta t})$ vanishes outside $[-T, +T]$, any sampled value will have an effect on the signal for ever (see Fig. 7).

Numerical mathematics provides algorithms to reconstruct the signal if a number of sample values are lacking (for instance, due to some flaw in material of the CD).[24]

A question that is of interest in Harmonic Analysis (pure mathematics) is whether the collection $\{k\Delta t\}$ can be replaced by another subset of $\mathbb{R}$ that, in some sense, contains 'less' points.

The theorems (and their variants for more dimensions) are also of importance for Crystallography: the points $k\Delta t$ will then represent the positions of atoms in a crystal.

[24] The Reed-Solomon code only corrects a bit if most of the neighboring bits are correct. In case of, for instance, a flaw in the material a whole series of bits will be incorrect. The Reed-Solomon codes can track the error, but they do not provide a correction then.

7.B Information on a signal with bounded bandwith

Communication technologists assume that a signal with bandwidth Ω and time interval $2T$ can carry approximately $4\Omega T$ different messages. The assumption "the dimension of the space of all signals f for which $f = 0$ outside $[-T,+T]$ and $\widehat{f} = 0$ outside $[-\Omega,+\Omega]$ is approximately $4\Omega T$ " looks like the mathematical translation of this assumption (signals f and $5f$ differ only in volume and consequently(?) carry the same information!). Unfortunately, except for the trivial signal, there is no signal with both bounded bandwidth and bounded time support. Since the devices that the engineers design do their job, we may assume that there is another translation. The following theorem might be a suitable candidate.

7.10 Theorem. *Assume $2\Omega T \in \mathbb{N}$. Let f be a signal with bandwidth $\leq \Omega$. Then, there is an $a \in [0, \frac{1}{2\Omega})$ such that, with $f_a(t) \equiv f(t-a)$, we have that*

$$\|f_a - \sum_{|k|\leq 2\Omega T} f_a(\frac{k}{2\Omega})\,\mathrm{sinc}(2\,t\,\Omega - k)\|_2 \leq \sqrt{\int_{|t|\geq T} |f(t)|^2 \;\mathrm{d}t}.$$

Proof. From 7.3 and 3.8 (interchanging roles of t and ω) we conclude that

$$\int_0^{\frac{1}{2\Omega}} \|f_a - \sum_{|k|\leq 2\Omega T} f_a(\frac{k}{2\Omega})\,\mathrm{sinc}(2\,t\,\Omega - k)\|_2^2 \;\mathrm{d}a \leq$$
$$\int_0^{\frac{1}{2\Omega}} \frac{1}{2\Omega} \sum_{|k|>2\Omega T} |f_a(\frac{k}{2\Omega})|^2 \;\mathrm{d}a \leq \frac{1}{2\Omega} \int_{|t|\geq T} |f(t)|^2 \;\mathrm{d}t.$$

For the last estimate, we exploited the observation that $|\frac{k}{2\Omega} - a| \geq |\frac{k}{2\Omega}| - a \geq \frac{2\Omega T+1}{2\Omega} - \frac{1}{2\Omega} = T$ if $|k| > 2\Omega T$. This leads to the statement in the theorem. □

Except for a *delay* a, the signals f and f_a are identical (cf., Fig. 13). If the signal f carries $\varepsilon \equiv \int_{|t|>T} |f(t)|^2 \;\mathrm{d}t$ energy outside the time interval $[-T,+T]$ then the difference between the slightly delayed signal f_a and some signal in the span of the $4\Omega T + 1$ functions $\mathrm{sinc}(2\,t\,\Omega - k)$ $(k = -2\Omega T, -2\Omega T + 1, \ldots, 2\Omega T - 1, 2\Omega T)$ has energy less than ε.

7.C Signal reconstruction

The questions in this subsection are harder to answer than the ones in the two preceding subsections. We will not give details here, but we will indicate what techniques can be used for a proper treatment. Questions and analysis techniques show that also in fields of application of Fourier theory a wide mathematical knowledge and expertise is necessary for successful research.

A signal can not be of bounded bandwidth and at the same time last only for a limited period of time. In spite of this fact, of signals f with bandwidth Ω often only values in a bounded time interval $[-T,+T]$ are know (or f-values from a certain time interval are lacking due to deficiencies in the material or perturbations by noise,...). It is remarkable that a signal f of bounded bandwidth is completely determined by its values on each arbitrary time interval $[t_0, t_1]$ where $t_0 < t_1$, no matter how small the interval is.

7.11 Property. *Let $t_0 < t_1$. If f_1 and f_2 are signals of bounded bandwidth such that $f_1 = f_2$ on $[t_0, t_1]$, then $f_1 = f_2$ on $\mathbb{R}$.*

Proof. Apply 7.2 with $f = f_1 - f_2$. □

Now, one may wonder how to reconstruct f from its values on a bounded time interval. The problem resembles the 'sample problem' of §7.A. However, the answer to the present problem is more complicated and requires results from Hilbert space theory. We look for a solution by means of an orthonormal basis of appropriate functions.[25]

We will use Plancherel's formula (see (73) in §3).

7.12 The space of all signals with bandwidth $\leq \Omega$ is denoted by $\mathcal{B}_\Omega$:

$$\mathcal{B}_\Omega \equiv \{f \in L^2(\mathbb{R}) \,\big|\, f \text{ has bandwidth } \leq \Omega\}.$$

Define $D_T, B_\Omega : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ as follows:

$$D_T(f) \equiv f\Pi_T \quad \text{and} \quad B_\Omega(f) \equiv f * \widehat{\Pi_\Omega} \qquad (f \in L^2(\mathbb{R}).$$

With $g \equiv B_\Omega(f)$, we have that $\widehat{g} = \widehat{f}\,\widehat{\widehat{\Pi}_\Omega} = \widehat{f}\,\Pi_\Omega = D_\Omega(\widehat{f})$. Here, we used that Π_Ω is an even function: $\widehat{\widehat{\Pi}_\Omega}(\omega) = \Pi_\Omega(-\omega) = \Pi_\Omega(\omega)$.

We collect some elementary and useful properties.

7.13 Properties. *Let $f, g \in L^2(\mathbb{R})$. Then, with $(\cdot,\cdot)$ the $L^2(\mathbb{R})$ inner product, we have that*

(i) $(D_T(f), g) = (f, D_T(g)\,)$ *and* $(B_\Omega(f), g) = (f, B_\Omega(g)\,)$,

(ii) $D_T D_T(f) = D_T(f)$ *and* $B_\Omega B_\Omega(f) = B_\Omega(f)$,

(iii) *if* $D_T B_\Omega(f) = 0$, *then* $B_\Omega(f) = 0$, *and if* $B_\Omega D_T(f) = 0$, *then* $D_T(f) = 0$.

Proof. (i) $(D_T(f), g) = \int_{-T}^{T} f(t)g(t)\ \mathrm{d}t = (f, D_T(g)\,)$.

$(B_\Omega(f), g) = (\widehat{B_\Omega(f)}, \widehat{g}) = (D_\Omega \widehat{f}, \widehat{g}) = \ldots$.

(ii) Clearly the first equality is correct. The second one follows by taking the Fourier transform.

(iii) The first claim is a consequence of 7.11, the second one follows from the first claim via the Fourier transform. □

The eigenfunctions of $B_\Omega\, D_T\, B_\Omega$ lead to the desired orthonormal basis.

7.14 An eigenvalue problem.

We are interested in the eigenfunctions ψ in $L^2(\mathbb{R})$ of $B_\Omega\, D_T\, B_\Omega$ with eigenvalue $\lambda \in \mathbb{R}$:

$$B_\Omega\, D_T\, B_\Omega(\psi) = \lambda\,\psi.$$

The following result can be proved with techniques from Hilbert space theory. The representation of the operator $B_\Omega\, D_T\, B_\Omega$ as an integral operator (see (122) below)

[25] Orthonormal basis often play a central role. The fact that the functions $t \rightsquigarrow \exp(2\pi i t \frac{k}{T})$ form an orthogonal basis for the space of the L^2_T-functions is one reason why the theory on Fourier series is so elegant. The Fourier transform of $t \rightsquigarrow \mathrm{sinc}(2\,\Omega\, t - k)$ is precisely $\omega \rightsquigarrow \frac{1}{2\Omega}\exp(2\pi i \frac{k}{2\Omega}\omega)\Pi_\Omega(\omega)$ (see the proof of Shannon–Whittaker's theorem). This and Parseval's formula imply that the functions $t \rightsquigarrow \mathrm{sinc}(2\,\Omega\, t - k)$ form an orthonormal basis of the space of signals of bandwidth $\leq \Omega$. This basis is well-suited for reconstruction from sampled function values. For reconstruction from functions on intervals, we need another orthonormal basis.

implies that the operator is compact. From 7.13 it can be deduced that this operator is selfadjoint and positive definite. The following theorem holds for compact positive definite operators on Hilbert spaces. We will not give a proof here.

7.15 Theorem.
There is a sequence (ψ_n) *in* $L^2(\mathbb{R})$ *and a sequence* (λ_n) *in* $[0,\infty)$ *such that*
(i) $B_\Omega\, D_T\, B_\Omega(\psi_n) = \lambda_n\, \psi_n$ *for all* $n \in \mathbb{N}$,
(ii) *the* ψ_n *form an orthonormal basis for* $L^2(\mathbb{R})$. □

The ψ_n in $\mathcal{B}_\Omega$ will provide us with a basis that can be used to solve our signal reconstruction problem.

7.16 Lemma. *Let* $\mathbb{E} \equiv \{n \in \mathbb{N} \,\big|\, \lambda_n \neq 0\}$.
If $n \in \mathbb{E}$, *then* $B_\Omega(\psi_n) = \psi_n$ *and* $\psi_n \in \mathcal{B}_\Omega$. *If* $n \notin \mathbb{E}$, *then* $B_\Omega(\psi_n) = 0$.

Proof. If $n \notin \mathbb{E}$, then $B_\Omega\, D_T\, B_\Omega(\psi_n) = 0$, and $B_\Omega(\psi_n) = 0$ follows from (iii) in 7.13.

Property (ii) of 7.13 implies that

$$\lambda_n\, B_\Omega(\psi_n) = B_\Omega\, B_\Omega\, D_T\, B_\Omega(\psi_n) = B_\Omega\, D_T\, B_\Omega(\psi_n) = \lambda_n\, \psi_n$$

and $B_\Omega(\psi_n) = \psi_n$ if $\lambda_n \neq 0$. □

7.17 Theorem. *For each* $n \in \mathbb{E}$, *we put* $\widetilde{\psi}_n \equiv \frac{1}{\sqrt{\lambda_n}} D_T(\psi_n)$.
(i) *The functions* ψ_n *with* $n \in \mathbb{E}$ *form an orthonormal basis of* $\mathcal{B}_\Omega$.
(ii) *The functions* $\widetilde{\psi}_n$ *with* $n \in \mathbb{E}$ *form an orthonormal basis of* $\{D_T(f) \,\big|\, f \in \mathcal{B}_\Omega\}$.
If $f \in \mathcal{B}_\Omega$ *and* $f = \sum_{k\in\mathbb{E}} \beta_k \frac{1}{\sqrt{\lambda_k}} \psi_k$, *then* $D_T(f) = \sum_{k\in\mathbb{E}} \beta_k \widetilde{\psi}_k$.

Proof. (i) Let $f \in L^2(\mathbb{R})$.
By 7.15, we have that $f = \sum_{k=1}^{\infty} \alpha_k \psi_k$. Hence, $(f, \psi_n) = \sum_k \alpha_k (\psi_k, \psi_n) = \alpha_n$. If $f \in \mathcal{B}_\Omega$ and $n \notin \mathbb{E}$, then $\alpha_n = (f, \psi_n) = (B_\Omega(f), \psi_n) = (f, B_\Omega(\psi_n)\,) = 0$.
(ii) To show that the $\widetilde{\psi}_n$ form an an orthonormal system, note that, for $n, m \in \mathbb{E}$,

$$\begin{aligned}(D_T(\psi_n), D_T(\psi_m)\,) &= (D_T\, B_\Omega(\psi_n), D_T\, B_\Omega(\psi_m)\,)\\ &= (B_\Omega\, D_T\, D_T\, B_\Omega(\psi_n), \psi_m) = \lambda_n(\psi_n, \psi_m).\end{aligned}$$

Therefore, the orthonormality of the ψ_n implies orthonormality of the $\widetilde{\psi}_n$ $(n \in \mathbb{E})$.

If $f \in \mathcal{B}_\Omega$, then $f = \sum_{k\in\mathbb{E}} \alpha_k \psi_k$ and

$$D_T(f) = \sum_{k\in\mathbb{E}} \alpha_k\, D_T(\psi_k) = \sum_{k\in\mathbb{E}} \alpha_n \sqrt{\lambda_n} \widetilde{\psi}_k.$$

The last claim in the theorem follows with $\beta = \alpha_k \sqrt{\lambda_k}$ and we proved that the $\widetilde{\psi}_n$ form a complete system for $D_T(\mathcal{B}_\Omega)$. □

The functions ψ_n have a double orthogonality property which makes them attractive in many application. For instance, they solve our problem:

7.18 Corollary. *If $g = D_T(f)$ for some $f \in \mathcal{B}_\Omega$, then*

$$f = \sum_{k\in\mathbb{E}} \beta_k \frac{1}{\sqrt{\lambda_k}} \psi_k, \quad \textit{where} \quad \beta_k \equiv \int_{-T}^{T} g(t)\, \widetilde{\psi}_k(t)\ \mathrm{d}t.$$

Proof. If $D_T(f) = D_T(g) = g = \sum_{k\in\mathbb{E}} \beta_k \widetilde{\psi}_k$,
then $\int_{-T}^{T} g(t)\, \widetilde{\psi}_n(t)\ \mathrm{d}t = (g, \widetilde{\psi}_n) = \sum_k \beta_k (\widetilde{\psi}_k, \widetilde{\psi}_n) = \beta_n$. □

Formally we have now solved our signal reconstruction problem. However, before we can apply the above corollary in practice, some difficulties have to be solved. We mention a few obstacles.

In the first place, we have to learn how to deal with the eigenfunctions ψ_n. The observations the following paragraphs will be useful for this.

7.19 The eigenvalue problem revisited. We are interested in the eigenfunctions $\psi \in \mathcal{B}_\Omega$. For such an eigenfunction ψ with eigenvalue λ we have that $B_\Omega(\psi) = \psi$. Therefore,

$$B_\Omega\, D_T(\psi)(t) = (\Pi_T \psi) * \widehat{\Pi_\Omega}(t) = \int_{-T}^{+T} \psi(s)\, \frac{\sin(2\pi(t-s)\Omega)}{\pi(t-s)}\ \mathrm{d}s = \lambda\, \psi(t). \tag{122}$$

With $c \equiv \Omega\, T$, $\phi(y) \equiv \psi(Ty)$, and $t \equiv Ty$, we have that (substitute $s = Tx$)

$$\int_{-1}^{+1} \frac{\sin 2c\pi(y-x)}{\pi(y-x)}\, \phi(x)\ \mathrm{d}x = \lambda\, \phi(y). \tag{123}$$

We see that the eigenfunction ϕ and eigenvalue λ depend only on the product ΩT, on c, and not on the terms T and Ω separately. Note that (123) allows to compute $\phi(y)$ for $|y| > 1$ if $\phi(x)$ is available for $|x| \leq 1$.

The eigenfunction ϕ is also a solution of the differential eigenvalue problem

$$\frac{d}{dx}(1-x^2)\,\frac{d\phi}{dx} + (\frac{1}{\lambda} - 4\pi^2 c^2 x^2)\,\phi = 0.$$

This differential equation plays also a role in the analysis of the wave equation in certain spherical coordinates. Due to this analysis, many properties of the solutions ϕ are well-known (prolete spherical wave functions).

7.20 The values of the eigenvalues. Let $\lambda_1(c), \lambda_2(c), \ldots$ be the eigenvalues λ of the problem (122). We may assume that these eigenvalues are ordered such that

$$\lambda_1(c) \geq \lambda_2(c) \geq \ldots$$

(and that $\lambda_n = \lambda_n(\Omega T)$ for all n, where λ_n is as in 7.15).

The following can be proved (cf. Fig. 19):

- $\lambda_n(c) \approx \frac{1}{2}$ if $n \approx 4c$;
- $\lambda_n(c) \approx 1$ if $n \ll 4c$ (even for $n < 4c - \delta$, where $\delta \approx \log(c)$);
- $\lambda_n(c) \approx 0$ if $n \gg 4c$ (even if $n > 4c + \delta$).

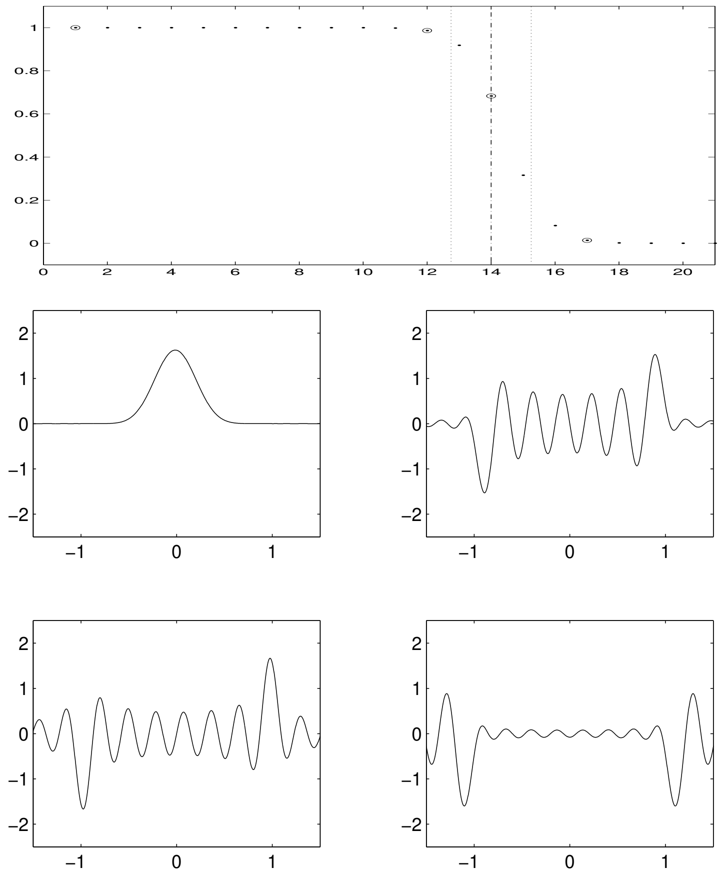


FIGURE 19. The top figure shows the value of the eigenvalues $\lambda_n(c)$ for $c = 3.5$. The dotted lines at $\log c$ of the value $4c$ (dashed-dot) mark the transition area from $\lambda \approx 1$ to $\lambda \approx 0$. The next four figures show the graph of eigenfunctions ϕ_n corresponding to the eigenvalues that are marked (with circles) in the top figure. The most left eigenvalue in the top picture corresponds to the left picture at the second row of pictures, etc.. Note that the eigenfunctions corresponding to $\lambda \approx 1$ are well-concentrated in $[-1, 1]$, whereas the eigenfunction with $\lambda \approx 0$ nearly vanishes on $[-1, 1]$. The eigenfunction with λ in the transition stage has significant values inside as well as outside the interval $[-1, 1]$.

The domain of the ns for which $\lambda_n(c) \not\approx 1$ and $\lambda_n(c) \not\approx 0$ is small. This domain is an interval around $n \approx 4c$ with width $\approx \log c$. Since $(D_T(\psi_n), D_T(\psi_n)) = \int_{-T}^{+T} |\psi_n(s)|^2 \, \mathrm{d}s = \lambda_n$ (cf. 7.17), this implies that for $n < 4c - \delta$, the signals ψ_n are almost completely concentrated in the interval $[-T, +T]$, while for $n > 4c + \delta$ the signal ψ_n have almost no energy in the interval $[-T, +T]$ (cf. Fig. 19. This figure is for $c = 3.5$. For $c = 10$, the $\lambda_1, \ldots, \lambda_{16}$ are up to 15 digits equal to 1, while $\lambda_{35} = 0.99986$: apparently, even λ_{35} is close to 1. Hence, even ψ_{35} is for 99.993% concentrated in $[-T, T]$.).

With the ψ_n and exploiting the insights obtained above, a "$4\Omega T$-theorem" can be formulated, that is, a theorem as in §7.B that expresses how a space of signals with bandwidth Ω and energy that is almost completely concentrated in $[-T, +T]$ can be viewed as the span of $\approx 4\Omega T$ linearly independent signals: the ψ_n for $n < 4\Omega T - \delta$ form an orthonormal system of signals that are almost completely concentrated in $[-T, +T]$.

To conclude this discussion, we mention that $\sum_{n=1}^{\infty} \lambda_n(c) = 4c$.

7.21 Practical objections. If $f \in \mathcal{B}_\Omega$ and f is known on $[-T, +T]$, then, in principle, f can be determined everywhere. Unfortunately, the values of f on $[-T, +T]$

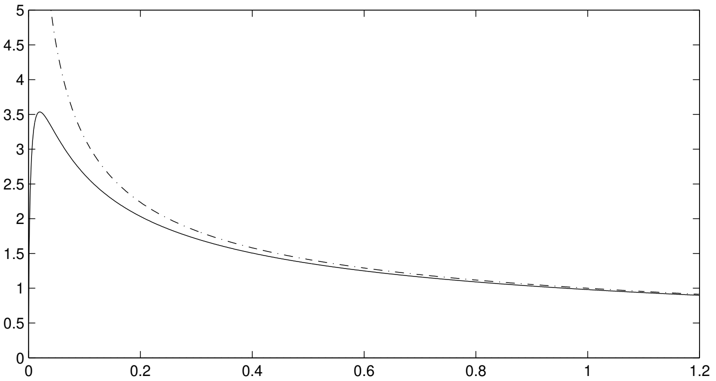


FIGURE 20. The function $\lambda \rightsquigarrow 1/\sqrt{\lambda}$ (dashed-dot) and its regularized variant $\lambda \rightsquigarrow \sqrt{\lambda}(\lambda+\mu)$ for $\mu = 0.02$.

that are available will usually be polluted by errors. Rather than $g = D_T(f)$, we have $\widetilde{g} = g + \varepsilon$, with ε some noise (function in $L^2(\mathbb{R})$) with $D_T(\varepsilon) = \varepsilon$ and $\bar{\varepsilon} \equiv \|\varepsilon\|_2$ small.

If $g = \sum_{k\in\mathbb{E}} \beta_k \widetilde{\psi}_k$, then $f = \sum_{k\in\mathbb{E}} \beta_k \frac{1}{\sqrt{\lambda_k}} \psi_k$ and $\|f\|_2^2 = \sum_{k\in\mathbb{E}} \frac{|\beta_k|^2}{|\lambda_k|} < \infty$, since $f \in L^2(\mathbb{R})$. For $\widetilde{f}$ we will take the function $\sum_{k\in\mathbb{E}} (\beta_k + \delta_k)\frac{1}{\sqrt{\lambda_k}} \psi_k$, with $\delta_k \equiv \int_{-T}^{+T} \varepsilon(t)\, \widetilde{\psi}_k(t)\ \mathrm{d}t$. Since the $\widetilde{\psi}_k$ form an orthonormal system for $k \in \mathbb{E}$, we have that $\sum_{k\in\mathbb{E}} |\delta_k|^2 \leq \bar{\varepsilon}$. However, since ε will not be in $\mathcal{B}_\Omega$, the error $\|f - \widetilde{f}\|_2$ can be unboundedly large, because $\|f - \widetilde{f}\|_2^2 = \sum_{k\in\mathbb{E}} \frac{|\delta_k|^2}{|\lambda_k|}$. The problem is said to be *ill conditioned*.

The produced solution $\widetilde{f}$ itself will be infinitely large in case the error is infinitely large ($\|\widetilde{f}\|_2 \geq \|\widetilde{f} - f\|_2 + \|f\|_2$). This is the reason that in practice a *regularisation* technique is employed: a small value of $\mu > 0$ is selected and the function f in $\mathcal{B}_\Omega$ is computed for which the expression

$$\|D_T(f) - D_T(\widetilde{g})\|_2^2 + \mu\|f\|_2^2$$

is minimal. If f^r is this minimizing f, then f^r can not have an infinite norm $\|f^r\|_2$.

With $r(\lambda) \equiv \frac{\sqrt{\lambda}}{\lambda+\mu}$ for $\lambda > 0$, we actually have that $f^r \equiv \sum_{k\in\mathbf{E}} \widetilde{\beta}_k\, r(\lambda_k)\, \psi_k$, where $\widetilde{\beta}_k = \int_{-T}^{+T} \widetilde{g}(t)\, \widetilde{\psi}_k(t)\ \mathrm{d}t$.

The above misery stems from the fact that $\frac{1}{\sqrt{\lambda_k}}$ is large if λ_k is small. Note that $r(\lambda) \approx \frac{1}{\sqrt{\lambda}}$ for larger λ, while $r(\lambda)$, also for smaller λ, will not be large; see Fig. 20

7.D Uncertainty relations

We have seen that a signal can not have bounded bandwidth and at the same time last only for a limited period of time. Nevertheless, it is important to have signals that are well concentrated both in time as well as in frequency. (Radio stations are close together on the radio band. One would like to have not only 'smooth' signals that transport music and speech, but one also would like to employ short pulses as dots and dashes in telegraphy and dots and non-dots in digital communication.)

The *uncertainty principle of Heisenberg* gives a lower bound on how well a signal can be concentrated both in time and in frequency.

7.22 Let f be a signal. Then

$$E \equiv \int_{-\infty}^{+\infty} |f(t)|^2 \; \mathrm{d}t = \int_{-\infty}^{+\infty} |\widehat{f}(\omega)|^2 \; \mathrm{d}\omega \tag{124}$$

is its *energy*. The *center* of the (energy distribution of) signal f is located at time

$$t_0 \equiv \frac{1}{E} \int_{-\infty}^{\infty} t \, |f(t)|^2 \; \mathrm{d}t. \tag{125}$$

The quantity σ_t, given by

$$\sigma_t^2 \equiv \frac{1}{E} \int_{-\infty}^{\infty} (t - t_0)^2 \, |f(t)|^2 \; \mathrm{d}t = \frac{1}{E} \int_{-\infty}^{\infty} t^2 \, |f(t)|^2 \; \mathrm{d}t - t_0^2, \tag{126}$$

is a measure for the concentration of the energy around the center: if f is well-concentrated around t_0, then σ_f will be small, whereas σ_f is large if f has large values far away from t_0.
Similar definitions can be given in the frequency domain: with

$$\omega_0 \equiv \frac{1}{E} \int_{-\infty}^{\infty} \omega \, |\widehat{f}(\omega)|^2 \; \mathrm{d}\omega, \quad \sigma_\omega^2 \equiv \frac{1}{E} \int_{-\infty}^{\infty} (\omega - \omega_0)^2 \, |\widehat{f}(\omega)|^2 \; \mathrm{d}\omega$$

ω_0 is the center of the spectral energy density, σ_ω is a measure for the spectral energy concentration.

The uncertainty principle of Heisenberg states that a signal can not be concentrated both in time as well as in frequency: σ_t and σ_ω can not both be small at the same time.

7.23 The uncertainty principle of Heisenberg. *Let f be a signal that is sufficiently smooth. Then*

$$\sigma_t \, \sigma_\omega \geq \frac{1}{4\pi}.$$

In addition, we have that

$$\sigma_t \, \sigma_\omega = \frac{1}{4\pi} \quad \Leftrightarrow \quad f(t) = c \, e^{\gamma(t-t_0)^2} \quad (t \in \mathbb{R}) \quad \textit{for some } c \textit{ and } \gamma \in \mathbb{C}, \ \mathrm{Re}(\gamma) < 0.$$

Proof. Without loss of generality, we may assume that $t_0 = 0$ and $\omega_0 = 0$.

We have that $\widehat{f'}(\omega) = 2\pi i \omega \widehat{f}(\omega)$. Cauchy–Schwartz' inequality and Plancherel's formula (see Th. 3.12) imply that

$$\begin{aligned} \left| \int_{-\infty}^{+\infty} t \, f(t) \, f'(t) \; \mathrm{d}t \right|^2 &\leq \int_{-\infty}^{+\infty} t^2 \, |f(t)|^2 \; \mathrm{d}t \cdot \int_{-\infty}^{+\infty} |f'(t)|^2 \; \mathrm{d}t \\ &= 4\pi^2 \int_{-\infty}^{+\infty} t^2 |f(t)|^2 \; \mathrm{d}t \cdot \int_{-\infty}^{+\infty} \omega^2 |\widehat{f}(\omega)|^2 \; \mathrm{d}\omega. \end{aligned}$$

Here, we also used Theorem 3.5. Integration by parts leads to

$$\int_{-\infty}^{+\infty} t f(t) f'(t) \; \mathrm{d}t = t \tfrac{1}{2} f^2(t) \Big|_{-\infty}^{+\infty} - \tfrac{1}{2} \int_{-\infty}^{+\infty} f^2(t) \; \mathrm{d}t = -\tfrac{1}{2} E$$

and Heisenberg's uncertainty relation follows by a combination of these two estimates.

The estimate can only be sharp if the estimate with the Cauchy–Schwartz inequality is sharp. This is the case if and only if the functions $t \rightsquigarrow tf(t)$ and f' are linearly dependent, that is, if there is a $\widetilde{\gamma} \in \mathbb{C}$ such that $f'(t) = \widetilde{\gamma}\, t f(t)$ for all $t \in \mathbb{R}$. Then, $(\log f)'(t) = \widetilde{\gamma} t$, and, therefore, $f(t) = c \exp(\frac{1}{2}\widetilde{\gamma} t^2)$. This f can only be in $L^2(\mathbb{R})$ if $\mathrm{Re}(\gamma) < 0$, where $\gamma \equiv \frac{1}{2}\widetilde{\gamma} < 0$. □

Heisenberg's uncertainty relation is famous from quantum physics. Quantum physicists do not interpret f as a signal but as a quantity that expresses the probability to find a particle in certain *place* t. In this interpretation, $\widehat{f}$ is the probability that the particle has impulse ω. The quantities σ_t and σ_ω are of stochastical origin in this application.[26]

7.24 The stochastical approach by the quantum physicists is less convenient for the communication technology, where one is more interested in the quantity $\alpha_T(f)$ given by

$$\alpha_T(f)^2 \equiv \frac{1}{E}\int_{-T}^{+T} |f(t)|^2 \,\mathrm{d}t \quad \text{with} \quad E \equiv \int_{-\infty}^{+\infty} |f(t)|^2 \,\mathrm{d}t.$$

The quantity $\alpha_T(f)^2$ is the fraction of the energy of the signal f between $-T$ and $+T$. One is interested in how large $\alpha_T(f)$ can be for signals f with bandwidth Ω, in other words, one wants to determine $\alpha_T \equiv \sup\{\alpha_T(f) \,\big|\, f \in \mathcal{B}_\Omega\}$.

Note that

$$\begin{aligned} \alpha_T^2 &= \sup\{\|D_T(f)\|_2^2 \,\big|\, f = B_\Omega(f),\ \|f\|_2 \le 1\} \\ &= \sup\{\|D_T\, B_\Omega(f)\|_2^2 \,\big|\, \|f\|_2 \le 1\} \\ &= \sup\{(B_\Omega\, D_T\, B_\Omega f, f) \,\big|\, \|f\|_2 \le 1\}. \end{aligned}$$

A result from Hilbert space theory (the Courant–Fischer Theorem) tells us that α_T^2 is the largest eigenvalue of the linear map $B_\Omega\, D_T\, B_\Omega$ (see 7.14). The eigenvalues and eigenfunctions from the preceding subsection appear to play a role here as well. Note that $1 \approx \alpha_T \le 1$ (see 7.20).

From the results in this subsection, we also know that α_T does not depend on T and Ω separately, but only on the product $T\Omega$ (see 7.19). In addition, it appears that ψ_1, corresponding to $\lambda_1 = \alpha_T^2$, is the signal in $\mathcal{B}_\Omega$ that is mostly concentrated in $[-T, +T]$. For a graph of ψ_1 for $T = 1$ and $\Omega = 40$, see Fig. 19.

---

[26] The quantum mechanical particle is described by a *Schrödinger wave packet* $\psi(x)$. This function describes a free quantum mechanical particle in one dimensional space with coordinate x at a fixed time. The square absolute value $|\psi(x)|^2$ represents the probability density for finding the particle at point x. In particular, we have that $\int |\psi(x)|^2 \,\mathrm{d}x = 1$ and the expected position of the particle is $\int x\,|\psi(x)|^2 \,\mathrm{d}x$. We can write $\psi(x)$ as a superposition of pure *De Broglie waves* $\exp(2\pi i p x/h)$, which corresponds to a Schrödinger function of particle momentum p. Here, h is *Planck's constant*. Notice that the 'frequency' of the De Broglie wave is p/h. We write the superposition of pure De Broglie waves as

$$\psi(x) = \frac{1}{\sqrt{h}}\int \phi(p) e^{2\pi i \frac{x}{h} p} \,\mathrm{d}p.$$

As an application of Plancherel's formula, we find that $\int |\phi(p)|^2 \,\mathrm{d}p = 1$. Therefore, $|\phi(p)|^2$ can be interpreted as the probability density that our particle has momentum p. Let us assume that $\int p|\phi(p)|^2 \,\mathrm{d}p = 0$: our particle is expected to be at rest. Theorem 7.23 can be applied with $\psi(x)$ instead of $f(t)$. With $\sigma_x^2 \equiv \int x^2\,|\psi(x)|^2 \,\mathrm{d}x$ and $\sigma_p^2 \equiv \int p^2\,|\phi(p)|^2 \,\mathrm{d}p$, this leads to $\sigma_x\sigma_p \ge h/(4\pi)$, which is viewed as the mathematical formulation of Heisenberg's uncertainty principle.

Exercises

Exercise 7.1. Prove that the functions $t \rightsquigarrow \mathrm{sinc}(2\,t\,\Omega - k)$ $(k \in \mathbb{Z})$ form an orthonormal system.

Exercise 7.2. Consider the functions $f(t) \equiv \sin(2\pi t\Omega)$ and, for some α, $0 < \alpha \ll 1$, $g(t) \equiv \exp(-\pi\alpha^2 t^2)$ $(t \in \mathbb{R})$.

(a) Show that $f \notin L^2(\mathbb{R})$ and $f * g \in L^2(\mathbb{R})$.

(b) Show that the bandwidth of $f * g$ is $> \Omega$.

Exercise 7.3. Suppose that the frequency band of a real-valued signal f is contained in $\mathcal{O} \equiv \{\omega \,\big|\, |\omega| \in [\ell\,\Omega, (\ell+1)\Omega)\}$, where $\Omega > 0$ and ℓ is a positive integer: $\mathcal{O}$ is symmetric about 0 and consists of two intervals of width Ω.

In this exercise, we will see that, although the bandwidth of f is $(\ell+1)\Omega$, it still suffices to sample f at $t_k \equiv k\Delta t$, where $\Delta t \equiv 1/(2\Omega)$, i.e., at a sample frequency of 2Ω as in the Shannon–Whittaker theorem. But with an adapted 'reconstruction function'. We will see that

$$f(t) = \sum_{k=-\infty}^{\infty} f(t_k)\,\cos(\pi(t-t_k)\Omega(2\ell+1))\,\mathrm{sinc}((t-t_k)\Omega) \qquad (t \in \mathbb{R}). \tag{127}$$

There is a 2Ω-periodic function F such that $F(\omega) = \widehat{f}(\omega)$ for all $\omega \in \mathcal{O}$ (see Exercise 3.15). Show that F is even. From Exercise 3.15 we know that

$$F(\omega) = \sum_{k=-\infty}^{\infty} \gamma_k\, e^{-2\pi i\omega k\Delta t} = \sum_{k=-\infty}^{\infty} \Delta t\, f(k\Delta t)\, e^{-2\pi i\omega k\Delta t} \qquad (\omega \in \mathbb{R}).$$

(a) Let $\Phi \in L^2(\mathbb{R})$ be such that $\widehat{\Phi}(\omega) = 1$ if $\omega \in \mathcal{O}$ and $\widehat{\Phi}(\omega) = 0$ elsewhere. Show that

$$\Phi(t) = 2\,\cos(2\pi t\Omega(\ell+\tfrac{1}{2}))\,\frac{\sin(\pi t\Omega)}{\pi t} \qquad (t \in \mathbb{R}).$$

Is this consistent with the formula for Ψ that we have in the proof of the Shannon–Whittaker theorem?

(b) Consider the function $g(t) \equiv \sum_{k=-\infty}^{\infty} \frac{1}{2\Omega} f(t_k)\,\Phi(t-t_k)$ $(t \in \mathbb{R})$. Show that

$$\widehat{g}(\omega) = \sum_{k=-\infty}^{\infty} \Delta t f(k\Delta t)\, e^{2\pi i\omega k\Delta t}\,\widehat{\Phi}(\omega) = F(\omega)\,\widehat{\Phi}(\omega) \qquad (\omega \in \mathbb{R}).$$

Conclude that $\widehat{g}(\omega) = \widehat{f}(\omega)$ for all $\omega \in \mathbb{R}$ and, finally, prove (127).

Exercise 7.4. Let $f \in L^2(\mathbb{R})$.

(a) Show that that

$$\int_{-\infty}^{\infty} (t-s)^2\,|f(t)|^2\,\mathrm{d}t$$

is minimal for $s = t_0$, where t_0 is the center of the energy distribution of f (see (126)).

(b) The uncertainty principle of Heisenberg is 'symmetric' with respect to f and $\widehat{f}$. The second statement in Theorem 7.23 involves a t_0, but does not seem to involve an ω_0. Explain this (seemingly) lack of symmetry.

(c) Derive the scaled version in Footnote 26 of the uncertainty principle.

8 Filtering

In fields as signal processing (with its three major subfields: audio signal processing, digital image processing and speech processing) *filtering* is an important operation. With analog or digital techniques, one wants to filter a part with limited bandwidth from a signal f: for instance, for $\Omega > 0$ one wants to produce a signal f_Ω from f such that $\widehat{f_\Omega} \equiv \widehat{f}\,\Pi_\Omega$ (to avoid aliasing, to let speakers produce the sound of only one radio channel,...). Or noise has to be filtered from a signal. In practice (specifically when using analogue techniques) filtering has to be done in time domain. It is not clear whether the 'ideal' filtered signal f_Ω can be produced in time domain. This issue will be addressed in §8.A. But, first we introduce some terminology.

Of course, in practice, one is not only interested in filtering for low frequencies.

8.1 Definition. Let H be a locally integrable bounded (even) function on $\mathbb{R}$. The map from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$ that maps the signal f to the signal g where $\widehat{g} = \widehat{f}H$ is called the *ideal H-filter*. The signal f is called the *input* signal, g is the *output* or *response* signal, and H is the so-called *transfer function* or *(frequency) response function*.

Consider an oscillation component $t \rightsquigarrow \widehat{f}(\omega)\exp(2\pi i\omega t)$ $(\omega \in \mathbb{R})$ of the input function. The filter changes the amplitude by a factor $|H(\omega)|$: $|H(\omega)|$ is the *gain* at frequency ω. H is complex-valued. The phase of this component is shifted by $\phi(\omega)$, where, ϕ is such that $H = |H(\cdot)|e^{-i\phi(\cdot)}$. Or, equivalently, the time is delayed by $\frac{\phi(\omega)}{2\pi\omega}$ seconds,

$$e^{2\pi i\omega t - i\phi(\omega)} = e^{2\pi i\omega(t-\frac{\phi(\omega)}{2\pi\omega})}:$$

$\frac{\phi(\omega)}{2\pi\omega}$ is the *phase delay* at frequency ω.

Here, we mainly focus on the ideal Π_Ω-filter. Insights for filtering low frequencies are also useful for filtering for other parts of the frequency band (if, for instance, $H = \Pi_\Omega - \Pi_\Gamma$ with $0 < \Gamma < \Omega$).

In order to see whether an ideal H-filter can be formed in the time domain, it is convenient to have a representation of the filtering process in time domain. If H is a response function then, according to Theorem 6.4, the ideal H-filter can be represented in the time domain by means of the convolution product $f \rightsquigarrow f{*}h$ $(f \in L^2(\mathbb{R}))$, where $h \in L^2(\mathbb{R})$ is such that $\widehat{h} = H$. This function h is also called the *impulse response* function (to understand the naming, consider the function $f_\delta \equiv \frac{1}{2\delta}\Pi_\delta$ for $\delta > 0$. Then $f_\delta{*}h(t) = \frac{1}{2\delta}\int_{t-\delta}^{t+\delta} h(s)\,\mathrm{d}s$ and certainly if h is continuous, we will have that $\lim_{\delta\to 0} f_\delta{*}h(t) = h(t)$: if the input signal is a "pulse" f_δ, then the response function will resemble h, the resemblance will be better for shorter pulses).

8.2 Example. The ideal low frequency band filter, the ideal Π_Ω-filter, can be represented in time domain as $f \rightsquigarrow f{*}\widehat{\Pi_\Omega}$. Here we used the fact that $\widehat{\Pi_\Omega}$ is even. Hence, $\widehat{f}(\omega)\widehat{\widehat{\Pi_\Omega}}(\omega) = \widehat{f}(\Omega)\Pi_\Omega(-\omega) = \widehat{f}(\omega)\Pi_\Omega(\omega)$. Recall that (see (b) in 3.7)

$$\widehat{\Pi_\Omega} = \frac{\sin(2\pi t\Omega)}{\pi t} = 2\Omega\,\mathrm{sinc}(2t\Omega).$$

A convolution product can be viewed as weighted averaging. Therefore, filtering in frequency domain corresponds to weighted averaging in time domain.

8.A Filters constructed with the window method

If a filter is to be formed in time domain, then one should realize that it is not possible to include signal values in the 'averaging process' that have not been received yet: in some applications, typically in audio signal processing, the function is available as a 'stream' of function values.[27] Then, the impulse response function h should be such that

$$f{*}h(t) = \int_{-\infty}^{t} f(s)\,h(t-s)\,\mathrm{d}s \quad \text{for each} \quad f.$$

This is possible only if $h(t) = 0$ for all $t < 0$.

8.3 Definition. An impulse response function h is *causal* if $h(t) = 0$ for all $t < 0$.

8.4 If h is an impulse response function for which, for some $s > 0$, $h(t) = 0$ for all $t < -s$, then h need not be causal. However, by delaying the input signal, we can average s seconds 'in future'.

For a signal g and $s \in \mathbb{R}$, we define the delayed signal g_s by $g_s(t) = g(t-s)$.

8.5 Property. *Let f be a signal and h an impulse response function. Then*

$$(f_s){*}h = f{*}(h_s) = (f{*}h)_s.$$

Proof. Express the three convolution products as integrals using the definition and apply appropriate substitution of variables. □

The output signal $f{*}h_s$ produced with a shifted impulse response function h_s is, except for a delay s, precisely the desired signal $f{*}h$. Note that, h_s is causal if $h(t) = 0$ for $t < s$. For ease of terminology, we will call such a response function causal as well.

8.6 The realistic low frequency band filter and Gibbs' phenomenon. The ideal low frequency band filter is not causal and can not be turned into a causal one by a shift. However, by taking T sufficiently large, we can approximate the ideal impulse response function $\widehat{\Pi_\Omega}$ with, for instance, $\widehat{\Pi_\Omega}\Pi_T$. If we shift this approximate impulse response function by T, then we have a causal impulse response function $(\widehat{\Pi_\Omega}\Pi_T)_T$. Except for the delay T, this response function produces the approximate signal $f{*}(\widehat{\Pi_\Omega}\Pi_T)$. We discuss how accurately this constructable signal (with delay) approximates the ideal signal. We compare the approximate response function $(\widehat{\Pi_\Omega}\,\Pi_T)\widehat{\ }$ with the ideal response function Π_Ω.

According to Theorem 6.4, we have that $(\widehat{\Pi_\Omega}\Pi_T)\widehat{\ } = \Pi_\Omega{*}\widehat{\Pi_T}$ and

$$\Pi_\Omega{*}\widehat{\Pi_T}(\omega) = \int_{-\Omega}^{+\Omega} \widehat{\Pi_T}(\omega-\rho)\,\mathrm{d}\rho = \int_{\omega-\Omega}^{\omega+\Omega} \frac{\sin(2\pi T\rho)}{\pi\rho}\,\mathrm{d}\rho.$$

With

$$U_T(\omega) \equiv \int_{-\infty}^{\omega} \frac{\sin(2\pi T\rho)}{\pi\rho}\,\mathrm{d}\rho \quad \text{for} \quad \omega \in \mathbb{R}$$

[27] That is, at time t, only the function values $f(s)$ for $s \leq t$ are available. For $s > t$, one has to wait $s - t$ seconds. Waiting may be undesirable if s is much larger than t.

In other fields, as digital image processing, all function values are already available at the beginning of the 'processing phase'.

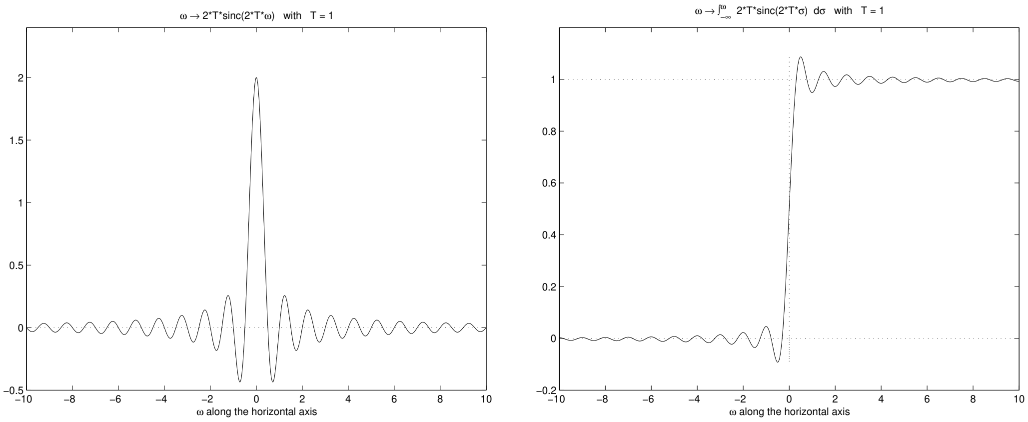


FIGURE 21. The graph of $\sin(2\pi T\omega)/(\pi\omega)$ (left) and its primitive (right).

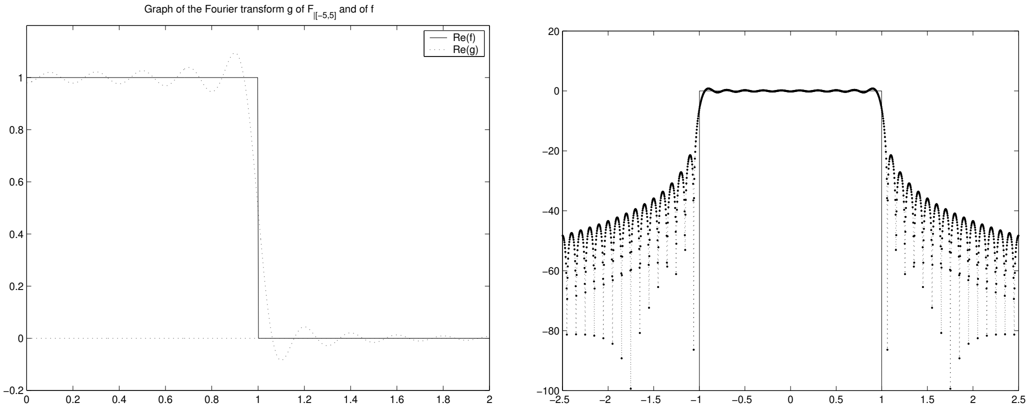


FIGURE 22. The ideal low frequency band filter with 'causal' approximate in frequency domain. The approximate vanishes in time domain outside the time interval $[-T, T]$ and is, strictly speaking, not causal (see §8.4). The left picture shows the ideal filter (solid line) and the causal approximate (dotted line) for $T = 5$. Both functions are even and only the graph for $\omega \geq 0$ is displayed. The right picture shows the ideal filter and the causal approximate on decibel scale.

we have that

$$\Pi_\Omega * \widehat{\Pi_T}(\omega) = U_T(\omega + \Omega) - U_T(\omega - \Omega).$$

For a sketch of the graph of $\widehat{\Pi_T}$ and U_T, see Fig. 21. We now can describe how $\Pi_\Omega * \widehat{\Pi_T}$ approximates the ideal Π_Ω (see Fig. 22). We see that, specifically near the endpoints of the interval $[-\Omega, +\Omega]$, the approximate filter exhibits a lot of deviating wiggles. By taking T larger the area shrinks where the wiggles are large, however, the height of the wiggles near $-\Omega$ and $+\Omega$ does not change. This can be understood as follows

$$U_T(\omega) = \int_{-\infty}^{\omega} \frac{\sin(2\pi T\rho)}{\pi T\rho}\, T\ \mathrm{d}\rho = \int_{-\infty}^{T\omega} \frac{\sin(2\pi\sigma)}{\pi\sigma}\ \mathrm{d}\sigma = U_1(T\omega).$$

Apparently, changing T only implies a rescale of the ω-axis. If T gets larger, the graph of U_T in Fig. 21 is compressed only in the horizontal direction; the height of the wiggles is not affected.[28]

[28] Also without explicit computation, properties of the graph of U_1 can be deduced. Obviously $U_1(-\infty) \equiv \lim_{\omega\to-\infty} U_1(\omega) = 0$. Since $\widehat{\Pi_T}$ is even (why?) it follows that $U_1(\omega) = U_1(\infty) - U_1(-\omega)$. In addition, we have that $U_1(\infty) = \int_{-\infty}^{+\infty} \widehat{\Pi_1}(\rho) e^{2\pi i 0\rho}\ \mathrm{d}\rho = \Pi_1(0) = 1$.

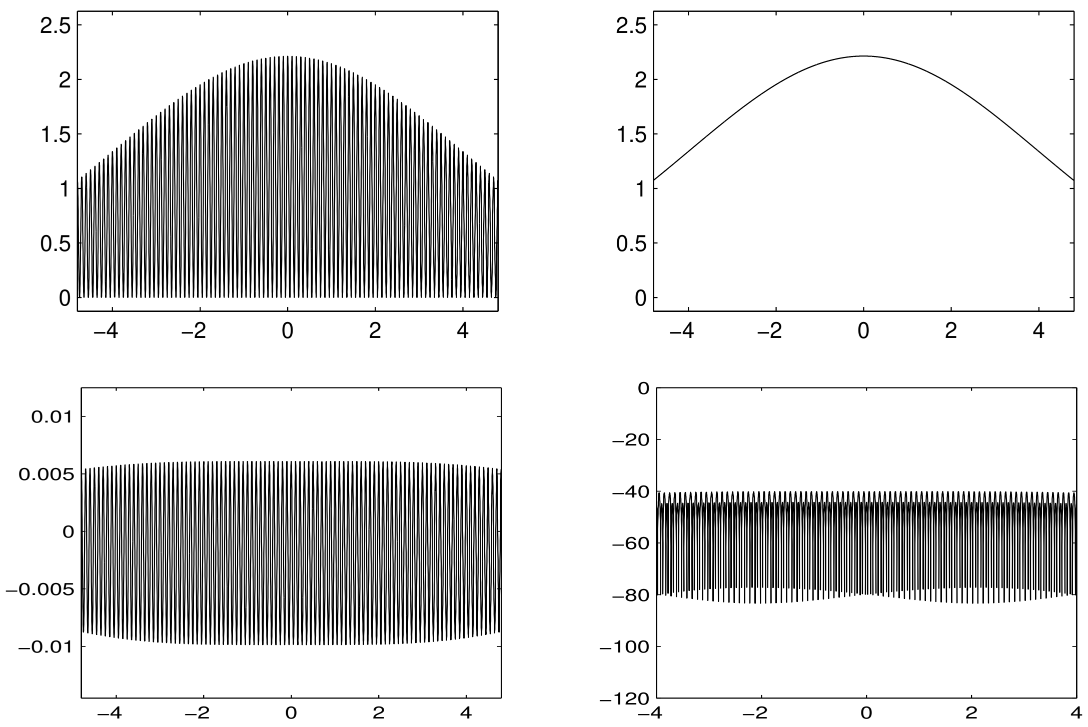


FIGURE 23. The top-left picture shows the original signal f in time domain, the top-right picture shows the filtered signal $f_0 = f * \widehat{\Pi}_\Omega$. Here $f_0(t) \equiv 3.5\sqrt{\alpha}\exp(-\pi\alpha^2 t^2)$, $f(t) \equiv 2f_0(t)\cos^2(2\beta\pi t)$, with $\alpha = 0.1$, $\beta = 5$ and $\Omega = 9$. The bottom-left picture shows the error $f_0 - \widetilde{f}_0$ when f_0 is computed with the causal approximate: $\widetilde{f}_0 = f * (\widehat{\Pi}_\Omega \Pi_T)$. The bottom-right figure also shows this error, but now on decibel scale ($20\log_{10}|f_0 - \widetilde{f}_0|$).

For a signal f that is filtered with this constructable approximation of the ideal Π_Ω-filter, this means that oscillation components with frequencies close to Ω and $-\Omega$ will be amplified by ≈ 10 % ('overshoot'), no matter how large T is. This is the so-called *Gibbs' phenomenon.* In addition, high frequency oscillation components (with $|\omega| > \Omega$) are damped, but for a large set of these components the damping is poor, certainly for acoustical applications. Not only oscillations with frequencies that are slightly larger than Ω are poorly damped but also many oscillations with frequencies that are considerably larger than Ω (see Fig. 23).
(Consider an $L > \Omega$ and, for $\delta > 0$, the signal f_δ with $\widehat{f_\delta}(\omega) = 1$ if $|\,|\omega| - |L|\,| \leq \delta$ and $\widehat{f_\delta}(\omega) = 0$ elsewhere. Let g_δ be the output signal associated with this input signal f_δ: $\widehat{g_\delta} = \widehat{f_\delta}(\Pi_\Omega * \widehat{\Pi_T})$. Then, for small δ, the ratio between the energy of the output and the input signal approximately equals $|U_T(L-\Omega)|^2$, or, in other words, is approximately $10\log|U_T(L-\Omega)|^2$ dB. This can be as much as ≈ -21 dB. From Fig. 22, we see that damping with not more than 20 à 30 dB occurs in a very wide part of the band of high frequencies.)

In the next paragraph we turn to the question whether the above approach can be used to construct filters with better filter properties.

Put $I_k \equiv \int_k^{k+1} \frac{\sin(\pi\sigma)}{\pi\sigma}\, \mathrm{d}\sigma$ for $k \in \mathbb{Z}$. Then
$$\ldots < -I_{-6} < +I_{-5} < -I_{-4} < +I_{-3} < -I_{-2} < +I_{-1} = I_0 > -I_1 > +I_2 > -I_3 > \ldots.$$
This implies that, for $k < 0$, $U_1(2k) = (I_{2k} + I_{2k-1}) + (I_{2k-2} + I_{2k-3}) + \ldots > 0$
and $U_1(2k+1) = (I_{2k+1} + I_{2k}) + (I_{2k-1} + I_{2k-2}) + \ldots < 0$, etc..

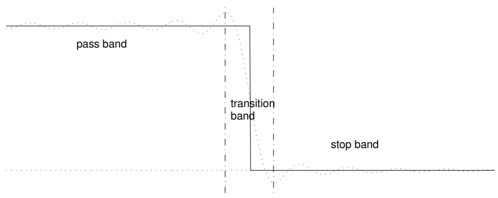


FIGURE 24. Pass band, transition band and stop band of a filter (dotted line) in the frequency domain.

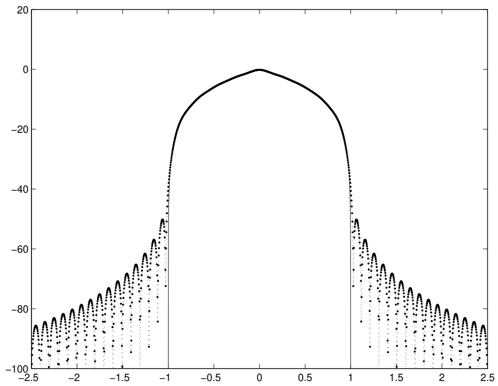


FIGURE 25. An ideal H filter (solid line) and a causal approximations (dotted line) for $T = 5$ in frequency domain. In the picture, H is the Bartlett window. All functions are on decibel scale.

8.7 Other low frequency band filters. In view of the arguments in the preceding subsection, it is clear that, if H is a response function, the ideal H-filter can be approximated by the $H*\widehat{\Pi_T}$-filter. Except for a delay T this filter has a causal impulse response function. This approach to form a causal filter from an ideal H-filter is called a *window-method*: in the weighted average $\int_{-T}^{+T} f(t-s)\,h(s)\,\mathrm{d}s$ (or in its the causal variant) only f-values from the "time window" $[-T,+T]$ are used.

Since $H*\widehat{\Pi_T}$ is continuous (see 6.3) if H is in $L^2(\mathbb{R})$, it is impossible to construct a step function as Π_Ω in this way. A response function $H*\Pi_\Omega$ that vanishes on the band of high frequencies can not be constructed either (see 7.2). Appropriate continuous approximation H of Π_Ω lead to realistic $H*\widehat{\Pi_T}$-filters with significantly better filter properties in the high frequency band and less overshoot than the $\Pi_\Omega*\widehat{\Pi_T}$-filter. The approximations H are also called *windows*: Π_Ω and also H are windows in frequency domain.

If H is continuous, then the ideal H-filter and the realistic $H*\widehat{\Pi_T}$-filter will not only either stop or pass harmonic oscillations but will also somewhat damp oscillations of certain frequencies. The collection of frequencies in this "transition" area is called the *transition band*. The collection of frequencies of the oscillations that are stopped to some acceptable degree are called the *stopband*. Frequencies of oscillations that are hardly damped form the *passband*. "Stop", "pass" and "damp" have to be understood in some relative sense on the decibel scale (cf. Fig. 24).

Continuous response functions H lead to realistic $H*\widehat{\Pi_T}$-filters of which the transition band is wider than that of the $\Pi_\Omega*\widehat{\Pi_T}$-filter but with better stop properties in the stop band (compare the Figures 22, 25, and 26).

Examples.

- *Bartlett* window: $H \equiv (1-\frac{\omega}{\Omega})\,\Pi_\Omega(\omega)$ (see the Fig. 25).
- *Welch* window: $H \equiv (1-(\frac{\omega}{\Omega})^2)\,\Pi_\Omega(\omega)$.

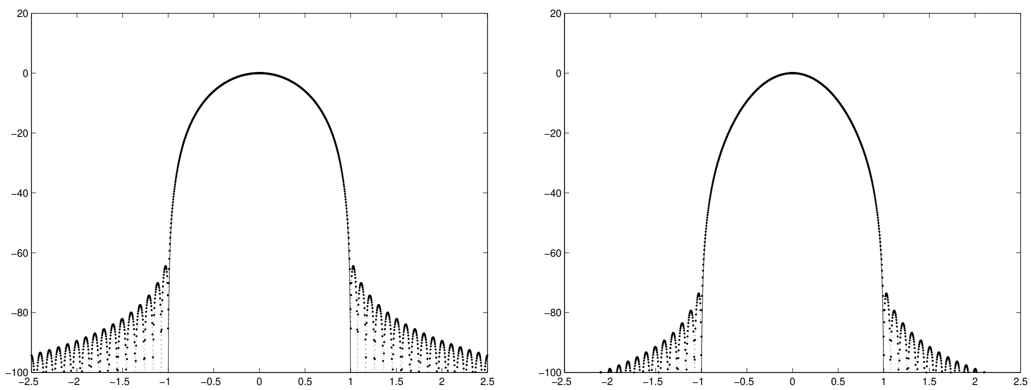


FIGURE 26. Ideal H filter (solid line) and a causal approximations (dotted line) for $T = 5$ in frequency domain. In the left picture, H is the Hann window, and, in the right picture, H is the Blackman window. All functions are on decibel scale.

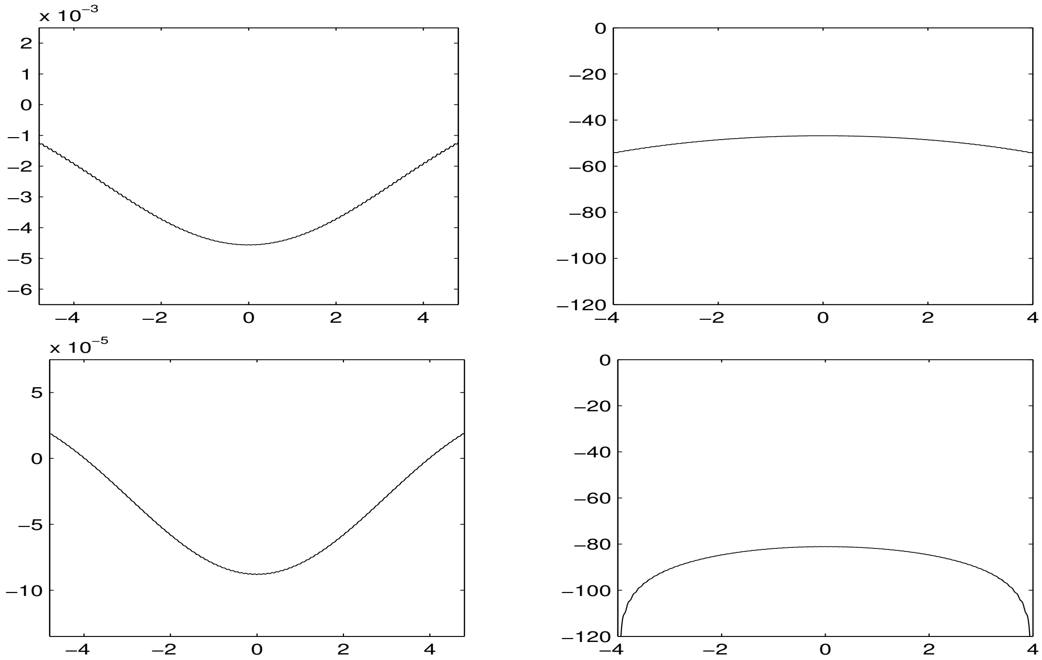


FIGURE 27. The function f of Fig. 23 is filtered here with a causal variant of Bartlett's window (top pictures) and Blackman's window (bottom pictures): $\widetilde{f_0} = f * (\widehat{H}\,\Pi_T)$. The errors $f_0 - \widetilde{f_0}$ are displayed in the left pictures on standard scale and in the right pictures on decibel scale.

• *Hann* window: $H \equiv \frac{1}{2}(\cos(\pi\frac{\omega}{\Omega}) + 1)\,\Pi_\Omega(\omega)$ (see the left picture in Fig. 26).
A *Hamming* window is a small variant of the Hann window:
$H \equiv (0.46\,\cos(\pi\frac{\omega}{\Omega}) + 0.54)\,\Pi_\Omega(\omega)$
• *Blackman* window: $H \equiv \frac{1}{2}(\cos(\pi\frac{\omega}{\Omega}) + 1)\,\Pi_\Omega(\omega) + 0.08(\cos(2\pi\frac{\omega}{\Omega}) - 1)\,\Pi_\Omega(\omega)$ (see the right picture in Fig. 26).
• The *Kaiser* window is based on the 0th order Bessel function.

8.8 Stability. For filters with a causal impulse response function, that we formed by means of a convolution product, we have to discuss the effect of noise on the output signal.
It is easy to see that $|\varepsilon * h(t)| \leq \bar{\varepsilon}\int_{-\infty}^{+\infty}|h(s)|\,\mathrm{d}s = \bar{\varepsilon}\,\|h\|_1$ if $|\varepsilon(s)| \leq \bar{\varepsilon}$ for each $s \in \mathbb{R}$.

For the L^1-impulse response functions h, as constructed with the window method, the noise gets amplified by at most $20\log_{10}\|h\|_1$dB. Although the amplification by the filter of the noise can be large, it can never be unlimited, in case of an L^1-impulse response function. (In view of this kind of stability arguments, we see that, in, for example, the $\Pi_\Omega * \widehat{\Pi_T}$-filter, we should not take the time window $[-T,+T]$ too long.)

8.B Analog filters with an infinitely long impulse response

In time domain, a filter can be represented by a convolution product in which signal values are averaged in some weighted way. The averaging for the ideal Π_Ω-filter involves very old signal values (as well as future ones). Since signals with bounded bandwidth are completely determinated by their values on any time interval (see 7.11 for more details) it is conceivable that there are 'filters' that, by exploiting the smoothness of a signal, realize the ideal filters in time domain better than those from the window methods.

In this section, we consider filters that, although in time domain they only need 'local information', their impulse response function has infinite duration ($h(t)\neq 0$ for arbitrarily large t). These filters assume input signals that have some smoothness.

8.9 The filter. Consider $a_0, a_1, \ldots, a_n$ and $b_0, b_1, \ldots, b_m$ in $\mathbb{R}$.
With, for instance, electronic circuits as in Application 2.6,[29] devices can be constructed that produce from a smooth input signal f, a smooth output signal g for which

$$a_0 g(t) + a_1 g'(t) + \ldots + a_n g^{(n)}(t) = b_0 f(t) + b_1 f'(t) + \ldots + b_m f^{(m)}(t) \quad \text{for} \quad t\in\mathbb{R}. \tag{128}$$

Put $p(\zeta) \equiv a_0 + a_1\zeta + \ldots + a_n\zeta^n$ and $q(\zeta) \equiv b_0 + b_1\zeta + \ldots + b_m\zeta^m$ for $\zeta\in\mathbb{C}$. Then we have that $p(2\pi i\omega)\widehat{g}(\omega) = q(2\pi i\omega)\widehat{f}(\omega)$. Suppose that the coefficients a_i and b_j are such that

$$p(\zeta)\neq 0 \quad \text{for all} \quad \zeta\in\{2\pi i\omega \,\big|\, \omega\in\mathbb{R}\} \quad \text{and} \quad n\geq m.$$

Define

$$H(\omega) \equiv \frac{q(2\pi i\omega)}{p(2\pi i\omega)} \quad \text{for} \quad \omega\in\mathbb{R}.$$

Then is H a bounded $C^{(\infty)}$-function and

$$\widehat{g} = H\widehat{f}.$$

[29] From `en.wikipedia.org`: The simplest electronic filters are based on combinations of resistors, inductors and capacitors. Since resistance has the symbol R, inductance the symbol L and capacitance the symbol C, these filters exist in so-called RC, RL, LC and RLC varieties. All these types are collectively known as *passive* filters, because they are activated by the power in the signal and not by an external power supply.

Here's how passive filters work: inductors block high-frequency signals and conduct low-frequency signals, while capacitors do the reverse. A filter in which the signal passes through an inductor, or in which a capacitor provides a path to earth, therefore transmits low-frequency signals more strongly than high-frequency signals and is a low-pass filter. If the signal passes through a capacitor, or has a path to ground through an inductor, then the filter transmits high-frequency signals more strongly than low-frequency signals and is a high-pass filter. Resistors on their own have no frequency-selective properties, but are added to inductors and capacitors to determine the time-constants of the circuit, and therefore the frequencies to which it responds.

Active filters are implemented using a combination of passive and active components. Operational amplifiers are frequently used in active filter designs.

If $n > m$, then $H \in L^2(\mathbb{R})$ and $g = f{*}h$ with $h \in L^2(\mathbb{R})$ such that $\widehat{h} = H$. We will see that h is also in $L^1(\mathbb{R})$ (see Theorem 8.13 and the subsequent discussion). In particular, we have that $g \in L^2(\mathbb{R})$ (see (110)).

Note that, in practice, initial value conditions for g are needed: if $f(t) = 0$ for all $t \leq 0$, then it seems reasonable to require that $g(0) = g'(0) = \ldots = g^{(n-1)}(0) = 0$. The function $f * h$ meets these conditions only if h is causal. This brings the question how we can tell from the polynomials p and q that is the case?

One can try to construct a low frequence band filter by selecting appropriate p and q and approximating Π_Ω by H. This kind of filters leads to problems that we did not encounter before. As an illustration, we consider an example with p and q as simple as possible. The resulting H happens to be complex, non-real, valued. Filtering with complex valued response functions is possible (and, even unavoidable, as we will see below): if, for instance, $H = \Pi_\Omega \exp(-i\phi)$ with $\phi(\omega) \in [0, 2\pi)$ for $\omega \in \mathbb{R}$, then, for any $\omega \in (-\Omega, +\Omega)$, the filter transforms the harmonic component $t \rightsquigarrow \widehat{f}(\omega) \exp(2\pi i t\omega)$ to the oscillation $t \rightsquigarrow \widehat{f}(\omega) \exp(2\pi i t\omega - i\phi(\omega))$: the phase of the oscillation shifts (by $\phi(\omega)$), while the amplitude is unaffected.

8.10 Example. We consider two filters that, as in (128), are given by

(a) $g + g' = f$ (b) $g - g' = f$.

The functions H for case (a) is not equal to the function H for case (b). However, in both cases, we have that $|H(\omega)|^2 = 1/(1+4\pi^2\omega^2)$: both Hs have the same amplification and damping properties. Nevertheless, the filtering properties are quite different as we will show now.

In practise, f may get perturbed. We will analyze the effect of such perturbations. For $\varepsilon_0 \in \mathbb{R}$ and $\delta > 0$, consider perturbations ε in $L^2(\mathbb{R})$ given by

$$\varepsilon(t) \equiv \varepsilon_0 \quad \text{for } t \in [-\delta, +\delta] \quad \text{and} \quad \varepsilon(t) \equiv 0 \quad \text{elsewhere.}$$

If g satisfies (a), $g + g' = f$, and $\widetilde{g}$ satisfies the perturbed equation, $\widetilde{g} + \widetilde{g}' = f + \varepsilon$, then the *error* $E \equiv \widetilde{g} - g$ satisfies the equation

$$E(t) + E'(t) = \varepsilon(t) \qquad \text{for } t \in \mathbb{R}. \tag{129}$$

The equation should hold in some weak sense, i.e., there is some function D and a constant c such that $E(t) = c + \int_0^t D(s)\,\mathrm{d}s$ and $\int_0^t |D(s)|\,\mathrm{d}s < \infty$ for all $t \in \mathbb{R}$ and $E(t) + D(t) = \varepsilon(t)$ for almost all t: D is the (weak) derivative of E. We put $E' \equiv D$. Note that E is continuous. Note also that, for any $c_1, c_2, c_3 \in \mathbb{R}$,

$$E(t) = c_1 e^{-t}\ (t < -\delta), \quad E(t) = \varepsilon_0 + c_2 e^{-t}\ (|t| < \delta), \quad E(t) = c_3 e^{-t}\ (t > \delta),$$

satisfies (129) if $|t| \neq \delta$. If we select the constants c_1, c_2 and c_3 such that E is continuous also at $t = \pm\delta$, then E solves (129) in the weak sense. In practice, perturbations will effect only future function values g and not past ones (why?). Therefore, c_1 will be 0. Hence, $c_2 = -\varepsilon_0 e^{-\delta}$ (then $E(-\delta) = 0$) and $c_3 = \varepsilon_0(e^\delta - e^{-\delta})$ (to have continuity also at $t = \delta$). Note that $c_3 e^{-t}$ is the effect of the perturbation ε for $t > \delta$. This effect vanishes if t increases. Note also that, with this choice for the c_i, the perturbation E is in $L^2(\mathbb{R})$.

Now, assume that g satisfies (b), $g - g' = f$, and $\widetilde{g}$ satisfies $\widetilde{g} + \widetilde{g}' = f + \varepsilon$, with ε as before. Following the above arguments and insisting on an error $E \equiv \widetilde{g} - g$ that

is 0 in the past, i.e., for $t < -\delta$, we find an E that is continuously differentiable and that, for some $c_3 \neq 0$, is equal to $c_3 e^t$ for $t > \delta$. Note that $|E(t)| \to \infty$ if $t \to \infty$. In particular, $E \notin L^2(\mathbb{R})$. This may seem a bit strange, since, for each input signal f (or ε), Equation (128) has precisely one solution that is also a signal (assuming that $n \geq m$ and p has no zeros on the imaginary axis). However, (128) itself does not require the solution to be a signal and, it appears that the condition '$E(t) = 0$ for $t < -\delta$' leads to this 'exploding' solution. We conclude that (b) is *unstable.*
The $L^2(\mathbb{R})$ solution E can be constructed by putting $c_3 = 0$, and selecting c_2 and c_1 such that E is continuous also at $t = \delta$ and $t = -\delta$, respectively. Then, for some $c_1 \neq 0$, we have that $E(t) = c_1 e^t$ for $t < -\delta$. In particular, $E \in L^2(\mathbb{R})$ and $|E(t)| \to 0$ if $t \to -\infty$. Unfortunately, this L^2-solution incorporates the effect of the perturbation ε on f *before* the perturbation became effective, which is impossible in practice.

In contrast to case (a), it appears that in case (b), the condition '$E \in L^2(\mathbb{R})$' (or, more appropriate, $\widetilde{g} \in L^2(\mathbb{R})$) and the condition '$E(t) = 0$ for $t < -\delta$' can not be satisfied simultaneously.

Instabilities as in (b) show up for large class of filters. The instability problem is a consequence of the desire to employ a filter that operates based on local information, but with a response function that has an infinitely long impulse response.

8.11 Stability. The instability in (b) in §8.10 comes from the fact that the homogeneous equation $g - g' = 0$ has a solution $e^{\lambda t}$ with $\mathrm{Re}(\lambda) \geq 0$. In this case $\lambda = 1$.

More generally, the homogeneous part of the differential equation (128) determines the stability. If the homogeneous part has a solution of the form $e^{\lambda t}$ and $\mathrm{Re}(\lambda) \geq 0$ then the differential equation is unstable and the associated filter is said to be *unstable.* Stability and instability can be expressed in terms of the roots of p (why?):

8.12 Definition. A filter with response function H of the form q/p is *stable* if $\mathrm{degr}(p) > \mathrm{degr}(q)$ and all zeros of p are in the left complex halfplane $\{\zeta \in \mathbb{C} \,\big|\, \mathrm{Re}(\zeta) < 0\}$.[30] $\mathrm{degr}(p)$ is the degree of the polynomial p.

The zeros of p are called the *poles of the filter.* The zeros of q are the *zeros of the filter.*

Consider the situation of §8.10. Note that $E = \varepsilon * h$. Therefore, taking $\varepsilon_0 = \frac{1}{2\delta}$ and driving δ to 0 leads to h: let E_δ be the L^2-solution corresponding to this $\varepsilon_\delta \equiv \varepsilon$, then $h = \lim_{\delta \to 0}(\varepsilon_\delta * h) = \lim_{\delta \to 0} E_\delta$ (Why?). Hence, $h(t) = e^{-t}$ for $t > 0$, $h(t) = 0$ for $t \leq 0$ in case (a), while $h(t) = -e^t$ for $t < 0$, $h(t) = 0$ for $t \geq 0$ in case (b). Note that in both cases h is also in $L^1(\mathbb{R})$. Only in case (a), h is causal. The conclusion holds more general:

8.13 Theorem. *The impulse response function h of a stable filter is causal.*

Proof. Any rational function q/p with $n > m$ and p has n mutually different zeros $\lambda_1, \ldots, \lambda_n$ can be written as

$$H(\zeta) = \frac{q(\zeta)}{p(\zeta)} = \sum_{j=1}^{n} \frac{\alpha_j}{\zeta - \lambda_j} \qquad (\zeta \in \mathbb{C}). \tag{130}$$

[30] If the notion of 'stability', has been formally defined in terms of the effects of perturbations, then the statement in the definition would have been a theorem.

Here, the α_j are appropriate constants in $\mathbb{C}$. Hence, h is the sum of h_j with h_j the inverse Fourier transforms of the term $\alpha_j/(2\pi i\omega - \lambda_j)$; h_j is equal to $\alpha_j e^{\lambda_j t}$ for $t > 0$ and 0 for $t \leq 0$ (see Exercise 3.3). The h_j are causal and, therefore, h is causal. Here, we used the fact that $\mathrm{Re}(\lambda_j) < 0$. If $\mathrm{Re}(\lambda_j) > 0$, then the h_j (in L^2) is given by $-\alpha_j e^{\lambda_j t}$ for $t < 0$ and 0 for $t \geq 0$.

Note that the term $\alpha_j/(\zeta - \lambda_j)$ corresponds to the differential equation $g' - \lambda_j g = \alpha_j f$: the 'special' cases in §8.10 appear to be quite general.

If p has roots, say λ_j, with multiplicity larger than 1, say ℓ, then the sum in (130) also contains terms $\alpha_{j,k}/(\zeta - \lambda_j)^k$ for $k = 1, \ldots, \ell$. The invers Fourier transform $h_{j,k}$ of such a term is equal to $t^k e^{\lambda_j t}$ for $t > 0$ and 0 for $t \leq 0$ (see (c) of Exercise 3.3). Therefore, also for this case, we can conclude that h is causal. $\square$

Note that the above proof implies that h_j and therefore the h not only belongs to $L^2(\mathbb{R})$ but also to $L^1(\mathbb{R})$.

8.14 Butterworth filters. The response function function of filter that is given by the equation

$$g + \frac{1}{(2\pi i\Omega)^n}\, g^{(n)} = f$$

is equal to $\omega \rightsquigarrow 1/(1 + \left(\frac{\omega}{\Omega}\right)^n)$. This function reasonable well resembles Π_Ω for large n. Unfortunately, the filter is unstable. Moreover, its coefficients are not real if n is odd. Stable variants with real coefficients are possible and ar called *Butterworth filters* (of order n).

We actually saw such a variant for $n = 1$: the filter in Example (a) in §8.10 has a stable response function H and $|H(\omega)| = 1/\sqrt{1 + (\frac{\omega}{\Omega})^2}$.

As an additional example, we consider the case $n = 2$.
The filter

$$g + \frac{\sqrt{2}}{2\pi\Omega}\, g' + \frac{1}{(2\pi\Omega)^2}\, g'' = f \tag{131}$$

has a stable response function H and $|H(\omega)| = 1/\sqrt{1 + (\frac{\omega}{\Omega})^4}$.

Since we are interested in stable filters, we learned from the preceding example that we have to deal with response functions H that are not purely real valued. The resulting phase shift leads to a surprising effect that we will discuss below.

8.15 Phase and group delay. If $H = |H|e^{-i\phi}$ is a response function, then the gain $|H(\omega)|$ and the phase delay $\phi(\omega)/(2\pi\omega)$ play a role (see §8.1), but also $1/(2\pi)$ times the derivative $\phi'(\omega)$ of the phase shift. Note that, $\phi(\omega)/\omega \approx \phi'(0)$ if $\omega \approx 0$.

First consider a signal f with bandwidth $\leq \Omega$ and a response function H for which $H(\omega) = e^{-ic\omega}$ if $|\omega| \leq \Omega$. Note that in this case $c = \phi(\omega)/\omega = \phi'(\omega)$: $\phi(\omega) = c\omega$. For the output signal g we have that

$$g(t) = \widehat{\widehat{f}H}(-t) = \int_{-\infty}^{+\infty} \widehat{f}(\omega) e^{2\pi i\omega t - ic\omega}\, \mathrm{d}\omega = f(t - \tfrac{c}{2\pi}) \quad \text{for each } t: \tag{132}$$

the filter delays the input signal by $\frac{c}{2\pi}$ seconds. This is not surprising, since we already know that each harmonic component of f is delayed by $\phi(\omega)/(2\pi\omega)$ seconds which, in this case, is the same amount of time: $\phi(\omega)/\omega$ is constant!

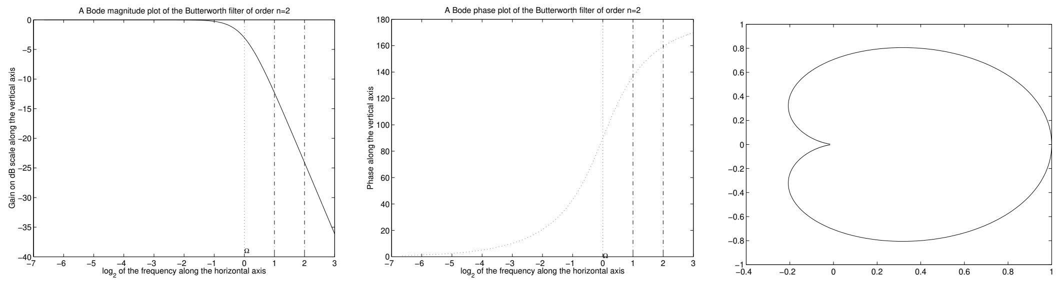


FIGURE 28. The pictures show several graphical ways of displaying a filter. Here, we consider the butterworth filter of order 2. The left picture is a so-called *Bode magnitude (or gain) plot*, that is, $|H(\omega)|$ is plotted along the vertical axis on dB scale ($20\log_{10}|H(\omega)|$) versus the frequency along the horizontal axis. The frequency is also on log scale. Here, we used the $\log_2$ scale, since on this scale one octave corresponds to an interval of unit 1. The value of $|H(\Omega)|$ at the so-called *cut off frequency* Ω is -3 dB (the dotted line); Ω is equal to 1 here. The transition band (right from the cut off frequency) slopes off towards $-\infty$ with slopes of -12dB per octave. An instance of an octave is indicated by the dashed-dotted lines. The frequency response is as flat as possible in the pass band (no ripples). The picture in the middle is a *Bode phase plot*. It shows the phase ϕ as function of the frequency: $H(\omega)=|H(\omega)|e^{-i\phi(\omega)}$. The frequency is again on $\log_2$-scale; ϕ is in degrees. Finally, the right picture is a *Nyquist plot*. It shows the curve $\omega \rightsquigarrow H(\omega)$ in the complex plane, where ω runs from $-\infty$ to ∞.

What if $\phi(\omega)/\omega$ is not constant? Then the situation is much more complicated, but for certain interesting signals, the derivative of the phase shift plays a role in the delay of the output signal.
To see this, consider a general smooth H and a signal of the form $f(t)=f_0(t)\,\exp(2\pi i t\Omega)$, where $\widehat{f_0}$ is concentrated on a small domain $[-\varepsilon,\varepsilon]$ around 0, that is, $\widehat{f_0}(\omega)=0$ (or ≈ 0) for $|\omega|>\varepsilon$. Note that $\widehat{f}$ is concentrated in a small domain around Ω: $\widehat{f}(\omega)=\widehat{f_0}(\omega-\Omega)$. Such a signal f is called a *wave packet* (*with frequency* Ω), f_0 is the *envelope*. Taylor approximations lead to $H(\omega)\approx|H(\Omega)|e^{-i(\phi(\Omega)+c(\omega-\Omega))}$ for $\omega\in[\Omega-\varepsilon,\Omega+\varepsilon]$. Here, $c\equiv\phi'(\Omega)$. Hence,

$$\begin{aligned} g(t) &= \widehat{\widehat{f}H}(-t)\approx\int_{-\infty}^{+\infty}\widehat{f}(\omega)|H(\Omega)|e^{-i(\phi(\Omega)-c\Omega)}e^{2\pi i\omega t-ic\omega}\,d\omega \\ &= |H(\Omega)|\,e^{-i(\phi(\Omega)-c\Omega)}\,f(t-\tfrac{c}{2\pi})=|H(\Omega)|\,f_0(t-\tfrac{\phi'(\Omega)}{2\pi})\,e^{2\pi i(t-\frac{\phi(\Omega)}{2\pi\Omega})\Omega}. \end{aligned} \tag{133}$$

Again, we see that the derivative of the phase shift gets expressed as a delay, now as a delay of the '*envelope*' f_0 by $\frac{1}{2\pi}\phi'(\Omega)$ seconds; $\frac{1}{2\pi}\phi'$ is the so-called *group delay*. The phase delay $\phi(\Omega)/(2\pi\Omega)$ is visible in the 'wave part' $\exp(2\pi i t\Omega)$ of the wave packet.

Note that, (133) is consistent with (132): if $\phi(\omega)=c\omega$, then $\phi(\Omega)/\Omega=c=\phi'(\Omega)$.

For real wave packets and even filters, we have a similar result.
For f of the form $f(t)=f_0(t)\,\cos(2\pi t\Omega)$ with $\widehat{f_0}$ concentrated on $[-\varepsilon,+\varepsilon]$ and H even (i.e., $H(-\omega)=\overline{H(\omega)}$, whence $|H|$ even, ϕ odd, and ϕ' even), we have that

$$g(t)=\widehat{\widehat{f}H}(t)\approx|H(\Omega)|\,f_0(t-\tfrac{\phi'(\Omega)}{2\pi})\,\cos(2\pi\left(t-\tfrac{\phi(\Omega)}{2\pi\Omega}\right)\Omega). \tag{134}$$

8.16 Chebyshev filters. *Chebyshev filters* are related to Butterworth filters. They are stable, but they have a steeper roll-off (the function is steeper in the transition band), 'at the cost' of more ripples in the passband or in the stopband. They have the property that they minimize the error between the idealized filter at the passband (Chebyshev filters of *type I*) or at the stopband (Chebyshev filters of *type II*) and the actual one, but as noted, there can be ripples in the passband (type I) or in the stopband (type II). For this reason filters which have a smoother response in the passband but a more irregular response in the stopbands are preferred for some applications.

The frequency (amplitude) characteristic of a Chebyshev filter of the nth order can be described mathematically by:

$$|H(\omega)|^2 = \frac{1}{1+\varepsilon^2\, T_n^2\left(\frac{\omega}{\Omega}\right)},$$

where $|\varepsilon| < 1$ and $|H(\Omega)| = 1/\sqrt{1+\varepsilon^2}$ is the gain at the cutoff frequency Ω, and T_n is a Chebyshev polynomial of the nth order.[31] (Note: Commonly the frequency at which the gain is -3dB is called the *cutoff frequency*. But that does not hold for Chebyshev filters!)

8.C Digital filters

In electronics, a *digital filter* is a filter in discrete time, that is implemented through digital computation. It operates on sampled signals (see §7.4 and (b) of Exercise 6.13).

One advantage of digital filters is the stability of their parameters. A computer is much less susceptible to environmental conditions than analog circuits. Computers are relatively cheap, and the advantages of digital signal processing are outweighing complex analog filters. (But, of course, computer chips consist of many very complex analog filters!) Although, for proper sampling an (analog) anti-aliasing filter is needed, which is usually a low-pass filter or band-pass filter.

In this section, we define

$$\widehat{f}(\omega) \equiv \sum_{k=-\infty}^{\infty} f_k\, e^{-2\pi i\omega k} \quad (\omega \in \mathbb{R})$$

for digital signals $f = (f_n) \in \ell^2(\mathbb{Z})$; $\widehat{f}$ is 1-periodic and $f_k = 2\int_0^1 \widehat{f}(\omega)\, e^{2\pi i\omega k}\, \mathrm{d}\omega$ (see §7.4, where here, for ease of notation, we took $\Omega = \frac{1}{2}$, i.e., $\Delta t = \frac{1}{2\Omega} = 1$).

We leave the proof of a number of claims in this section to the reader: these claims and proofs are analogues of the ones in §8.B.

8.17 The filter. A digital filter with input $f \in \ell^2(\mathbb{Z})$ and output g is typically of the form

$$\begin{aligned} \alpha_0 g_j &= (\beta_0 f_j + \beta_1 f_{j-1} + \ldots + \beta_m f_{j-m}) - (\alpha_1 g_{j-1} + \alpha_2 g_{j-2} + \ldots + \alpha_n g_{j-n}) \\ &= \sum_{k=0}^{m} \beta_k\, f_{j-k} - \sum_{k=1}^{n} \alpha_k\, g_{j-k}. \end{aligned} \tag{135}$$

Hence, g satisfies the equation

$$\alpha * g = \beta * f \qquad \text{on } \mathbb{Z}$$

where $\alpha \equiv (\alpha_0, \ldots, \alpha_n)$, $\alpha_0 \neq 0$ and $\beta \equiv (\beta_0, \ldots, \beta_m)$ are finite sequences of real numbers. We will assume that $\alpha_0 = 1$, which is no restriction. The function g can be computed in MATLAB with the command '`g=filter(beta,alpha,f)`'. The filter has

[31] The Chebyshev polynomials have been introduced in Exercise 2.8.

m so-called *feed-forward stages* and n *feedback stages.* The number n of feedback stages is called the *order* of the filter. If the order is 0 (i.e, $n = 0$: $\alpha_j = 0$ for all $j > 0$), then we have a so-called *finite impulse response* (FIR) filter. The number m of coefficients of a FIR filter is the *length of the filter* (also referred to as the number of *tabs* and sometimes as the order). We have an *infinite impulse response* (IIR) filter if $n > 0$: the response has infinite duration because feedback of the outputs is used to calculate other output values. Unlike the FIR filter, the IIR filter must have an 'initial condition' for n successive feedback values: for instance, if $f_k = 0$ for $k < 0$ then $g_{-1} = \ldots = g_{-n} = 0$ are appropriate 'initial values' for g.

With $p(\zeta) \equiv 1 + \alpha_1\zeta + \ldots + \alpha_n\zeta^n$ and $q(\zeta) \equiv \beta_0 + \beta\zeta + \ldots + \beta_m\zeta^m$ for $\zeta \in \mathbb{C}$, we have that $p(\exp(-2\pi i\omega))\,\widehat{g}(\omega) = q(\exp(-2\pi i\omega))\,\widehat{f}(\omega)$. Hence, the response function H is given by

$$H(\omega) = \frac{\sum_{k=0}^{n} \beta_k\, e^{-2\pi i\omega k}}{1 - \sum_{k=1}^{m} \alpha_k e^{-2\pi i\omega k}} = \frac{q(\zeta^{-1})}{p(\zeta^{-1})}, \quad \text{where} \quad \zeta = e^{2\pi i\omega}.$$

The function $z \rightsquigarrow q(z^{-1})/p(z^{-1})$ $(z \in \mathbb{C})$ is the so-called *z-transform* of the filter. In MATLAB: '`[H,omega]=freqz(beta,alpha,n)`'. Note that the z-transform is a rational function of z: with $N \equiv \max(n, m)$, both $z^N q(z^{-1})$ and $z^N p(z^{-1})$ are polynomials.

8.18 Stability. The filter is *stable* if $p(\lambda^{-1}) = 0$ implies that $|\lambda| < 1$, or, equivalently, if the zeros of the polynomial $z^N p(z^{-1})$ are in the open complex unit disc.

The definition is motivated by the following observation. If $g = (g_k)$ is a solution and $p(\lambda^{-1}) = 0$, then $\widetilde{g}_k \equiv g_k + \varepsilon\lambda^k$ is a solution as well. Stability guarantees that such a perturbation is relatively harmless.

A filter that is not stable, i.e., $|\lambda| \geq 1$ for some $\lambda \in \mathbb{C}$ for which $p(\lambda^{-1}) = 0$, is *unstable.*

The zeros of $p(z^{-1})$ are the *poles* of the filter, while the zeros of $q(z^{-1})$ are referred to as the *zeros* of the filter. The poles of the filter determine the stability. Both poles as well as zeros of the filter are important in the design of the filter: except for a scalar multiplicativity factor, the filter is completely determined by its poles and zeros. The multiplicativity factor does not play an essential role.

There is a close similarity between the analog filters discussed in §8.B and the digital filters that we have here. Note, however, the difference in the stability condition: $\mathrm{Re}(\lambda) < 0$ in case of analog filters, while $|\lambda| < 1$ for digital filters. This is precisely the difference that we know from discussions on differential equations versus difference equations: (135) is the general form of a linear difference equation with constant coefficients. These type of equations show up when, for instance, a ordinary differential equation (initial value problem) has been discretized with a so-called multistep method.

We did not discuss existence of g in $\ell^2(\mathbb{Z})$ yet. We will address this issue now. Note, however, that we are not only interested in having a square summable output, but the output may also have to satisfy initial conditions.

8.19 Causality. Suppose the filter is stable.
Then, $p(\zeta) \neq 0$ for all $\zeta \in \mathbb{C}$, $|\zeta| = 1$. Hence, H is a continuous 1-periodic function. In particular, $H \in L_1^2(\mathbb{R})$ and there is an $h = (h_k) \in \ell^2(\mathbb{Z})$ for which

$$\widehat{h}(\omega) = \sum_{k=-\infty}^{\infty} h_k\, e^{-2\pi i\omega k} = \sum_{k=-\infty}^{\infty} h_k\, \zeta^{-1} = \frac{q(\zeta^{-1})}{p(\zeta^{-1})} = H(\omega) \quad \text{where } \zeta \equiv e^{2\pi i\omega}.$$

With Fourier transforms, one easily verifies that, with $g \equiv f * h$, g satisfies (135). From the fact that p and q are polynomials, p non zero on the complex unit circle, one can even deduce that $h \in \ell^1(\mathbb{Z})$ (cf., Exercise 8.7 and Exercise 8.8). Hence, $g \in \ell^2(\mathbb{Z})$ whenever $f \in \ell^2(\mathbb{Z})$. However, it is not clear that $g_k = 0$ for all $k < 0$ if $f_k = 0$ for all $k < 0$. In other words, we do not know whether h is causal.

However, it can be shown that (see Exercise 8.7), for some sequence (h'_k) in $\ell^1(\mathbb{Z})$,

$$\frac{q(z)}{p(z)} = \sum_{k=0}^{\infty} h'_k\, z^k \quad \text{for all } z \in \mathbb{C} \text{ for which } |z| \leq \frac{1}{\rho},$$

where $\rho \equiv \max\{|\lambda| \,\big|\, \lambda \in \mathbb{C},\, p(\lambda^{-1}) = 0\}$. Stability of the filter implies that $\rho < 1$. In particular, we have convergence for $z = \zeta^{-1}$, where $\zeta = \exp(2\pi i\omega)$. Since two $\ell^2(\mathbb{Z})$-sequences that have the same Fourier transform, coincide, we have that, $h_k = h'_k$ for $k \geq 0$, and $h_k = 0$ for $k < 0$. In particular, h is causal: stability implies causality. Conversely, if the filter is not stable then h is not causal (cf., Exercise 8.8).

8.20 Discussion. If the filter is a FIR filter, then $p = 1$. There are no poles and the filter is stable. In this case $h = \beta$, h has finite duration: only finitely many h_k are non-zero. Here, we define $h_k \equiv 0$ for k for which β_k is not defined. The impulse response function h has an infinite duration (see Exercise 8.7) if the filter is a (stable) IIR filter.

A FIR filter can have a linear phase. It has linear phase if and only if its coefficients are symmetric about the center coefficient.

If the impulse response function of a FIR filter has a long duration, then it it may be attractive to apply (Fast) Fourier transform to compute $f * h$. There is no advantage in using Fourier transform to compute the output of IIR filters, that is, if Fourier transforms are considered for practical computation with IIR filters, then one could as well have designed a filter in frequency domain and not bother at all about its realization in time domain. There is no need to work with rational functions then. In the construction phase of an IIR filter, however, the Fourier transform does play an important role.

The sequences α and β of IIR filters are usually relatively short and the filter is often applied in situations in which the input signal f is available as a 'stream' of function values (as when playing a CD or on radio). The filter is causal: g_j can be computed as soon as f_k is available for $k \leq j$, but f_k for $k > j$ is not yet needed. With the feedback part α one wishes to achieve good filtering properties with short α and β that otherwise with FIR filters would require a long β.

8.21 Butterworth and Chebyshev filters. There are digital versions of the Butterworth filters as well as of the Chebyshev filters: the gain (that is, $|H(\omega)|$) for these versions is as described in §8.14 and §8.16, respectively, but the stability has to be understood in the 'digital' sense.

8.22 Equiripples filter. Approximation of the ideal Π_{Ω_0} filter leads to the Gibb's phenomenon. To avoid the disadvantages associated herewith, FIR filters are constructed that approximate as accurately as possible the value 1 for $|\omega| \leq \omega_{\text{pass}}$ and the value 0 for $|\omega| \geq \omega_{\text{stop}}$ and, of course, $|\omega| \leq \frac{1}{2}$: they do not 'worry' about the transition band (for frequencies ω with $\omega_{\text{pass}} < |\omega| < \omega_{\text{stop}}$. Here we assume that cutoff frequency

$\omega_{\rm pass}$ of the passband is smaller than the cutoff frequency $\omega_{\rm stop}$ of the stopband and $\omega_{\rm stop} \leq \frac{1}{2}$). If accuracy is measured in the sup-norm, with supremum taken over all $\omega \in [-\frac{1}{2}, +\frac{1}{2}]$ except for the ones in the transition band, then the best filter, the *equiripples* or *elliptic* filter, has ripples of equal (maximum and minimum) height, say δ, in passband and in stopband. The filter slopes almost linearly off (in dB scale) in the transition band. The height δ depends on the length N of the filter: $N \approx \frac{-\ln(\pi\delta)}{\ln(\cot\frac{\pi-\Delta\omega}{4})}$, where $\Delta\omega \equiv \omega_{\rm stop} - \omega_{\rm pass}$ is the width of the transition band. The so-called Rémez algorithm can be used to compute the best filter for any length N. If accuracy is measured in another norm (as the 2-norm over the frequency area of interest), then other filters arise.

8.23 Application: MP3 compression. A bird that flies in front, or almost in front, of the sun is not visible. For the same reason tones are not audible if they sound at approximately the same time as a louder tone with approximately the same pitch. These softer tones can be blocked without affecting the quality of the piece of music. Audio compression techniques as MP3 (MPEG/audio Layer III) exploit this fact.

Psychoacoustics. The resolving power of the human auditory system appears to be frequency dependent. Experiments show that this dependency can be expressed in terms of 27 critical frequency bands $\mathcal{O}_i$ with widths which are less than 50 Hz for the lowest audible frequencies and approximately 5 kHz at the highest. For each critical band there is a 'mask' that can be used to block softer tones in neighbouring frequency bands, or, in a more mathematical formulation, for each critical band $\mathcal{O}_i$ there is a positive function $\omega \rightsquigarrow \delta_i(\omega)$ (the mask) with $\delta_i(0) = 1$ and decaying to zero for increasing $|\omega|$ such that f and F sound the same. Here, f and F are signals such that $\widehat{F}(\omega) = 0$ if $\omega \notin \mathcal{O}_i$ and $|f(\omega)| \leq \max\{|\widehat{f}(\mu)|\,\delta_i(\omega-\mu) \,\big|\, \mu \in \mathcal{O}_i\}$, and $\widehat{F}(\omega) = f(\omega)$ elsewhere.

MPEG/audio. The MPEG/audio algorithms use 32 frequency bands O_i of equal width, (on linear scale) whereas the width of the 27 critical bands $\mathcal{O}_i$ fits equality better on a log-scale. However, equal width on linear scale allows more efficient computation as we will see below. These algorithms split a digital audio signal f into 32 signal f_i, where (ideally) f_i has frequencies only in O_i. Sample values $f_{i'}(t_{j'})$ are replaced by zero if signal components $f_i(t_j)$ in a neighbouring frequency band O_i and nearby time domain ($t_{j'} \approx t_j$ and $t_j \leq t_{j'}$) carry relatively much energy: softer tones are blocked. The blocking threshold depends on amount of energy and band numbers. Layer III (MP3) is more sophisticated in computing the blocking thresholds than the layer I and layer II approaches: it applies MDCT (see Exercise 5.9) to each of the f_i to obtain more detailed spectral information.[32] In addition, each of the signals f_i is '*quantized*', that is, each (16 bit) function value $f_i(t_j)$ is replaced by an m-bit value with $m = 8$ or less.[33] Quantization (that is, the value for m) also depends on the band and is as

[32] The splitting of a signal into 32 signals with a limited frequency band has been preserved in MP3 for compatibility reasons. Other modern compression techniques as Advanced Audio Control (AAC) rely on MDCT only.

[33] Quantization is preceded by a transformation of the function value $x = f_i(t_j)$ according to the so-called *μ-law*: $y = \mathrm{sign}(x)\ln(1+\mu|x|/X)/\ln(1+\mu)$, here X is the maximum absolute value that the function values can take, μ is an appropriate constant (as $\mu = 255$). The resulting logarithmic step spacing obtained after quantization represents low-amplitude audio samples with greater accuracy than higher-amplitude values, which is not the case for linear quantization (i.e., quantization of x/X). So, this strategy avoids that soft parts of a piece music get rounded to 0.

coarse as inaudibility allows.[34]

Polyphase filter bank. The splitting exploits a so-called *polyphase filter bank*:

$$f_i(t_k) = f * h_i(t_k) = \sum_{j=-\infty}^{\infty} f(t_{k-j})\, h_i(j), \ \text{ where } h_i(k) = 2C(k)\cos(\tfrac{\pi}{64}(2i+1)k), \quad (136)$$

where C is the impulse response function of an approximate ideal Π_δ filter with $\delta \equiv \frac{1}{128}$.

The choice for this 'analysis' filter bank for the encoder can be understood as follows. First recall that, with sample frequency $2\Omega = 1/\Delta t$, (Ω is the bandwidth of f),

$$\widehat{f}(\omega) = \Delta t \sum f(k\Delta t)\, e^{2\pi i k\,\Delta t\,\omega} = \Delta t \sum f(k\Delta t)\, e^{2\pi i\, k\,\widetilde{\omega}}, \ \text{ where } \widetilde{\omega} \equiv \Delta t\,\omega = \frac{\omega}{2\Omega}.$$

The range of ω is $(-\Omega, +\Omega)$, whence, $\widetilde{\omega}$ ranges in $(-\frac{1}{2}, +\frac{1}{2})$, and with $\nu_i \equiv \frac{2i+1}{128}$,

$$\widetilde{O}_i \equiv \{\widetilde{\omega} \,\big|\, -\delta \le |\widetilde{\omega}| - \nu_i < \delta\} \qquad (i = 0, \dots, 31)$$

defines a partitioning of the scaled frequency interval $(-\frac{1}{2}, \frac{1}{2})$ into 32 bands $\widetilde{O}_i$ of equal width 2δ (since $\nu_{i+1} - \nu_i = 2\delta$, $\nu_0 - \delta = 0$, $\nu_{31} + \delta = \frac{1}{2}$). Note that the $\widetilde{O}_i$ are symmetric about 0 ($\widetilde{\omega} \in \widetilde{O}_i \Rightarrow -\widetilde{\omega} \in \widetilde{O}_i$) and $O_i = \{2\Omega\widetilde{\omega} \,\big|\, \widetilde{\omega} \in \widetilde{O}_i\}$. If C filters for frequencies in $(-\delta, \delta)$, then $k \rightsquigarrow C(k)[\exp(2\pi i k\nu_i) + \exp(-2\pi i k\nu_i)] = h_i(k)$ filters for frequencies in $\widetilde{O}_i$, since $\widehat{h}_i(\widetilde{\omega}) = \widehat{C}(\widetilde{\omega} + \nu_i) + \widehat{C}(\widetilde{\omega} - \nu_i)$.

For C a sequence of 512 coefficients is used.

Note that $k \rightsquigarrow f_0(t_k)$ has (ideally) maximum frequency 2δ. Hence, its Nyquist ratio is $1/(2 \cdot 2\delta) = 32$. Therefore, it suffices to subsample $f_0(t_k)$ at $k = 32\ell$. But it also suffices to subsample the other f_i by 32 (see Exercise 7.3). This implies that it suffices to replace each set of 32 input samples as $f(t_0), \dots, f(t_{31})$ with a set of only 32 output samples as $f_0(t_0), \dots, f_{31}(t_0)$: the outcome of the filter bank does not require more storage than the original signal!

Of course, the 'prototype' filter $\widehat{C}$, does not have a sharp cutoff at frequency δ: $\widehat{C}$ is not equal to Π_δ. So, when the filter outputs are subsampled by 32, there is a considerable amount of aliasing. However, the design of the prototype filter $\widehat{C}$ results in a complete alias cancellation at the output of the decoder's 'synthesis' filter bank. Another consequence of the lack of sharpness at cutoff is that some frequencies get represented at two bands. The MDCT in the Layer III approach allows to remove some artifacts caused by this overlap.

Exercises

Exercise 8.1. Let h be an pulse response function.

(a) Prove that the following three properties (in continuous time) are equivalent.
  (i) h is causal
  (ii) $f * h(t) = 0$ for all $t < 0$, whenever $f(t) = 0$ for all $t < 0$ ($f \in L^2(\mathbb{R})$).
  (iii) $f * h(t) = \int_{-\infty}^{t} f(s)\, h(t-s)\, \mathrm{d}s$ for all t and all $f \in L^2(\mathbb{R})$.

(b) Formulate and prove the analogue of the above statement in discrete time.

Exercise 8.2. Consider the situation as in §8.10.

[34] All these techniques replace function values by 0 or by values that can be represented by much less bits than the original values. In the layer III approach an additional compression is obtained by the Huffman coding. This is a lossless compression technique.

Assume f is perturbed by ε that is now given by

$$\varepsilon(t) \equiv \varepsilon_0(t^2 - \delta^2) \quad \text{for } |t| < \delta \quad \text{and} \quad \varepsilon(t) \equiv 0 \quad \text{elsewhere.}$$

Note that ε is continuous and in $L^2(\mathbb{R})$.

(a) Prove that the continuous solution E is continuously differentiable.

(b) Determine the error E that is in $L^2(\mathbb{R})$ for case (a) of §8.10. The same question for case (b).

Exercise 8.3. Prove (134).

Exercise 8.4. **Optimum filter.**

Consider a signal f. Suppose that the signal is perturbed by white noise n: $\widetilde{f} \equiv f + n$. We are interested in a filter that gives the best signal-to-noise ratio, i.e., that is such that after filtering the ratio N/S of the energy in the noise (N) and in the signal (S) is small as possible.

The energy in noise and signal after filtering is given by

$$\int_{-\infty}^{\infty} |\widehat{n}(\omega)\, H(\omega)|^2 \, \mathrm{d}\omega \quad \text{and} \quad \int_{-\infty}^{\infty} |\widehat{f}(\omega)\, H(\omega)|^2 \, \mathrm{d}\omega$$

respectively. Here, H is the frequency response function of the filter. For white noise, we may assume that

$$\int_{-\infty}^{\infty} |\widehat{n}(\omega)\, H(\omega)|^2 \, \mathrm{d}\omega = \kappa \int_{-\infty}^{\infty} |H(\omega)|^2 \, \mathrm{d}\omega$$

for some constant κ. Why?

(a) Prove that $N/S \geq \kappa$ and that equality holds only if $|H|$ is a multiple of $|\widehat{f}|$. This is the *matched filter theorem*. (Hint: apply Cauchy–Schwartz to S/N).

It appears that, for the best filter, the power spectrum of the signal needs to be know in advance. This filter could not be implemented in hardware, but can only be (approximately) applied by software after the data has been collected.

Exercise 8.5. **Butterworth filters.**

Consider a butterworth filter of order n.

(a) Compute $H(\omega)$ in case $n = 2$ and show that the filter is stable.

(b) Show that $|H(\omega)|$ at cut off frequency $\omega = \Omega$ is ≈ -3dB (or to be more precise, $-3.0103\ldots$dB), independent of the order n.

(c) Show that the slope in the transition band approximates n times $-6.0206\ldots$dB for large frequencies ω.

Exercise 8.6. **The stability of digital filters..**

Consider the digital filter defined by $p(z) = 1 + \alpha_1 z + \ldots + \alpha_n z^n$ and $q(z) = \beta_0 + \beta_1 z + \ldots + \beta_m z^m$ with input signal f and output signal g.

(a) Prove that $(g_j + \varepsilon\lambda^j)$ is a solution if (g_j) is a solution and $p(\lambda^{-1}) = 0$.

(b) Assume that $p(z) \neq 0$ if $|z| \leq 1$. Prove that $\varepsilon\lambda^j \to 0$ if $j \to \infty$.

Exercise 8.7. **The output of digital filters.**

Let $\alpha_1 = -\lambda_1$ be such that $|\lambda_1| < 1$. Consider the digital filter

$$g_j = \beta_m\, f_{j-m} - \alpha_1 g_{j-1} \quad (j \in \mathbb{Z})$$

with input signal f and output signal g.

(a) Verify that $q(z) = \beta_m\, z^m$ and $p(z) = 1 - \lambda_1\, z$.

(b) Show that the filter is stable (also in the sense that errors at, say g_0, are not amplified).

(c) Show that for $|z| < 1/\lambda_1$, we have that

$$\frac{q(z)}{p(z)} = \sum_{k=0}^{\infty} \beta_m\, z^m\, (\lambda_1\, z)^k.$$

(d) Let $h_{k+m} \equiv \beta_m\, \lambda_1^k$ for $k \geq 0$ and $h_{k+m} = 0$ if $k < 0$. Prove that $g = h$ if $f_0 = 1$, $f_k = 0$ if $k \neq 0$ and $g_k = 0$ if $k < 0$. Note that the sequence h is in $\ell^1(\mathbb{Z})$ and that it has an *infinite duration*: $h_k \neq 0$ for all $k \geq 0$.

Consider the general digital filter defined by $p(z) = 1 + \alpha_1 z + \ldots + \alpha_n z^n$ and $q(z) = \beta_0 + \beta_1 z + \ldots + \beta_m z^m$ with input signal f and output signal g.

(e) Prove that $p(z) = \Pi_{j=0}^{n}(1 - \lambda_j\, z)$, where λ_j are the zeros of $p(z^{-1})$.

(f) Show that there is an $h \in \ell^1(\mathbb{Z})$ such that

$$\frac{q(z)}{p(z)} = \sum_{k=0}^{\infty} h_k\, z^k \quad \text{ for all } z \in \mathbb{C},\ |z| < \frac{1}{\rho},$$

where $\rho \equiv \max\{|\lambda_j|\} < 1$.

(Hint: Since $\frac{1}{p(z)} = \sum \frac{w_j}{1-\lambda_j z}$ for appropriate weights w_j, we have that $\frac{q(z)}{p(z)}$ is a linear combination of terms of the form $\beta_{\widetilde{m}}\, z^{\widetilde{m}}/(1 - \lambda_j\, z)$.)

(g) Prove that $g = h$ if $f_0 = 1$, $f_k = 0$ if $k \neq 0$ and $g_k = 0$ if $k < 0$ and conclude that

$$\widehat{h}(\omega) = \frac{q(\zeta^{-1})}{p(\zeta^{-1})} \quad \text{for } \zeta \equiv e^{2\pi i \omega}.$$

(h) Show that $g = f * h$ and $g \in \ell^2(\mathbb{Z})$ if $f \in \ell^2(\mathbb{Z})$.

(i) Show that the impulse response function h has an infinite duration, i.e., for each $k_0 \in \mathbb{N}$, there is a $k \in \mathbb{N}$, $k > k_0$ such that $h_k \neq 0$.

Exercise 8.8. The output of digital filters 2.

Consider the digital filer

$$g_n = f_n - 2 g_{n-1}.$$

(a) Compute the response function H. Is H continuous and, therefore, in $L_1^2(\mathbb{R})$?

(b) Show that

$$\frac{q(z^{-1})}{p(z^{-1})} = \frac{1}{1 - \frac{z}{2}} = \frac{\frac{1}{2} z}{\frac{1}{2} z - 1} = \sum_{k=1}^{\infty} 2^{-k}\, z^k \quad \text{ if } |z| < 2.$$

(c) Compute the impulse response function h for this filter.

(d) Is the filter stable? Is the impulse response causal?

Exercise 8.9. FIR filters with linear phase.

Prove that a digital FIR filter has a linear phase if and only if its coefficients are symmetric about its center coefficient.

Exercise 8.10. Polyphase quadrature filter.

Consider the polyphase filter bank from (136) with h_0 only 512 non-zero coeeficients, i.e., $h_0(k) = 0$ if $k < 0$ or $k > 511$.

(a) Show that

$$\begin{aligned} f_i(t_k) = f * h_i(t_k) &= \sum_{j=0}^{511} f(t_{k-j})\, 2h_0(j)\, \cos\!\left(\frac{\pi}{64}(2i+1)j\right) \\ &= \sum_{n=0}^{63} M_{i,n} \sum_{l=0}^{7} \widetilde{h}_0(n + 64l)\, f(t_{k-n-64l}), \end{aligned}$$

where $M_{i,n} \equiv \cos(\frac{\pi}{64}(2i+1)n)$ and $\widetilde{h}_0(n+64l) \equiv 2(-1)^l h_0(n+64l)$.

(b) Show that the first sum requires 32×512 multiplications to compute $f_i(t_k)$ $(i = 1, \ldots, 32)$ for each k, whereas the second sum only requires 5×512 multiplications per k. Here we have averaged the total number of multiplications over all k and assumed that k is large.

Exercise 8.11. **Resampling.**

Let $f \in L^2(\mathbb{R})$ such that $\widehat{f}(\omega) = 0$ if $|\omega| > \Omega$. Let $\Delta t \equiv 1/(2\Omega)$, and $t_n = n\Delta t$. Suppose that only f-values $f_n \equiv f(t_n)$ at the sample points t_n are available and, for some $p, q \in \mathbb{N}$, we want to compute f values at the sequence $n\Delta t'$, where $\Delta t' \equiv \frac{q}{p}\Delta t$: we want to *resample* the sequence (f_n) of f-values.

(a) Let $g \in L^2_T(\mathbb{R})$ with Fourier coefficients γ_k. Note that g is also pT-periodic. Let $\widetilde{\gamma}_k$ be the Fourier coefficients of g as a pT-periodic function. Show that

$$\widetilde{\gamma}_{kp} = p\gamma_k \quad (k \in \mathbb{Z}) \qquad \text{and} \qquad \widetilde{\gamma}_j = 0 \ \text{ if } j \in \mathbb{Z}\backslash p\mathbb{Z}.$$

(b) Let F be such that $\widehat{F}(\omega) = \widehat{f}(\omega - 2k\Omega)$ if $|\omega| \leq p\,\Omega$ with $k \in \mathbb{Z}$ such that $|\omega - k\Omega| \leq \Omega$ and $\widehat{F}(\omega) = 0$ elsewhere. Put $h \equiv \Delta t/p$. Show that

$$F(nph) = pf(n\Delta t) \quad (n \in \mathbb{Z}) \qquad \text{and} \qquad F(jh) = 0 \text{ if } j \in \mathbb{Z}\backslash p\mathbb{Z}.$$

(c) To have $f(j\Delta t')$ we proceed by subsequentially (i) upsample, (ii) filter, (iii) downsample.

(i) First we *upsample* (f_n) by p, denoted by $(\uparrow p)(f_n)$: with $(\widetilde{f}_j) \equiv (\uparrow p)(f_n)$

$$\widetilde{f}_{np} \equiv f(n\Delta t) \ \text{ for } \ n \in \mathbb{Z} \qquad \text{and} \qquad \widetilde{f}_j \equiv 0 \ \text{ if } \ j \in \mathbb{Z}\backslash p\mathbb{Z}.$$

In other words, the sequence (f_n) is extended to a sequence $(\widetilde{f}_j)$ by adding $p-1$ zeros after every f_n. Show that $\widetilde{f}_j = \frac{1}{p}F(jh)$ $(j \in \mathbb{Z})$. Show that the Fourier transform of $(\widetilde{f}_j)$ is the $2p\Omega$-periodic function that is equal to $\widehat{F}$ on $[-p\Omega, p\Omega]$.

(ii) Secondly, we *apply a filter* to get rid of the high frequencies. Ideally, in the frequency domain, we multiply the Fourier transform of $\widetilde{f}$ by the $2p\Omega$-periodic function Π_Ω, i.e., $\Pi_\Omega(\omega) = 1$ if $|\omega| \leq \Omega$ and $\Pi_\Omega(\omega) = 0$ if $\Omega < |\omega| \leq p\Omega$: $g = \widetilde{f} * \widehat{\Pi}_\Omega$. In practice, a digital filter (IIR) is applied to $(\widetilde{f}_j)$ that approximates the effect of the ideal filter. Show that, in the ideal case, the filtered result, say (g_j), is such that $g_j = \frac{1}{p}\,f(jh)$ for all $j \in \mathbb{Z}$.

(iii) Finally, we *downsample* the sequence (g_j) by q, denoted by $(\downarrow q)(g_j)$, i.e., with $(\widetilde{g}_j) = (\downarrow q)(g_j)$, we have that $\widetilde{g}_j \equiv g_{jq}$ $(j \in \mathbb{Z})$. Show that $\widetilde{g}_j = \frac{1}{p}f(j\Delta t')$ $(j \in \mathbb{Z})$.

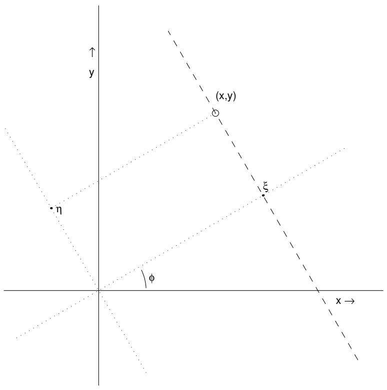


FIGURE 29. The dashed line is at distance ξ from the origin and has angle $\phi + \frac{1}{2}\pi$ with the x-axis. The point (x, y) has coordinates (ξ, η) in the rotated axis system: $x = \xi \cos(\phi) - \eta \sin(\phi)$, $y = \xi \sin(\phi) + \eta \cos(\phi)$.

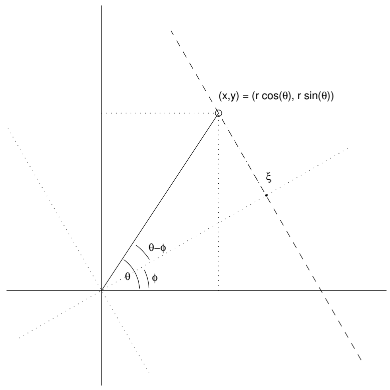


FIGURE 30. If (r, ϑ) are the polar coordinates of (x, y), $(x, y) = (r\cos(\vartheta), r\sin(\vartheta))$, then (x, y) has polar coordinates $(r, \vartheta - \phi)$ in the rotated system, where the rotation of the Cartesian coordinates is over an angle ϕ.

9 Computerized Tomography (CT)

Before we give the physical background in §9.2 of X-ray computerized tomography, we firstly give the mathematical formulation of the problem.

9.1 A reconstruction problem. Consider a complex-valued map f on $\mathbb{R}^2$. For each $\phi \in [0, \pi)$ and each $\xi \in \mathbb{R}$, let $p_\phi(\xi)$ be defined by

$$p_\phi(\xi) \equiv \int f(x(\eta), y(\eta))\ \mathrm{d}\eta, \tag{137}$$

where $x(\eta) \equiv \xi\cos(\phi) - \eta\sin(\phi)$ and $y(\eta) \equiv \xi\sin(\phi) + \eta\cos(\phi)$ $(\eta \in \mathbb{R})$. The map $\eta \rightsquigarrow (x(\eta), y(\eta))$ describes a straight line in the (x, y)-plane at distance ξ from the origin. The line has angle $\phi + \frac{1}{2}\pi$ with the x-axis (see Fig. 29). The function f in (137) is integrated along this line: p_ϕ can be viewed as a projection of f on a straight line with angle ϕ with the x-axis (the 'ξ-axis'). For $\phi = 0$, we have that $p_0(\xi) = \int f(\xi, \eta)\ \mathrm{d}\eta$. The map $(\xi, \phi) \rightsquigarrow p_\phi(\xi)$ is known as the *Radon transform* of f. The 2-dimensional graph of the Radon transform $(\xi, \phi) \rightsquigarrow p_\phi(\xi)$ is called the *sinogram of* f.

Now, suppose that only the values $p_\phi(\xi)$ are available ($\xi \in \mathbb{R}, \phi \in [0, \pi)$). Then the question is how to reconstruct f from these values $p_\phi(\xi)$. It turns out that this is possible via two *one* dimensional Fourier transforms (see Theorem 9.4). However, the proof of this result relies on two dimensional Fourier transforms (see 9.3).

9.2 CT scans. We took the text in the following two paragraphs from [4], pp.325–326. For more details, we refer to [6, Ch.3].

As a beam of x-rays passes through a uniform slab of material, the intensity of the beam decreases according to $I = I_0 e^{-\kappa d}$, where I_0 is the initial intensity, d is the thickness of the slab, and κ is an appropriate coefficient. κ is often called an absorption coefficient, but there are actually many processes occurring that diminish the intensity of the beam. The detector is constructed to measure the intensity of the beam that passes straight through the object, so that x-rays that are scattered will not be detected. The reduction of the signal in *computerized tomography* (CT) is primarily

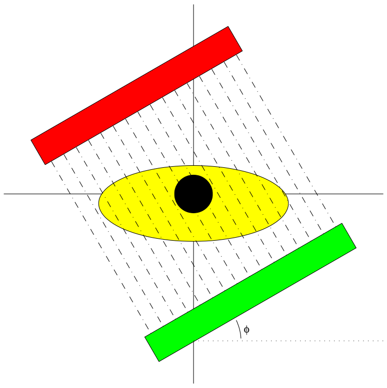


FIGURE 31. x-rays are directed in a series of thin, pencil like parallel beams through the image plane, and the attenuation of each beam is measured.

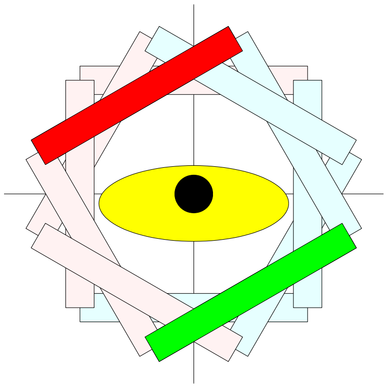


FIGURE 32. The scanner will circle around the object. The attenuation of beams is measured for a range of angles ϕ. The light gray rectangles mark past and future positions of the scanner (beamer (gray) and detector (dark gray)).

due to scattering out of the forward direction. We will refer to the loss of intensity, to whatever cause, as *attenuation*.

In CT, a two dimensional slice through an object is imaged. In its simplest configuration, x-rays are directed in a series of thin, pencil like parallel beams through the image plane, and the attenuation of each beam is measured (as picturized in Fig. 31). Since the object is nonuniform, the attenuation of the beam varies from point to point, and the total attenuation is given as an integral along the path of the beam. That is, if f is an 'absorption' function on $\mathbb{R}^2$, then the intensity at the detector of the beam at distance ξ from the origin at angle $\phi+\frac{1}{2}\pi$ with the x-axis will be of the form

$$I = I_0\, e^{-\int f(x(\eta),y(\eta))\,\mathrm{d}\eta},$$

where $\eta \rightsquigarrow (x(\eta),y(\eta))$ is the parametrization of the beam as in 9.1. From the collection of beams a profile of the object can be built. The 'projection at ϕ' can be obtained by evaluating the logarithm of the measured I/I_0 ratio,

$$p_\phi(\xi) = -\ln\left(\frac{I}{I_0}\right) = \int f(x(\eta),y(\eta))\,\mathrm{d}\eta.$$

Note that f can not be reconstructed if only the values $p_\phi(\xi)$ are available at a fixed angle ϕ (then the 'bone' can be in the front as well in the back): the scanner has to circle around the object (see Fig. 32). Note that the result of a scan at ϕ and at $\phi+\pi$ will be the same.

A three dimensional image can be obtained by piling the images of two dimensional slices.

9.3 To reconstruct f from the $p_\phi(\xi)$ values, we first represent the Fourier transform with respect to a rotated basis.

Note that, for a fixed ϕ, each pair (x,y) can be written as

$$x = \xi\cos(\phi) - \eta\sin(\phi), \quad y = \xi\sin(\phi) + \eta\cos(\phi) \tag{138}$$

for an appropriate pair (ξ,η) (see Fig. 29): (138) rotates the axis over an angle ϕ. In two dimensions, (82) reads as

$$\widehat{f}(\omega_1,\omega_2)=\int\!\!\int f(x,y)\,e^{-2\pi i(x\omega_1+y\omega_2)}\,\mathrm{d}x\;\mathrm{d}y.$$

We substitute the expressions (138) for x and y in this integral. To simplify the resulting expression for $x\omega_1+y\omega_2$ in the exponential function, we also rotate the axis in the (ω_1,ω_2)-plane over the angle ϕ:

$$\omega_1=\rho_1 c_\phi-\rho_2 s_\phi,\quad \omega_2=\rho_1 s_\phi+\rho_2 c_\phi,\quad \text{where } c_\phi\equiv\cos(\phi)\text{ and } s_\phi\equiv\sin(\phi).$$

Then, $x\omega_1+y\omega_2=\xi\rho_1+\eta\rho_2$. To see this, recall that the inner product between vectors does not change by representing these vectors with respect to another orthonormal basis. Hence,

$$\begin{array}{l}\widehat{f}(\rho_1 c_\phi-\rho_2 s_\phi,\rho_1 s_\phi+\rho_2 c_\phi)=\\ \qquad\iint f(\xi c_\phi-\eta s_\phi,\xi s_\phi+\eta c_\phi)\,e^{-2\pi i(\xi\rho_1+\eta\rho_2)}\,\mathrm{d}\xi\;\mathrm{d}\eta.\end{array}$$

In particular, for $\rho_2=0$,

$$\widehat{f}(\rho_1 c_\phi,\rho_1 s_\phi)=\int\!\!\int f(\xi c_\phi-\eta s_\phi,\xi s_\phi+\eta c_\phi)\,e^{-2\pi i\xi\rho_1}\,\mathrm{d}\xi\;\mathrm{d}\eta.$$

Swapping the order of integration leads to

$$\begin{array}{rcl}\widehat{f}(\rho_1 c_\phi,\rho_1 s_\phi)&=&\iint f(\xi c_\phi-\eta s_\phi,\xi s_\phi+\eta c_\phi)\,e^{-2\pi i\xi\rho_1}\,\mathrm{d}\eta\;\mathrm{d}\xi\\ &=&\int\left[\int f(\xi c_\phi-\eta s_\phi,\xi s_\phi+\eta c_\phi)\,\mathrm{d}\eta\right]e^{-2\pi i\xi\rho_1}\,\mathrm{d}\xi.\end{array}$$

Now, note that the expression between square brackets is precisely $p_\phi(\xi)$. Hence,

$$\widehat{f}(\rho\,c_\phi,\rho\,s_\phi)=\int p_\phi(\xi)\,e^{-2\pi i\xi\rho}\,\mathrm{d}\xi. \tag{139}$$

Here, we dropped the index 1 of ρ to simplify notation. This last integral is the one dimensional Fourier transform of the function p_ϕ at frequency ρ: $\widehat{f}(\rho\,c_\phi,\rho\,s_\phi)=\widehat{p}_\phi(\rho)$.

Although, we derived expression (139) for a fixed ϕ, the expression is correct for any ϕ: ϕ was fixed, but arbitrary. This observation turns (139) into a remarkable result, since any point (ω_1,ω_2) in $\mathbb{R}^2$ can be written in polar coordinates as $(\rho\,c_\phi,\rho\,s_\phi)$ for appropriate ρ and ϕ. In other words, the one dimensional Fourier transforms in (139) yields $\widehat{f}$ in any (ω_1,ω_2) of $\mathbb{R}^2$.

Now, in principle, we can apply the two dimensional inverse Fourier transform to find f. Although this approach is theoretically correct, in practice it is inaccurate. In practice, $p_\phi(\xi)$ will only be available for a discrete set of points $\phi=\phi_k=k\Delta\phi$ and $\xi=\xi_\ell=\ell\Delta\xi$. Since, to compute $\widehat{f}$, we will apply FFT to (139), $\widehat{f}$ will be computed at a grid that is rectangular in polar coordinates, that is, at $(\rho\,c_\phi,\rho\,s_\phi)$ for $(\rho,\phi)=(\rho_\ell,\phi_k)=(\ell\Delta\rho,k\Delta\phi)$ (see Fig. 33). To compute f from $\widehat{f}$, using inverse FFT based on (83), we rather would like to have the function values of $\widehat{f}$ at a grid that is rectangular in Cartesian coordinates, that is, at (ω_1,ω_2) for $(\omega_1,\omega_2)=(\omega_{1,i},\omega_{2,j})=(i\Delta\omega,j\Delta\omega)$. Computing $\widehat{f}$ at the points of this 'rectangular Cartesian' grid with some interpolation formula from the 'rectangular polar' grid introduces errors and these errors will be amplified by the subsequent inverse Fourier transform. Note that the distribution of

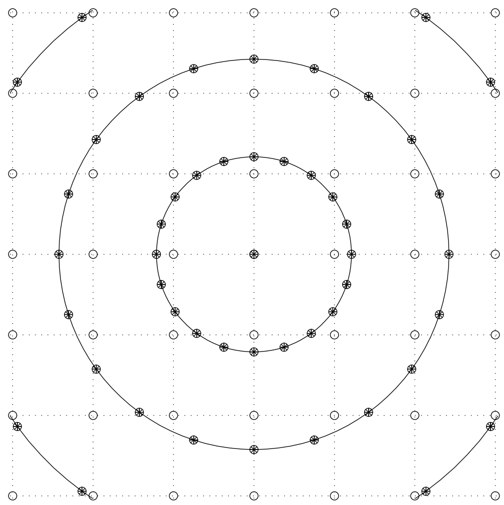


FIGURE 33. The solid dots mark a grid that is rectangular in polar coordinates, while the circles mark a grid that is rectangular in Cartesian coordinates. Note that the density of the grid points in the grid that is rectangular in polar coordinates is non uniform in Cartesian coordinates. The density is high close to the origin and low towards infinity.

the grid points of the rectangular polar grid is very nonuniform with respect to Cartesian coordinates (see Fig. 33).

Therefore, as a more accurate alternative, we will express the inverse Fourier transform

$$f(x,y)=\iint \widehat{f}(\omega_1,\omega_2)\,e^{2\pi i(x\omega_1+y\omega_2)}\;\mathrm{d}\omega_1\;\mathrm{d}\omega_2 \tag{140}$$

(see (83)) in polar coordinates: $(\omega_1,\omega_2)=(\rho\,c_\phi,\rho\,s_\phi)$. We will let ρ range from $-\infty$ to ∞ and ϕ in $[0,\pi)$, whereas the standard choice is $\rho\in[0,\infty)$, $\phi\in[0,2\pi)$. The absolute value of the determinant of the Jacobian matrix of the transform to polar coordinates is $|\rho|$. Again, in order to simplify $x\omega_1+y\omega_2$, we will use polar coordinates for (x,y) as well: $(x,y)=(rc_\vartheta,rs_\vartheta)$. Then

$$x\,\omega_1+y\,\omega_2=r\,c_\vartheta\,\rho\,c_\phi+r\,s_\vartheta\,\rho\,s_\phi=r\rho\,(c_\vartheta c_\phi+s_\vartheta s_\phi)=r\rho\,c_{\vartheta-\phi}.$$

Hence, (140) reads as

$$f(rc_\vartheta,rs_\vartheta)=\int_0^\pi\int_{-\infty}^\infty \widehat{f}(\rho\,c_\phi,\rho s_\phi)\,e^{2\pi i r\rho\,c_{\vartheta-\phi}}\,|\rho|\;\mathrm{d}\rho\;\mathrm{d}\phi.$$

Put $\xi\equiv rc_{\vartheta-\phi}$. Note that this definition is consistent with the use of ξ above (see Fig. 30). Then,

$$f(rc_\vartheta,rs_\vartheta)=\int_0^\pi\left[\int_{-\infty}^\infty |\rho|\,\widehat{f}(\rho\,c_\phi,\rho s_\phi)\,e^{2\pi i\xi\rho}\;\mathrm{d}\rho\right]\;\mathrm{d}\phi$$

If we denote the expression in square brackets by $\widetilde{p}$, then we have that

$$f(rc_\vartheta,rs_\vartheta)=\int_0^\pi \widetilde{p}_\phi(\xi)\;\mathrm{d}\phi,\quad \text{where}\quad \widetilde{p}_\phi(\xi)\equiv\int_{-\infty}^\infty |\rho|\,\widehat{f}(\rho\,c_\phi,\rho s_\phi)\,e^{2\pi i\xi\rho}\;\mathrm{d}\rho.$$

The following theorem summerizes our results.

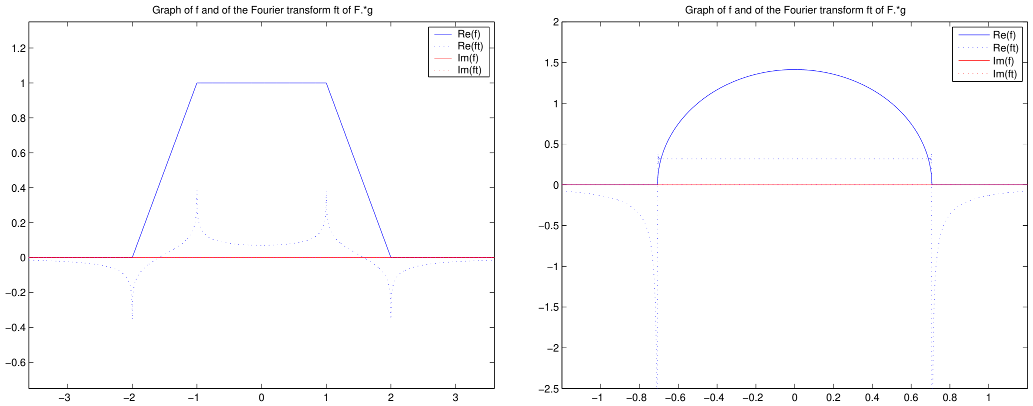


FIGURE 34. The pictures show the projection p_ϕ (the solid line) and the filtered projection $\widetilde{p}_\phi$ (the dotted line) of a function f for some angle ϕ. In the left picture, $\phi\in(0,\pi/4)$ and f is constant (non zero) on some square $[-a,a]^2$ and 0 elsewhere. In the right picture, f is equal to 1 on the disc $\{(x,y): x^2+y^2\leq 0.5\}$ and 0 elsewhere. Note that in this second situation the projection p_ϕ is the same for each ϕ. The filtered projection $\widetilde{p}_\phi$ in the right picture is constant on the interval of ξ's with $\xi^2\leq 1/2$. The Gibb's phenomenon in $\widetilde{p}_\phi$ at the end of the interval is the due to the fact that for the computation a discretized version of the Fourier transform has been used.

9.4 Theorem *For $f\in L^2(\mathbb{R}^2)$, $\xi\in\mathbb{R}$ and $\phi\in[0,\pi)$, let*

$$p_\phi(\xi)\equiv\int_{-\infty}^{\infty} f(x(\eta),y(\eta))\,\mathrm{d}\eta, \tag{141}$$

where $x(\eta)\equiv\xi\cos(\phi)-\eta\sin(\phi)$, $y(\eta)\equiv\xi\sin(\phi)+\eta\cos(\phi)$ $(\eta\in\mathbb{R})$. Then, with

$$\widehat{p}_\phi(\rho)=\int_{-\infty}^{\infty} p_\phi(\xi)\,e^{-2\pi i\xi\rho}\,\mathrm{d}\xi \quad \textit{and} \quad \widetilde{p}_\phi(\xi)\equiv\int_{-\infty}^{\infty}|\rho|\,\widehat{p}_\phi(\rho)\,e^{2\pi i\xi\rho}\,\mathrm{d}\rho, \tag{142}$$

we have that

$$f(r\cos(\vartheta),r\sin(\vartheta))=\int_0^\pi \widetilde{p}_\phi(r\cos(\vartheta-\phi))\,\mathrm{d}\phi \qquad (r\in[0,\infty),\vartheta\in[0,2\pi)) \tag{143}$$

□

Note that, due to the additional factor $|\rho|$, $\widetilde{p}_\phi$ is not exactly the inverse Fourier transform of $\widehat{p}_\phi$. Nevertheless, both $\widehat{p}_\phi$ and this 'modified' projection $\widetilde{p}_\phi$ are one dimensional Fourier transforms and FFT and inverse FFT, respectively, can be employed for efficient computation. Note that the map $p_\phi\rightsquigarrow\widetilde{p}_\phi$ can be viewed as a filter operation: $\widetilde{p}_\phi$ is the inverse Fourier transform of $\widehat{p}_\phi\,H$, where $H(\rho)\equiv|\rho|$.

If p_ϕ is the projection of f, then it might be reasonable to call

$$\widetilde{f}(r\cos(\vartheta),r\sin(\vartheta))\equiv\int_0^\pi p_\phi(r\cos(\vartheta-\phi))\,\mathrm{d}\phi \qquad (r\in[0,\infty),\vartheta\in[0,2\pi))$$

the back projection (why?) and (143) is a *filtered back projection* (FBP). The back projection is a blurred version of f, whereas the filtered back projection is exactly equal to f.

Suppose that, for a given angle ϕ, the projection values $p_\phi(\xi_\ell)$ are available for $\xi_\ell=\ell\Delta\xi$. Then, note that the FFTs will yield values of the 'modified' projection $\widetilde{p}_\phi$ at the same points ξ_ℓ.

Exercises

Exercise 9.1. Computerized Tomography project.
Select an image array (x_i, y_j), that is, for a set $x_i = x_0 + i\Delta x$ and $y_j = y_0 + j\Delta y$, we want to compute the image values $f_{i,j} \equiv f(x_i, y_j)$ at (x_i, y_j) (for a more precise definition, see (144)): the values for $f_{i,j}$ that we obtain by computation form the 'reconstructed' or 'output' image.

We suppose that the projected values $p_\phi(\xi)$ are available (can be measured) for a set of angles $\phi = \phi_k = k\Delta\phi$ and, for each ϕ in this set, for a set of points $\xi = \xi_\ell = \ell\Delta\xi$.

Alg. 1 uses this input information to reconstruct the image.

A. Initialize the output image to 0 ($f_{i,j} = 0$ for all i, j).

B. For each angle ϕ in the collection $\{\phi_k\}$ do:

1. 'Measure' the projections $p_\phi(\xi)$ at the points $\xi = \xi_\ell$.

2. Use FFT to compute $\widehat{p}_\phi(\rho)$ (see (142)).

3. Use inverse FFT to compute $\widetilde{p}_\phi(\xi)$ at the points $\xi = \xi_\ell$ (see (142)).

4. Use the results for this one projection to update the output image (see (143)), that is, for each (x, y) in the set (x_i, y_j) do:

a. Determine ξ from (138).

b. Approximate the modified projection at ξ by interpolation of the $p_\phi(\xi_\ell)$ (linear interpolation is usually sufficient).

c. Add this contribution to the ϕ integral (143) to the image.

5. Display the image obtained thus far (optional).

ALGORITHM 1. An outline of an image reconstruction algorithm based on the theory in §9. For a motivation to use linear interpolation in step 4.b, see the Paragraph "Computing the projections" below.

In practice, of course, the projections in step 1 would be obtained from the CT scanning device. For our purposes, we compute them from an "input" image.[35] The reconstructed image is the output image.

Use the outline in Alg. 1 to write a computer program to reconstruct images.[36] You may assume that the images are contained in a circle of unit radius.[37]

Write a report that is readable for someone who is familiar with Fourier Theory but who does not have a copy of the lecture notes. In your report you may address the following issues:

— Investigate the effect of changing the resolution (i.e., the size of Δx and Δy. You may take $\Delta x = \Delta y$. Consider, for instance $\Delta x = 2/100, 2/200, \ldots$). Do you need the same step sizes Δx and Δy for the reconstruction as for the projections? How do you relate the step size Δx and Δy for reconstruction to the step size $\Delta\xi$ and $\Delta\phi$?

— Discuss the advantages of this approach with one dimensional inverse Fourier transforms followed by an integral over ϕ as compared to an approach based on one two dimensional inverse Fourier transform.

— What do you expect from replacing linear interpolation in step 4b by cubic interpolation? (i.e., approximate the value $\widetilde{p}_\phi(\xi)$ by the value $P(\xi)$ of the cubic polynomial P for which $P(\xi_{\ell-1}) = \widetilde{p}_\phi(\xi_{\ell-1})$, $P(\xi_\ell) = \widetilde{p}_\phi(\xi_\ell)$, $P(\xi_{\ell+1}) = \widetilde{p}_\phi(\xi_{\ell+1})$, $P(\xi_{\ell+2}) = \widetilde{p}_\phi(\xi_{\ell+2})$, where ℓ is such that $\xi \in (\xi_\ell, \xi_{\ell+1})$. How do you define P near the boundary?).

— FFT 'assumes' that the input function is periodic. This periodicity can be achieved by first restricting the image function f to a rectangle (multiplication by a 2-d tophat function?),

[35] The code for computing the projections will be provided for.

[36] You can use existing codes for FFT and inverse FFT.

[37] The MATLAB file `Images.m` can be used for producing images. The file gives several ways of producing image files that can be used in testing the image reconstruction code.

followed by a periodic extension (after an even extension?). What is the effect of the restriction and the extension on the reconstructed image? The function $H(\rho) \equiv |\rho|$ that defines the 'filter operation' (multiplcation of $\widehat{p}_\phi$ by $|\rho|$) is not bounded nor in $L^2(\mathbb{R})$. Is this a problem?

For more details, we refer to [6, Ch.3].

Discretization. In practice, the values of $p_\phi(\xi)$, for $\xi \in \mathbb{R}$, $\phi \in [0, \pi)$, or for $\xi = \xi_\ell = \ell\Delta\xi$ and $\phi = \phi_k = k\Delta\phi$, are available from measurements. For the design and the analyse of an effective numerical solution process it is usefull to have a discrete model of the problem and to have discrete versions of the operations that are involved.[38]

Averaging versus sampling. A picture of which each pixel is either black or white gives the impression of being composed of patches of different shades of gray when viewed from a distance. When in an approximation that uses larger pixels the color is determined by *sampling*, the shades of gray disappear and black and white blocks become visible. This is undesirable. Averiging leads to shades of gray. Therefore, discretization by *averaging* seems to be a more appropriate approach here: specifically in Image Processing, discretization by averaging leads to pictures that are more appealing to the eye, then when discretized by sampling.

If the real-valued function f that is defined on the square $[-a, a] \times [-a, a]$ represents a picture,[39] then f can be approximated as follows by a piece-wise constant function. Select a step size Δx in the x-direction and a step size Δy in the y direction ($\Delta x = 2a/N_x$ and $\Delta y = 2a/N_y$). Then, $(x_i, y_j) \equiv ((i - \frac{1}{2})\Delta x, (j - \frac{1}{2})\Delta y) - (a, a)$ is the center of the pixel area $I_{i,j} \equiv [x_i - \frac{1}{2}\Delta x, x_i + \frac{1}{2}\Delta x) \times [y_j - \frac{1}{2}\Delta y, y_j + \frac{1}{2}\Delta y)$ $(i = 1, \ldots, N_x, j = 1, \ldots, N_y)$, and we compute the image value $f_{i,j}$ for this pixel by averaging,

$$f_{i,j} \equiv \frac{1}{\Delta x\, \Delta y} \iint_{I_{i,j}} f(x, y)\ \mathrm{d}x\ \mathrm{d}y : \tag{144}$$

f is approximated by the function f_d that has the value $f_{i,j}$ on $I_{i,j}$ $(i = 1, \ldots, N_x, j = 1, \ldots, N_y)$.

This approach suggests the following discretization for the projection $p_\phi(\xi)$. Select a step size $\Delta\xi = 2/(N_\xi - 1)$. Then, for $\xi_\ell \equiv -1 + (\ell - 1)\Delta\xi$, define

$$p_{\ell,\phi} \equiv \frac{1}{\Delta\xi} \int_{\xi_\ell - \frac{1}{2}\Delta\xi}^{\xi_\ell + \frac{1}{2}\Delta\xi} p_\phi(\xi)\ \mathrm{d}\xi \qquad (\ell = 1, \ldots, N_\xi). \tag{145}$$

In the exposition below we will assume that f is equal to the discretized version: $f = f_d$.

Computing the projections. Note that p_ϕ is the sum of all projected pixels. To be more precise, if $\chi_{i,j}$ is the function with value 1 on $I_{i,j}$ and zero elsewhere, then

$$p_\phi(\xi) = \sum_{i,j} f_{i,j} \int \chi_{i,j}(x(\eta), y(\eta))\ \mathrm{d}\eta. \tag{146}$$

Here, we sum over all $i = 1, \ldots, N_x$ and all $j = 1, \ldots, N_y$, and $x(\eta)$ and $y(\eta)$ are as defined in §9.1. Except for the color value $f_{i,j}$, the function

$$\pi_\phi^{(i,j)}(\xi) \equiv \int \chi_{i,j}(x(\eta), y(\eta))\ \mathrm{d}\eta$$

is the projected pixel. For a graph of this function, see Fig. 35. Note that the integral of π_ϕ

[38] The discretization as explained here, is also used in the MATLAB code for this assigment, the code that computes the 'measured' projection values $p_\phi(\xi)$.

[39] We assume the picture to be defined on a square contained in the unit disk $\{(x, y) \,\big|\, |x|^2 + |y|^2 \le 1\}$, that is, $a \equiv \frac{1}{2}\sqrt{2}$. Obviously, this is not a restriction, since each picture can be scaled to this situation. Working on the unit disk slightly simplifies the coding when using polar coordinates.

For a full color picture, we need three color functions f_r, f_g, and f_b, where, for instance, f_r defines the red intensities, f_g the green intensities, and f_b the blue intensities. However, in CT the colors are artificial anyway. We, therefore, assume the image to be defined by one color function f only. The function value $f(x, y)$ represent the intensity of the color at (x, y).

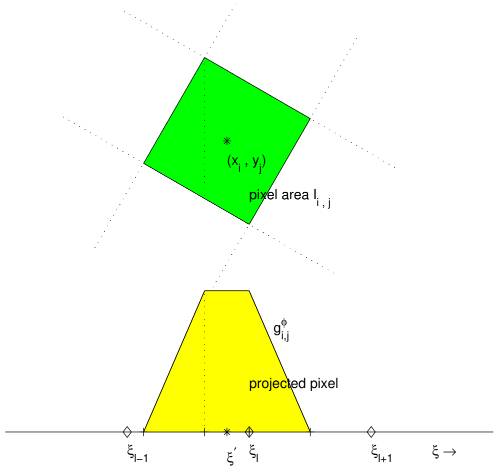


FIGURE 35. The picture represents a pixel (with center (x_i, y_j)) and its projection π_ϕ on a ξ-axis. The projection is on the line with angle ϕ with the x axis (ϕ is here $-\pi/3 = 2\pi/3$). Rather than rotating the projection line over an angle ϕ, the pixel area has been rotated over an angle $-\phi$ in this picture.

over ξ is precisely equal to $\Delta x\,\Delta y$ (why?). The contribution of the (i,j)th pixel to $p_{\ell,\phi}$ is equal to

$$f_{i,j}\,\frac{1}{\Delta\xi}\int_{\Xi_\ell}\pi_\phi^{(i,j)}(\xi)\,\mathrm{d}\xi., \quad \text{where } \Xi_\ell \equiv [\xi_\ell - \tfrac{1}{2}\Delta\xi, \xi_\ell + \tfrac{1}{2}\Delta\xi).$$

A matrix representation. For coding purposes, it is convenient to number the pixels as well as the gridpoints in the (ξ,ϕ)-plane. We suggest lexicographical numberings: the mth pixel is the pixel with center (x_i, y_j) with i and j such that $m = i + (j-1)N_x$; similarly, the nth point of the (ξ,ϕ)-grid is (ξ_ℓ, ϕ_k) with $n = \ell + (k-1)N_\xi$. Then the picture f corresponds to a vector $\mathbf{x}$ with coordinate x_m equal to the image value $f_{i,j}$, where $m = i + (j-1)N_x$. The projected values p_{ℓ,ϕ_k} are represented by the vector $\mathbf{b}$ with nth coordinate b_n equal to p_{ℓ,ϕ_k}, where $n = \ell + (k-1)N_\xi$. Note that now the reconstruction problem can be formulated as a matrix-vector equation:

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad \text{where } A_{n,m} \equiv \frac{1}{\Delta\xi}\int_{\Xi_\ell}\pi_{\phi_k}^{(i,j)}(\xi)\,\mathrm{d}\xi, \tag{147}$$

and m, i, j and n, ℓ, k related as before.

The nth point in the (ξ,ϕ)-grid corresponds to the strip of beams that, at angle ϕ_k, hit the ℓth ξ-interval Ξ_ℓ: b_n is the projected result from this nth strip of beams. The multiplication $\mathbf{A}_{n,:}\,\mathbf{x}$ of the nth row $\mathbf{A}_{n,:}$ of the matrix $\mathbf{A}$ with $\mathbf{x}$ representes the effect of the beams in this nth strip. The matrix entry $A_{n,m}$ is equal to the scaled surface of the interesection of the mth pixel and the nth strip of beams. Since most of the strips of beams mis most of the pixels, each row of $\mathbf{A}$, and therefore $\mathbf{A}$ itself, is sparse.

Further simplifications. If $\pi_\phi^{(i,j)}$ is zero outside the ξ-interval Ξ_ℓ, then the contribution of the (i,j)th pixel to $p_{\ell,\phi}$ is precisely equal to $f_{i,j}\,\frac{\Delta x\,\Delta y}{\Delta\xi}$. Of course, it depends on the diameter of the pixel (on Δx and Δy) and on $\Delta\xi$, but, unfortunately, in general, $\pi_\phi^{(i,j)}$ will not be non-zero on one interval only.[40] To speed up computations, we will not compute what part of $\pi_\phi^{(i,j)}$ is non-zero on each of the intervals Ξ_ℓ, but, we will simply follow the following strategy.

If ξ' is the ξ-coordinate of the center (x_i, y_j) of the (i,j)th pixel (i.e., $x_i = \xi' c - \eta' s$, $y_j = \xi' s + \eta' c$ where $c \equiv \cos(\phi)$ and $s \equiv \sin(\phi)$), and ℓ is such that $\xi' \in [\xi_\ell, \xi_{\ell+1})$, then we add

[40] If $\Delta x + \Delta y \ll \Delta\xi$ then most of the π_ϕ will be non-zero on only one interval. However, certainly in the reconstruction phase, the pixels will be selected such that their diameters are comparable to $\Delta\xi$. Then, $\pi_\phi^{(i,j)}$ will generally be non-zero on two intervals.

the fraction $f_{i,j}\,\frac{\Delta x\,\Delta y}{\Delta\xi}\,\frac{\xi'-\xi_\ell}{\xi_{\ell+1}-\xi_\ell}$ to $p_{\ell+1,\phi}$ and the remaining fraction $f_{i,j}\,\frac{\Delta x\,\Delta y}{\Delta\xi}\,\frac{\xi_{\ell+1}-\xi'}{\xi_{\ell+1}-\xi_\ell}$ to $p_{\ell,\phi}$: we follow a 'linear interpolation' type of strategy.

Alternative algebraic approaches. As explained in (147), the problem of reconstructing the $f_{i,j}$ values from the values of $p_\phi(\xi)$ at $\phi=\phi_k$ and $\xi=\xi_\ell$ can be formulated as a matrix-vector equation. This formulation suggests to use numerical linear algebra methods for obtaining a solution.

Unfortunately, it is unlikely that $\mathbf{A}$ is square, $\mathbf{A}$ may not even have full column rank. Moreover, due to errors in the measurements and errors from the discretization $\mathbf{b}$ will not be equal to $\mathbf{Ax}$. Therefore, we can not simply apply a straight-forward method as Gauss elimination. The solution of the matrix-vector equation may not exist, if it exists it may not be unique and if it is exists uniquely, it may not be the one that we are interested in. Therefore, we first have to decide on the 'solution' that we want to compute. If we decide to follow the algebraic approach, we actually decide to discart the structure in the problem (the structure that allows for a solution using Fourier transforms) and the 'solution' that we will try to compute has to be characterized by algebraic properties.

Finding the solution with smallest error is impossible since the exact solution is unknown. As an alternative, we can try to compute a solution $\widetilde{\mathbf{x}}$ with smallest residual norm. The *residual* $\mathbf{r}$ is defined by $\mathbf{b}-\mathbf{A}\widetilde{\mathbf{x}}$. The residual can be computed and if there is an $\mathbf{x}$ for which $\mathbf{Ax}=\mathbf{b}$ then the residual is $\mathbf{A}$ times the error $\mathbf{x}-\widetilde{\mathbf{x}}$.

Note that

$$\mathbf{x}=\text{minarg}\{\|\mathbf{b}-\mathbf{A}\widetilde{\mathbf{x}}\|_2\,\Big|\,\widetilde{\mathbf{x}}\}\quad\Leftrightarrow\quad\mathbf{b}-\mathbf{Ax}\perp\mathbf{A}\widetilde{\mathbf{x}}\quad\forall\widetilde{\mathbf{x}}\quad\Leftrightarrow\quad\mathbf{A}^{\mathrm{T}}\mathbf{Ax}=\mathbf{A}^{\mathrm{T}}\mathbf{b}:$$

the *minimal residual solution* or *least square solution* is equal to the solution of the *normal equations*. The matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is square. However, if $\mathbf{x}$ is a minimal residual solution and if $\mathbf{Az}=\mathbf{0}$ then $\mathbf{x}+\mathbf{z}$ is also minimal residual. To have uniqueness, one often computes the minimal residual solution $\mathbf{x}$ with smallest norm. This solution is characterized by the property $\mathbf{x}\perp\mathbf{z}$ if $\mathbf{Az}=\mathbf{0}$, or equivalently, $\mathbf{x}=\mathbf{A}^{\mathrm{T}}\mathbf{y}$ for some $\mathbf{y}$ (why?). The *minimum residual minimum norm solution* exists and is unique. Iterative methods for computing a minimum residual solution (as Conjugate Gradients for the Normal Equations (CGNE)) produce the minimum residual minimum norm solution. Unfortunately, due to errors in $\mathbf{b}$, the min.res. min.norm solution is often a noisy version of the true image, and regularization has to be included to damp the noise.

The Fourier approach requires a homogenuous distribution of beams, but is fast and avoids the problems mentioned above. The algebraic approach is applicable also in cases the beams are not nicely distributed as in SPECT (Single Positron Emission CT) and in Acoustic Tomography and Seismic Tomography.

Exercise 9.2. Let $f:\mathbb{R}^2\to\mathbb{R}$ be 1 on $\{(x,y)\,\big|\,x^2+y^2\leq 1\}$ and 0 elsewhere. The right picture in Fig. 34 show the graph of some of the functions in this exercise.

(a) Prove that $p_\phi(\xi)=2\sqrt{1-\xi^2}$ if $|\xi|\leq 1$ and $\phi(\xi)=0$ if $|\xi|\geq 0$.

(b) Let $\widetilde{f}(r\cos(\vartheta),r\sin(\vartheta))\equiv\int_0^\pi p_\phi(r\cos(\vartheta-\phi))\,\mathrm{d}\phi$ be the (unfiltered) back projection. Show that $\widetilde{f}(r\cos(\vartheta),r\sin(\vartheta))=\widetilde{f}(r,0)$. Show that $\widetilde{f}$ is strict positive and decreasing on $[0,\infty)$ and $\widetilde{f}(r,0)\sim\frac{2\pi}{r}$ for $r\to\infty$.

(c) Suppose that $\widetilde{p}_\phi$ is analytic in $(-1,1)$. Show that there is a $\gamma>0$ such that $\widetilde{p}_\phi(\xi)=\gamma$ for all ξ, $|\xi|<1$. Compute γ.

Exercise 9.3. Let $f:\mathbb{R}^2\to\mathbb{R}$ be 1 on $[-1,1]^2$ and 0 elsewhere.
Let $g_T:\mathbb{R}\to\mathbb{R}$ be defined by $g_T(\xi)\equiv T-|\xi|$ if $|\xi|\leq T$ and $g_T(\xi)=0$ elsewhere. The left picture in Fig. 34 show the graph of some of the functions in this exercise.

(a) Prove that, for each $\phi\in(0,\frac{1}{4}\pi)$, p_ϕ is a linear combination of $g_{T(\phi)}$ and $g_{T(-\phi)}$ with $T(\phi)\equiv\cos(\phi)+\sin(\phi)$.

(b) Show that $g_2=\Pi_1*\Pi_1$. Compute $\widehat{g_2}$. Show that $\rho\rightsquigarrow\widehat{g_2}(\rho)|\rho|$ is in $L^2(\mathbb{R})$.

(c) Put $h_\Omega(t)\equiv\mathrm{Re}\left(\int_0^\Omega\frac{\sin^2(\pi\omega)}{\omega}\,e^{2\pi it\omega}\,\mathrm{d}\omega\right)\quad(t\in\mathbb{R})$. Compute $h'_\Omega(t)$.

It can be show that, for all $\tau > 0$, $\int_\tau^\infty \frac{\cos(2\pi\Omega t)}{t}\,\mathrm{d}t \to 0$ for $\omega \to \infty$. Use this result to show that,

$$h_\infty(\tau) \equiv \lim_{\Omega\to\infty} h_\Omega(\tau) = \lim_{\Omega\to\infty} \int_\tau^\infty h'_\Omega(t)\,\mathrm{d}t = \frac{1}{4}\ln\left(\frac{\tau^2-1}{\tau^2}\right) \qquad (\tau \in \mathbb{R}).$$

Compute the inverse Fourier transform of $\rho \rightsquigarrow \widehat{g}_2(\rho)|\rho|$.

(d) For $\phi \in (0, \frac{1}{4}\pi)$, compute $\widetilde{p}_\phi$.

Index

References

[1] Juan Arias de Reyna. *Pointwise convergence of Fourier series*, volume 1785 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002.

[2] D. C. Champeney. *Fourier transforms and their physical applications*. Academic Press, London, New-York, 1973.

[3] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965.

[4] Paul L. DeVries. *A first course in computational Physics*. Wiley & sons, 1994.

[5] R. E. Edwards. *Fourier Series, a modern introduction*. Holt, Rinehart and Winston, Inc., New York, etc., 1967.

[6] A. C. Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[7] T. W. Körner. *Fourier analysis*. Cambridge University Press, Cambridge, 1988.

[8] A. Papoulis. *The Fourier Integral and its Applications*. McGraw-Hill Book Company, 1962.

[9] A. Papoulis. *Signal Analysis*. McGraw-Hill Book Company, 1977.

[10] J. Price, editor. *Fourier techniques and applications*, New York, 1985. Plenum Press.

[11] P. L. Walker. *The Theory of Fourier Series and Integrals*. Wiley, Chichester, 1986.

Computer session I: Convergence of Fourier series

I.A Introduction

In this exercise, $f : \mathcal{I} \to \mathbb{R}$ is bounded and integrable on the bounded interval $\mathcal{I} \equiv [a, b]$. We assume f to be periodically extended to $\mathbb{R}$ with period $T \equiv b - a$:

$$f(t) = f(t + kT) \qquad (k \in \mathbb{Z}, t \in \mathbb{R}).$$

For $k \in \mathbb{Z}$, $\gamma_k(f)$ is the kth Fourier coefficient of f, and $S_n(f)$ is the nth partial Fourier series $(n \in \mathbb{N}_0)$ as defined in §2.2. We are interested in the approximation quality of $S_n(f)$ (cf. §2.4).

We will also consider the best sup-norm approximation $P_n(f)$ of f of the form

$$P_n(f)(t) = \sum_{k=-n}^{n} \mu_k \, e^{2\pi i t \frac{k}{T}} \qquad (t \in \mathbb{R}),$$

with $\mu_k \in \mathbb{C}$. $P_n(f)$ is a so-called *trigonometric polynomial* of degree n.

We put

$$F_n(f) \equiv \|f - S_n(f)\|_\infty \quad \text{and} \quad E_n(f) \equiv \|f - P_n(f)\|_\infty \qquad (n \in \mathbb{N}_0).$$

For some $N \in \mathbb{N}$, we will also consider the *discrete Fourier coefficients* $\widetilde{\gamma}_k(f)$,[1] for $|k| \le N$, and the nth partial discrete Fourier series $\widetilde{S}_n(f)$, for $n \in \mathbb{N}, n \le N$. These quantities are given by (cf. §5)

$$\widetilde{\gamma}_k(f) \equiv \frac{1}{N} \sum_{n=0}^{N-1} f(t_n) \, e^{-2\pi i \frac{nk}{N}} \quad \text{and} \quad \widetilde{S}_n(f)(t_m) \equiv \sum_{k=-n}^{n} \widetilde{\gamma}_k(f) \, e^{2\pi i t_m \frac{k}{T}}. \qquad (1)$$

Notation. For $\mu, \nu : \mathbb{Z} \to \mathbb{R}$ we write

$$\mu \sim \nu \quad \text{if} \quad \lim_{|k|\to\infty} \frac{\mu(k)}{\nu(k)} \quad \text{exists and is non-zero:}$$

then (μ_k) and (ν_k) have the same asymptotic behaviour for $|k| \to \infty$.

In this exercise, we study the relation between the 'speed' by which the sequence $(\gamma_k(f))$ of Fourier coefficients approaches 0 if $|k| \to \infty$ and the smoothness of f. For instance, we will try to find out when

$|\gamma_k(f)| \sim k^a$ for some $a < 0$,
$|\gamma_k(f)| \sim \lambda^k$ for some $\lambda \in [0, 1)$, or
$|\gamma_k(f)| \sim \frac{1}{k!}$.

For this purpose, we try to find a function $\mu : \mathbb{C} \times \mathbb{Z} \to \mathbb{R}$ such that $\lim_{|k|\to\infty} \mu(\gamma_k(f), k)$ exists and is non-zero: if, for instance, $|\gamma_k(f)| \sim k^{-2}$, then the choice $\mu(\gamma, k) = k^2\gamma$, implies that $\mu(\gamma_k(f), k) \sim 1$, with $\mu(\gamma, k) = \log(\gamma)/k$ we have that $\mu(\gamma_k(f), k) \sim 1$ if $\gamma_k(f) \sim \lambda^k$. We employ scaling functions μ to make the asymptotic behaviour of $(\gamma_k(f))$ better visible in graphical displays.

[1] Actually the **Matlab** code that we use does not compute the Fourier coefficients $\gamma_k(f)$ but a discrete variant: it computes the discrete Fourier coefficients for $N = 2^l > 8n + 100$, where n is the largest value of $|k|$ that is of interest to us. These values can be computed with the so-called Fast Fourier Transform (FFT) technique. The technique is fast if N is a power of 2. The program uses the defining formulas if N is not a power of 2.

I.B Exercises

The **Matlab** *program* `FSstart.m` *can be used for all exercises in this session. This program has to be edited to change parameters. To run* `FSstart.m`*, type* `FSstart` *in* **Matlab***'s execution screen. Instructions that are program-dependent are in footnotes in order to keep the main text of the exercises independent of a specific program. Since* **Matlab** *does not accept Greek letters as input, we spelled out the names of the Greek letters: for instance, we used* `mu` *for* μ *and* `gamma` *for* γ*. You may have to hit the 'space' bar once in a while to get the next picture; the* `pause` *statement has been included in the* **Matlab** *programs between pictures. Further references in this text to the program* `FSstart.m` *will be in footnotes.*

Make notes: insights and results will be used in subsequential exercises.

Computer-exercise I.1. Take for f the function $f(t) \equiv |t|$ for $t \in \mathcal{I} \equiv [-2, +2]$. Plot the Fourier coefficients.[2] The Fourier coefficients appear to be real and symmetric ($k \rightsquigarrow \gamma_k(f)$ is even). Could this be anticipated? Find a function μ such that $\mu_k \equiv \mu(\gamma_k(f), k)$ is more or less constant for large k.[3] The speed with which the Fourier coefficients of the function $f(t) = |t - 1|$ on $[-2, +2]$ converge to 0 is slower (the absolute value of the Fourier coefficients are larger for large $|k|$).[4] Why is the asymptotic behaviour for $|k| \to \infty$ of the Fourier coefficients of $|t|$ and of $|t-1|$ on $[-2, +2]$ so different? (A heuristic argument suffices here; a detailed explanation will be derived in the exercises below.) The Fourier coefficients of $f = |t - 1|$ on $[-1, 3]$ appear to be purely imaginary, but their asymptotic behaviour is the same as for $f = |t|$ on $[-2, +2]$.[5] Can you explain this?

Computer-exercise I.2. In this exercise, $\mathcal{I} = [-2, +2]$. The functions are defined for $t \in \mathcal{I}$.

Take for f the function $f(t) = |t|$. Design a function μ such that $\mu_k = \mu(\gamma_k(f), k)$ is more or less constant for large k, k odd. Design also a function μ such that $\mu(F_k(f), k)$ is more or less constant for larger k.[6]

Investigate the asymptotic behaviour of $(\gamma_k(f))$ and of $(F_k(f))$ for the function f defined by $f(t) = |\sin(\frac{1}{4}\pi t)|$.[7] Does the asymptotic behaviour depend on the number of singularities in one period?

Find the asymptotic behaviour of $(\gamma_k(f))$ and of $(F_k(f))$ for the function $f(t) = \sqrt{|\sin(\frac{1}{4}\pi t)|}$.[8] What do you expect for the speed by which $(\gamma_k(f))$ an $(F_k(f))$ converge towards 0 in case $f(t) = |\sin(\frac{1}{4}\pi t)|^a$ for some $a > 0$? Check your conjectures for several $a > 0$. Why are $a \in \mathbb{N}_0$, a even, exceptional values?

[2] Check whether the values in the executable lines of this program are correct: `f='abs(t)'`, `I=[-2,2]`, and, for the moment, take `ok=0`, `delta=0`, `mu='1*gamma'`, and `k=-50:50`. 'Run' the program. First, the graph of f will be displayed. Then, if you hit the space bar, the graph of the Fourier coefficients will be plotted in the next figure window. To be more precise, you will see the graph of $k \rightsquigarrow \mu_k \equiv \mu(\gamma_k(f), k)$, with $\mu(\gamma, k) = \gamma$, for $k \in \mathbb{Z}$, $|k| \leq 50$: the line `mu='1*gamma'` sets the choice for μ and the line `k=-50:50` defines the collection of ks for which μ_k will be shown.

[3] Try a μ of the form $\mu(\gamma, k) = \gamma\, k^i$ for $i = 1, 2, \ldots$: `mu='gamma.*(k.∧1)'`, `mu='gamma.*(k.∧2)'`, ...

[4] Take `f='abs(t-1)'`. Try another i.

[5] `I=[-1,3]`.

[6] Now take `ok=1;`. As with `ok=0`, the graph f and of $k \rightsquigarrow \mu(\gamma_k(f), k)$ will be displayed. Another hit of the space bar will show the graph $k \rightsquigarrow \mu(F_k(f), k)$ in the same figure as $k \rightsquigarrow \mu(\gamma_k(f), k)$. Adjust `mu` to capture the correct asymptotic behaviour.

[7] `f='abs(sin(pi*t/4))'`.

[8] Take `f='abs(sin(pi*t/4)).∧0.5';`.

Consider the functions $f = |\sin(\cos(\frac{1}{2}\pi t))|$ and $f = \sin(\cos(\frac{1}{2}\pi t))$ as well: if f' has a singularity, then the asymptotic behaviour of the sequences $(\gamma_k(f))$ and $(F_k(f))$ seems to depend only on the type of the singularity and not on the specific values of f away from the singular point (can you explain this? We will return to this issue in Computer-exercise I.6 below).

Can you predict the asymptotic behaviour of $(\gamma_k(f))$ and $(F_k(f))$ for $f(t) = |t|^{\frac{1}{2}}$ and for $f(t) = |t|^{\frac{3}{2}}$?

Computer-exercise I.3. Consider the convergence behaviour of $(\gamma_k(f))$ and $(F_k(f))$ for $f(t) \equiv \exp(\cos(t)) \quad (t \in [-\pi, \pi])$.[9] The function f is smooth and fast convergence could have been anticipated. Why? However, the fast convergence for the $(\gamma_k(f))$ seems to get lost for $k \geq 13$. Can you explain the behaviour of $(\gamma_k(f))$ and of $(F_k(f))$ for $k \geq 13$?

From the graphs of $k \rightsquigarrow \mu(\gamma_k(f), k)$ and $k \rightsquigarrow \mu(F_k(f), k)$ it appears that, for $k \leq 12$, $|F_k(f)| \leq 2\,|\gamma_k(f)|$ and $|F_k(f)| \approx 2\,|\gamma_{k+1}(f)|$. Can you explain this?

Computer-exercise I.4. For compression purposes, one may consider to approximate f by $P_n(f)$ rather than by $S_n(f)$. In this exercise, we investigate whether this is an attractive idea.

Consider $f = \exp(\cos(t))$ for $t \in [-\pi, \pi]$. Compare the approximation quality of $S_n(f)$ and $P_n(f)$ for $n = 10$.[10] Is it worthwhile to use $P_n(f)$ instead of $S_n(f)$?

Investigate the same questions for $f(t) = |t| \quad (t \in [-2, 2])$. How much can we gain with $P_n(f)$ for, say, $n = 10$? What about $f(t) = |t-1|$ on $[-2, 2]$? What is your conclusion?

Computer-exercise I.5. We investigate the effect of random perturbations on f in this exercise: we compute the partial Fourier series of $f^\star \equiv f + \varepsilon$, where, for some $\delta > 0$, the "white" noise $\varepsilon : \mathcal{I} \to \mathbb{R}$ takes values that are uniformly randomly distributed between $-\delta$ and δ.

First take $f(t) = 0 \quad (t \in \mathcal{I} = [-1, +1])$ and $\delta = 1$.[11]

Consider the Fourier coefficients $\gamma_k(f^\star)$ and $F_k^\star(f) \equiv \|f - S_k(f^\star)\|_\infty$.[12] (These Fourier coefficients have not been formally defined. Why not? However, we use discrete Fourier coefficients in our program $\widetilde{\gamma}_k(f^\star)$ for $M = 2^l > 8N+200$; see (1).). The $\gamma_k(f^\star)$ seem to be randomly distributed between $[-\gamma, \gamma]$ for some $\gamma > 0$ (with $\gamma \approx \frac{\delta}{2\sqrt{M}}$?).

It can be show that $F_k^\star(f) \leq \lambda_k \delta$ with λ_k the so-called Lebesgue's constant: $\lambda_k \approx 0.1 + \frac{4}{\pi^2}\log(2|k|+1)$. Can you see this upper bound in the graph of $k \rightsquigarrow F_k^\star(f)$? The upper bound for $F_k^\star(f)$ by a function proportional to $\log(2|k|+1)$ appears to be an over estimate for smaller values of k ($k \leq \sqrt{M}$?): a linear growth seems to be more realistic here. Can you derive such a linear estimate?

Now, take $f(t) = \exp(\cos(\pi t)) \quad (t \in [-1, +1])$ and $\delta = 0.01$.
Consider the Fourier coefficients $\gamma_k(f^\star)$ and $F_k^\star(f)$.[13] Initially $k \rightsquigarrow F_k^\star(f)$ decreases, but for larger k it slowly increases. For what value of k do you get the best approximation of f? Since the sequence $(\gamma_k(f))$ converges rapidly to 0, the approximation error decreases

[9] `ok=1`, `f='exp(cos(t))'`, `I=[-pi,pi]`, and `k=-30:30`. Take `mu='log10(abs(gamma))'` or another suitable function.

[10] Take `ok=3` to get a graph of $P_n(f)$ and $f - P_n(f)$. The value of n will be equal to the maximum absolute value of the values appearing in the sequence `k`.

[11] `f='0.*t'` and `delta=1.0`.

[12] Take `ok=1`, `mu='gamma'`, and `k=0:100`.

[13] Take `mu='log10(abs(gamma))'`, etc..

rapidly if k increases. The effects of the noise increase with increasing k. Therefore, the total error $F_k^\star(f)$ takes a minimal value for some $k = k_0$. From the figure, it appears that, by working with some $S_{k_0}(f^\star)$, the effect of the noise can be reduced by, at least, a factor $\approx \frac{k_0}{\sqrt{M}}$. Such a reduction of the noise can be proved for smooth functions f.

(This property is exploited in practice: if, from a smooth function f, only N function values are known from measurements, then the effect of the errors in the measurements can be diminished by working only with an nth partial Fourier series —or, to be more precise, an nth discrete partial Fourier series— with, typically, $n \leq \sqrt{N}$: the complete (discrete) Fourier series is not used.)

For a function $f : \mathcal{I} \to \mathbb{R}$ en $t \in \mathcal{I}$ we use the notation $f(t+)$ if the right-limit $\lim_{\varepsilon>0,\varepsilon\to 0} f(t+\varepsilon)$ of f at t exists. Then, $f(t+)$ is the value of this limit. Similarly, with $f(t-)$ we refer to the left limit.

Computer-exercise I.6. In this exercise, we investigate how accurately the Fourier series converge in the neighborhood of discontinuities of f. (The insights that we gain here are also helpful to explain the convergence behaviour of Fourier series close to discontinuities of f', f'': differentiation brings the discontinuities in lower derivatives.)

Take for f on $\mathcal{I} = [-2, 2]$ the top-hat function Π_1 defined by

$$\Pi_1(t) = \begin{cases} 1 & \text{voor} \quad |t| < 1 \\ 0 & \text{voor} \quad |t| \geq 1. \end{cases}$$

The sequence $(\gamma_k(f))$ converges to 0 (why?). Check graphically that $|\gamma_{2k+1}(f)| \sim \frac{1}{2k+1}$ and $F_n(f) \sim 1$ (why is $F_k(f) \geq \frac{1}{2}$ for all k?).[14]

Pay careful attention to meaning of the various notions of convergence and conclude from the graphs that [15]

(i) $(S_k(f))$ converges point-wise to $\frac{1}{2}[f(\cdot+)+f(\cdot-)]$ (note that this function coincides with f itself in each $t \in [-2,2]$, $|t| \neq 1$),

(ii) for each $\varepsilon > 0$, $(S_k(f))$ converges uniformly to f on $\{t \in [-2,2] \,\big|\, ||t|-1| \geq \varepsilon\}$,

(iii) the sequence $(\mathcal{G}_n)$, with $\mathcal{G}_n \equiv \{(t, S_n(f)(t)) \,\big|\, t \in \mathcal{I}\}$ the graph of $S_n(f)$, does not converge to the subset $\mathcal{G}$ of $[-2,2] \times \mathbb{R}$ in the left picture in Fig. 36, but it does converge to the subset $\mathcal{G}$ in the right picture (here $(t, y) \in \mathbb{R}^2$ is a limit point of $(\mathcal{G}_n)$ if each disk with center (t, y) in $[-2,2] \times \mathbb{R}$ intersects the graph of $\mathcal{G}_n$ for all sufficiently large n. Why is this type of convergence not in conflict with the one in (i)?).

Comment to (iii). The graphs $\mathcal{G}_n$ exhibit more wiggles with increasing n. For each n, the tops and the bottoms of these wiggles appear to have exactly the same heights; the wiggles are only compressed in the t-direction towards the points of discontinuity if n increases. There is no change in the vertical direction. This effect is known as the *Gibb's phenomenon.*

Do the same for $f = g\Pi_1$, where g is a smooth 4-periodic function with $g(1) \neq 0$.[16] Note that in the neighbourhood of the discontinuity ($t = 1$) only the scaling of the

[14] Take `ok=0;`, `delta=0;`, `f='(abs(t)<1)'`, and `I=[-2,2]`.

[15] Take `ok=2;` and subsequentially for `k` `k=0:10;`, `k=0:20;`, `k=0:40;`, `k=0:80;`. Pay special attention to the height of the first top of $S_k(f)$ in the neighbourhood of the discontinuity of f.

[16] For instance, `f='exp(-cos(4*pi*t)).*(abs(t<1)'`.

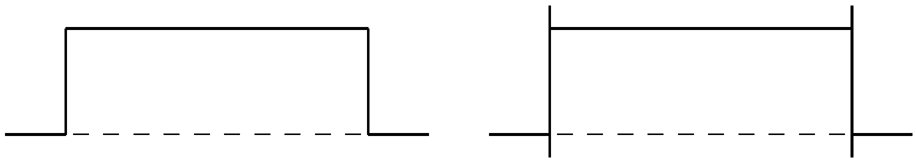


FIGURE 36. Limit functions?

phenomenon changes (with a factor $g(1)$). Can you explain this? (Hint: $f = h + \kappa\Pi_1$ for some $\kappa \in \mathbb{R}$ and a function h that is smoother than f. Note that, now, differentiation leads to an explanation of the phenomenon as observed in the last part of Computer-exercise I.2)

Computer-exercise I.7. Let $\widetilde{\gamma}_k(f)$ be the kth discrete Fourier coefficient of f $(|k| \leq N)$. For each $k \in \mathbb{Z}$, $|k| < N$, we have that

$$\widetilde{\gamma}_k(f) = \gamma_k(f) + \sum_{j\in\mathbb{Z}, j\neq 0} \gamma_{k+jN}(f),$$

assuming f is sufficiently smooth: if the Fourier coefficients of f rapidly converge to 0, then the $\widetilde{\gamma}_k(f)$ form a good approximation to $\gamma_k(f)$ (for $|k| \ll N$).

Check this claim experimentally and observe that the approximation is also good if k is not much smaller than $\frac{1}{2}N$.[17] Check the quality of the approximation also in case the Fourier coefficients decrease slowly.[18]

Compare the efficiency of FFT and the naive approach to evaluate the discrete Fourier transform.[19] How is the time needed to computed the Fourier coefficients related to the number of coefficients in case of FFT and in case of the naive approach?

Computer-exercise I.8.

For $a > 0$, the function $f \in C(\mathcal{I})$ is defined by

$$f(t) := \begin{cases} |t|^a \cos\frac{\pi}{|t|} & \text{for} \quad t \in [-1, +1], t \neq 0 \\ 0 & \text{als} \quad t = 0. \end{cases}$$

The Fourier series converges uniformly if $a > 1$. Why? There is no theorem on uniform convergence that can be applied in case $a \in (0, 1)$. Check graphically whether the Fourier series converges for, for instance, $a = 1$ and $a = 0.2$.[20] The sequence $(\gamma_k(f))$ of Fourier coefficients should converge for each choice of $a \in (0, 1)$. Why? Can you observe this in the graphs?

Consider the case where the Fourier series does not seem to converge uniformly. Display the graph of $f - S_k(f)$ for several values of k and observe that the Fourier series

[17] Take `ok=4`, for f, say, `f='1./(25*cos(pi*t).∧2+1)'`, `mu='log10(abs(gamma))'`, and `k=-100:100`. If you run the program, the graphs of $k \rightsquigarrow \mu(k, \gamma_k(f))$ and, after hitting the space bar, the graphs of $k \rightsquigarrow \mu(k, \widetilde{\gamma}_k(f))$ and of $k \rightsquigarrow \mu(k, \gamma_k(f) - \widetilde{\gamma}_k(f))$ will be displayed. The discrete Fourier coefficients are computed for $N = n$, where n is the maximum absolute value in the sequence `k`.

[18] Take, for instance, `f='(1-t.*t).*cos(10*pi*t).∧2'`, etc..

[19] The program does not compute the exact Fourier coefficients, but for $\gamma_k(f)$, it takes discrete Fourier coefficients for $N = 2^\ell > 8n + 100$. These coefficients are evaluated with FFT. The timings are shown in the **Matlab** window. Take for `k=-1000:1000`, `k=-2000:2000`, ...

[20] take `ok=1`, `k=0:100`, `mu='log10(abs(gamma))'` and, for instance, `f='(abs(t)+10∧(-12)).∧0.2.* cos(pi./(abs(t)+10∧(-12)))'` (to avoid "overflow" by division by 0, we suggest to add 10^{-12} to `abs(t)`).

converges at t, for each t ($\lim_{n\to\infty} S_n(f)(t) = f(t)$) and even uniformly on $[-1,-\varepsilon]\cup[\varepsilon,1]$,[21] (cf., Computer-exercise I.6). In what sense does the 'uniform convergence' fail?

I.C Best approximations

I.1 Trigonometric polynomials. A function p on $\mathbb{R}$ is a *trigonometric polynomial* of degree $\leq n$ with period T, if, for some scalars α_k, we have that

$$p(t) = \sum_{k=-n}^{n} \alpha_k\, e^{2\pi i t \frac{k}{T}} = \sum_{k=-n}^{n} \alpha_k\, \zeta^k \quad \text{where } \zeta \equiv \zeta_t \equiv e^{2\pi i \frac{t}{T}} \qquad (t\in\mathbb{R}).$$

Let $\mathcal{T}_T^n$ be the space of all T-periodic trigonometric polynomials of degree $\leq n$.

Note that $S_n(f)\in\mathcal{T}_T^n$ for each $f\in L_T^1(\mathbb{R})$.

I.2 The best trigonometric approximation. Consider a real-valued T-periodic function $f\in C(\mathbb{R})$. We call p the *best approximation* of f in $\mathcal{T}_T^n$ with respect to the sup-norm if $p\in\mathcal{T}_T^n$ and $\|f-p\|_\infty \leq \|f-q\|_\infty$ for all $q\in\mathcal{T}_T^n$. Without proof we mention that the best approximation exists and is unique. This best approximation is denoted by $P_n(f)$. We put $E_n(f)\equiv\|f-P_n(f)\|_\infty$: $E_n(f)$ is the sup-norm *distance* from f to $\mathcal{T}_T^n$.

If f is real-valued, then the best approximation in $\mathcal{T}_T^n$ is real-valued as well.

In order to achieve better compression, one may consider to approximate a T- periodic function f by its best approximation $P_n(f)$ in $\mathcal{T}_T^n$: instead of storing a collection of sampled function values $f(n\Delta t)$, one can store the largest coefficients α_k of $P_n(f)$. Standardly, a partial Fourier series of f is used. This polynomial can relatively easily by computed using FFT. Sup-norm best approximations are hard to compute. But, one may hope that $\|f-P_n(f)\|_\infty \leq \|f-S_m(f)\|_\infty$ for some n that is (much) smaller than m.

One can show the following. We leave the proof as an exercise to the reader (see Exercise I.3).

I.3 Theorem

$$\|f-P_n(f)\|_\infty \leq \|f-S_n(f)\|_\infty \leq \|f-P_n(f)\|_\infty(1+T\lambda_n),$$

where λ_n is Lebesgue's constant:

$$\lambda_n \equiv \int_0^1 \frac{\sin\pi(2n+1)t}{\sin\pi t}\,\mathrm{d}t \approx 0.1 + \frac{4}{\pi^2}\log(2|n|+1). \qquad \square$$

For $n\to\infty$, Lebesgue's constant tends to ∞. However, the convergence is very slow and, for moderate values of n, Lebesgue constant is rather small. Nevertheless, the theorem is rather pessimistic in case of rapid convergence of the Fourier series. Then the following theorem can be useful.

[21] Take `ok=2` and subsequentially, for instance, `k=0:10`, `k=0:20`, `k=0:40`, `k=0:80`.

I.4 The Theorem of de la Vallée–Poussin.
Let f be a real-valued function in $L^1_T(\mathbb{R})$. Consider some real-valued $p \in \mathcal{T}^n_T$. If there is an increasing sequence $t_0, \ldots, t_{2n+1}$ such that $t_{2n+1} - t_0 < T$ and

$$G_k \equiv (-1)^k[f(t_k) - p(t_k)] > 0 \quad \textit{for} \quad k = 0, \ldots, 2n+1,$$

then

$$\min_k G_k \leq E_n(f) \leq \|f - p\|_\infty.$$

Proof. The last equality is trivially true by definition of $E_n(f)$.

Now, assume that $E_n(f) < G_k$ for all $k = 0, \ldots, 2n+1$.
Then, for all $k = 0, \ldots, 2n+1$, we have that

$$(-1)^k(P_n(f) - p)(t_k) = (-1)^k(f - p)(t_k) - (-1)^k(P_n(f) - f)(t_k) \geq G_k - E_n(f) > 0.$$

Put $q \equiv P_n(f) - p$. q is real-valued and in $\mathcal{T}^n_T$. If we put $t_{2n+2} \equiv t_0 + T$, then, by T-periodicity of q, we have that $(-1)^k q(t_k) > 0$ for all $k = 0, \ldots, 2n+2$. Since q is real-valued and continuous, q has at least one zero between each pair of consecutive t_k's. In other words, there are $2n+2$ different zeros $s_0, \ldots, s_{2n+1}$ of q in (t_0, t_0+T). q is of the form $Q(\zeta_t)$ with $Q(\zeta) \equiv \sum_{k=-n}^n \beta_k \zeta^k$ $(\zeta \in \mathbb{C})$ and $\zeta_t \equiv \exp(2\pi i t/T)$. Hence, Q as well as the polynomial $\zeta^n Q$ of degree at most $2n$ has $2n+2$ zeros at the ζ_{s_k}. A corollary of the main theorem of the algebra states that non-trivial polynomials of degree N can have at most N zeros in the complex plain. Consequently, Q and, therefore, q must be zero. Apparently $P_n(f) = p$. But this violates the assumption that $E_n(f) < G_k$ for all k. We can conclude that $E_n(f) \geq G_k$ for some k. □

Note that the number of points t_k that are required by the above theorem is one more than the number of terms involved in the trigonometric polynomial p.

In our applications, $S_n(f)$ will be the approximating trigonometric polynomial p. If the $\gamma_k(f)$ converge quickly towards 0 if $|k| \to \infty$, then we will have that

$$f - S_n(f) \approx 2\,\mathrm{Re}(\gamma_{n+1}(f)\, e^{2\pi i \frac{(n+1)t}{T}}) = 2\,|\gamma_{n+1}(f)|\, \cos(2\pi \left(\frac{(n+1)t}{T} + \phi\right)),$$

where ϕ is such that $\gamma_{n+1}(f) = |\gamma_{n+1}(f)|\, e^{2\pi i \phi}$. Since $\cos(k\pi) = (-1)^k$, we see that $f(t_k) - S_n(f)(t_k) \approx (-1)^k |\gamma_k(f))|$ at $t_k = (\frac{1}{2}k - \phi)\frac{T}{n+1}$ for $k = 0, \ldots, 2n+1$. Since we will also have that $\|f - S_n(f)\|_\infty \approx 2\,|\gamma_{n+1}(f)|$, it follows that in this situation of quickly decreasing $|\gamma_k(f)|$ the best approximation $P_n(f)$ will not provide a much better approximation than $S_n(f)$ (see Exercise I.4).

As a corollary we have that $p \in \mathcal{T}^n_T$ is the best approximation of f if there is a increasing sequence $t_0, \ldots, t_{2n+1}$ in $[t_0, t_0+T)$ such that $(-1)^k[f(t_k) - p(t_k)] = \|f - p\|_\infty$ for all $k = 0, \ldots, 2n+1$; we say that $f - p$ has the *alternation property* at $2n+2$ points. The converse is also correct if f is continuous. For a proof we refer to the literature.

I.5 Alternation theorem. *Let f be a real-valued continuous function in $L^1_T(\mathbb{R})$, and let $p \in \mathcal{T}^n_T$. Then, p is the best sup-norm approximation of f in $\mathcal{T}^n_T$ if and only if there is an increasing sequence $t_0, \ldots, t_{2n+1}$ in $[t_0, t_0 + T)$ such that*

$$(-1)^k[f(t_k) - p(t_k)] = \|f - p\|_\infty \quad \textit{for all } k = 0, \ldots, 2n+1. \quad \Box$$

Exercises

Exercise I.1. Consider a T-periodic function f. Suppose that p is the best approximation of f in $\mathcal{T}_T^n$.
(a) If $q \in \mathcal{T}_T^n$, then $\frac{1}{2}(q+\overline{q}) \in \mathcal{T}_T^n$.
(b) If f is real-valued then p is real valued. Prove this.
(Hint: If $a \in \mathbb{R}$ and $b \in \mathbb{C}$, then $|a - \mathrm{Real}(b)| \leq |a-b|$.)

Exercise I.2.
(a) Show that

$$S_n(f)(s) = \sum_{|k|\leq n} \gamma_k(f)\, e^{2\pi i s \frac{k}{T}} = \int_0^T f(t) \sum_{|k|\leq n} e^{2\pi i (s-t)\frac{k}{T}} \,\mathrm{d}t.$$

Apparently, with $D_n(t) \equiv \sum_{|k|\leq n} \exp(2\pi i t k)$, we have that $S_n(f)(s) = \int_0^T f(t) D_n(\frac{s-t}{T})\,\mathrm{d}t$. D_n is the so-called *Dirichlet kernel*. Note that D_n is real-valued, even and 1-periodic.
(b) Prove that

$$\|S_n(f)\|_\infty \leq \|f\|_\infty \, T\lambda_n \quad \text{where} \quad \lambda_n \equiv \frac{1}{T}\int_0^T |D_n(\frac{t}{T})|\,\mathrm{d}t = 2\int_0^{\frac{1}{2}} |D_n(t)|\,\mathrm{d}t.$$

λ_k is *Lebesgue's constant.*
(c) Show that, with $\zeta \equiv \exp(2\pi t)$, we have that

$$D_n(t) = \sum_{|k|\leq n} \zeta^k = \zeta^{-n}\sum_{k=0}^{2n} \zeta^k = \zeta^{-n}\frac{\zeta^{2n+1}-1}{\zeta - 1} = \frac{\zeta^{n+\frac{1}{2}} - \zeta^{-n-\frac{1}{2}}}{\zeta^{\frac{1}{2}} - \zeta^{-\frac{1}{2}}} = \frac{\sin(\pi(2n+1)t)}{\sin(\pi t)}.$$

(d) Show that, $|D_n(t)| \leq \pi(n+\frac{1}{2})$ if $t \in [0, \frac{1}{2n+1}]$, and $|D_n(t)| \leq \frac{1}{2t}$ if $t \in [0, \frac{1}{2}]$. (Hint: $\sin(\pi t) \geq 2t$ for $|t| \leq \frac{1}{2}$). Now, show that

$$\lambda_n \leq \pi + \log(n + \tfrac{1}{2}).$$

The sharper upper bound in Theorem I.3 requires more careful estimates. However, note that this simple derivation already shows the logarithmical dependence on n.

Exercise I.3. Consider a $f \in L_T^1(\mathbb{R})$.
(a) Prove that $\|f - P_n(f)\|_\infty \leq \|f - S_n(f)\|_\infty$.
(b) Prove that $S_n(p) = p$ for all $p \in \mathcal{T}_T^n$.
(c) Prove that $\|f - S_n(f)\|_\infty \leq \|f-p\|_\infty + \|S_n(f-p)\|_\infty$ for all $p \in \mathcal{T}_T^n$.
(d) Prove that $\|f - S_n(f)\|_\infty \leq \|f - P_n(f)\|_\infty(1 + T\lambda_n)$, where λ_n is as defined in Exercise I.2.

Exercise I.4. Let f be real-valued and in $L_T^1(\mathbb{R})$. Suppose there is some $\vartheta \in [0,1)$ such that

$$2(1-\vartheta)\,|\gamma_{n+1}(f)| \leq \sum_{|k|>n} |\gamma_k(f)| \leq 2(1+\vartheta)\,|\gamma_{n+1}(f)|.$$

(a) Prove that $2(1-\vartheta)\,|\gamma_{n+1}(f)| \leq \|f - S_n(f)\|_\infty \leq 2(1+\vartheta)\,|\gamma_{n+1}(f)|$.
(b) Show that there is an increasing sequence $t_0, \ldots, t_{2n+1}$ in $[t_0, t_0+T)$ such that $(-1)^k[f(t_k) - S_n(f)(t_k)] \geq 2\,(1-\vartheta)|\gamma_{n+1}(f)|$ for all $k = 0, \ldots, 2n+1$.
(c) Prove that $E_n(f) \leq \|f - S_n(f)\|_\infty \leq \frac{1+\vartheta}{1-\vartheta} E_n(f)$.

Exercise I.5. Use de la Vallée–Poussin's theorem to prove that a $p \in \mathcal{T}_T^n$ is the best sup-norm approximation of a real-valued function $f \in L_T^1(\mathbb{R})$ if the 'error' function $f - p$ has the alternation property at at least $2n+2$ points (that is, prove the 'if' part of the alternation theorem).