MATHEMATICAL TECHNIQUES FOR IMAGE ANALYSIS

2DMM10 COURSE NOTES, 5TH REVISED EDITION

Luc Florack

MATHEMATICAL TECHNIQUES FOR IMAGE ANALYSIS

Code: 2DMM10

Author: Luc Florack

Version: December 14, 2020

CONTENTS

PRELIMINARIES

Abbreviations

a.e.	almost everywhere
h.o.t.	higher order term(s)
iff	if and only if
l.h.s.	left hand side
r.h.s.	right hand side
s.t.	such that

Number Fields

- $\mathbb{N} = \{1, 2, 3, \ldots\}$: The set of natural numbers. If we wish to include 0 we write $\mathbb{N}_0$.
- $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$: The set of integer numbers.
- $\mathbb{Q} = \{p/q \mid p, q \in \mathbb{Z},\ q \neq 0\}$: The set of rational numbers.
- $\mathbb{R}$: The set of real numbers.
- $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$: The set of complex numbers: $i^2 = -1$.
- $\mathbb{K}$: Shorthand for either $\mathbb{R}$ or $\mathbb{C}$, depending on context.
- $\mathbb{D} = \{a + b\epsilon \mid a, b \in \mathbb{K}\}$: The set of dual numbers: $\epsilon^2 = 0$.
- $\mathbb{H} = \{a + bi + cj + dk \mid a, b, c, d \in \mathbb{R}\}$: The set of quaternions: $i^2 = j^2 = k^2 = ijk = -1$.

Function Classes

We abbreviate $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\int_\Omega f(x)dx = \int_\Omega f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$ for $\Omega \subset \mathbb{R}^n$. The essential supremum $\operatorname{ess\,sup}_{x\in\Omega} f(x)$ of a $\mathbb{R}$-valued function f is the smallest $M \geq 0$ s.t. $f(x) \leq M$ a.e. The

essential infimum ess $\inf_{x\in\Omega} f(x)$ of a $\mathbb{R}$-valued function f is the largest $m \geq 0$ s.t. $f(x) \geq m$ a.e. Unless explicitly stated otherwise, or if the context implies otherwise, functions are assumed to be $\mathbb{K}$-valued.

- $C(\Omega)$: The set of continuous functions with domain Ω.
- $C^k(\Omega)$, $k \in \mathbb{N}_0$: The set of k-fold continuously differentiable functions on Ω, with $C^0(\Omega) = C(\Omega)$.
- $C^\infty(\Omega)$: The set of smooth, i.e. infinitely differentiable functions with domain Ω.
- $C^\omega(\Omega)$: The set of analytical functions with domain Ω.
- $L^p(\Omega)$, $p \geq 1$: The set of functions for which $\int_\Omega |f(x)|^p \, dx$ exists. Notation: $\|f\|_p = \left(\int_\Omega |f(x)|^p \, dx\right)^{1/p}$.
- $L^\infty(\Omega)$: The set of functions f for which ess $\sup_{x\in\Omega} |f(x)|$ exists. Notation: $\|f\|_\infty = \text{ess}\sup_{x\in\Omega} |f(x)|$.
- $\mathscr{S}(\mathbb{R}^n)$: The set of smooth functions of rapid decay on $\mathbb{R}^n$.
- $\mathscr{S}'(\mathbb{R}^n)$: The topological dual of $\mathscr{S}(\mathbb{R}^n)$, a.k.a. the set of tempered distributions on $\mathbb{R}^n$.
- $\mathscr{D}(\Omega)$: The set of smooth bump functions with compact support in $\Omega \subset \mathbb{R}^n$.
- $\mathscr{D}'(\Omega)$: The topological dual of $\mathscr{D}(\Omega)$, a.k.a. the set of distributions on $\Omega \subset \mathbb{R}^n$.

Standard Functions

Some frequently used standard $C^\omega(\mathbb{R})$-functions:

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i}, \quad \cos x = \frac{e^{ix} + e^{-ix}}{2}, \quad \sinh x = \frac{e^{x} - e^{-x}}{2}, \quad \cosh x = \frac{e^{x} + e^{-x}}{2} \qquad (x \in \mathbb{R}).$$

Useful Inequalities

Hölder Inequality. Let $1 \leq p, q \leq \infty$ such that $1/p + 1/q = 1$, $f \in L^p(\Omega)$, $g \in L^q(\Omega)$, then $\|fg\|_1 \leq \|f\|_p \, \|g\|_q$.

The Hölder Inequality for the special case $p = q = 2$ is known as the *Schwartz Inequality*.

Young Inequality. Let $1 \leq p, q, r \leq \infty$ with $1/p + 1/q = 1 + 1/r$, $f \in L^p(\Omega)$, $\phi \in L^q(\Omega)$, then $\|f * \phi\|_r \leq \|f\|_p \|\phi\|_q$.

See e.g. Wheeden and Zygmund [31] for proofs.

Common Coordinate Reparametrisations

Polar coordinates for $(x, y) \in \mathbb{R}^2$, with $(r, \phi) \in \mathbb{R}^+ \times [0, 2\pi)$:

$$\begin{cases} x(r,\phi) &= r\cos\phi\,, \\ y(r,\phi) &= r\sin\phi\,. \end{cases}$$

Jacobian determinant:

$$\det \frac{\partial(x,y)}{\partial(r,\phi)} = r\,.$$

Cylindrical coordinates for $(x,y,z) \in \mathbb{R}^3$, with $(r,\phi,\zeta) \in \mathbb{R}^+ \times [0,2\pi) \times \mathbb{R}$:

$$\left\{ \begin{array}{rcl} x(r,\phi,\zeta) & = & r\cos\phi\,, \\ y(r,\phi,\zeta) & = & r\sin\phi\,, \\ z(r,\phi,\zeta) & = & \zeta\,. \end{array} \right.$$

Jacobian determinant:

$$\det \frac{\partial(x,y,z)}{\partial(r,\phi,\zeta)} = r\,.$$

Spherical coordinates (ISO convention) for $(x,y,z) \in \mathbb{R}^3$, with $(r,\theta,\phi) \in \mathbb{R}^+ \times [0,\pi] \times [0,2\pi)$:

$$\left\{ \begin{array}{rcl} x(r,\phi,\theta) & = & r\sin\theta\cos\phi\,, \\ y(r,\phi,\theta) & = & r\sin\theta\sin\phi\,, \\ z(r,\phi,\theta) & = & r\cos\theta\,. \end{array} \right.$$

Jacobian determinant:

$$\det \frac{\partial(x,y,z)}{\partial(r,\theta,\phi)} = r^2\sin\theta\,.$$

In general, if

$$\left\{ \begin{array}{rcl} x_1 & = & x_1(x'_1,\ldots,x'_n)\,, \\ \vdots & & \vdots \\ x_n & = & x_n(x'_1,\ldots,x'_n)\,, \end{array} \right.$$

is a coordinate transformation (aka substitution of variables), then its Jacobian determinant is defined as

$$\det \frac{\partial(x_1,\ldots,x_n)}{\partial(x'_1,\ldots,x'_n)} = \det \begin{bmatrix} \dfrac{\partial x_1}{\partial x'_1} & \cdots & \dfrac{\partial x_1}{\partial x'_n} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial x_n}{\partial x'_1} & \cdots & \dfrac{\partial x_n}{\partial x'_n} \end{bmatrix}\,.$$

The following observation may sometimes be helpful: $\det \dfrac{\partial(x_1,\ldots,x_n)}{\partial(x'_1,\ldots,x'_n)} = 1/\det \dfrac{\partial(x'_1,\ldots,x'_n)}{\partial(x_1,\ldots,x_n)}$.

Dirac function (minuscule/majuscule letters refer to free, respectively fixed point coordinates, l.h.s. and r.h.s. correspond to Cartesian, respectively polar/cylindrical/spherical representations):

$$\begin{array}{rcl} \delta(x-X)\delta(y-Y) & = & \dfrac{\delta(r-R)\delta(\phi-\Phi)}{r}\,, \\ \delta(x-X)\delta(y-Y)\delta(z-Z) & = & \dfrac{\delta(r-R)\delta(\phi-\Phi)\delta(\zeta-Z)}{r}\,, \\ \delta(x-X)\delta(y-Y)\delta(z-Z) & = & \dfrac{\delta(r-R)\delta(\phi-\Phi)\delta(\theta-\Theta)}{r^2\sin\theta}\,. \end{array}$$

In the above cases, $(X,Y) \in \mathbb{R}^2$, respectively $(X,Y,Z) \in \mathbb{R}^3$, do not coincide with the coordinate origin.

Unit Balls and Spheres

Let $\mathbb{B}^n = \{x \in \mathbb{R}^n \mid \|x\| < 1\}$ be the unit ball in $\mathbb{R}^n$, and $\mathbb{S}^{n-1} = \partial\mathbb{B}^n = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ the corresponding unit n-sphere, then the volume, respectively surface area of $\mathbb{B}^n$ and $\mathbb{S}^{n-1}$ are given by

$$\begin{aligned} \operatorname{vol}_n \mathbb{B}^n &= \frac{\sqrt{\pi}^n}{\Gamma(\frac{n}{2}+1)}, \\ \operatorname{vol}_{n-1} \mathbb{S}^{n-1} &= \frac{2\sqrt{\pi}^n}{\Gamma(\frac{n}{2})}. \end{aligned}$$

Recursions:

$$\begin{aligned} \operatorname{vol}_{n+1} \mathbb{B}^{n+1} &= \frac{\operatorname{vol}_n \mathbb{S}^n}{n+1} \qquad \text{with } \operatorname{vol}_0 \mathbb{S}^0 = 2, \\ \operatorname{vol}_{n+1} \mathbb{S}^{n+1} &= 2\pi \operatorname{vol}_n \mathbb{B}^n \qquad \text{with } \operatorname{vol}_0 \mathbb{B}^0 = 1. \end{aligned}$$

Trigonometric Identities

$$\begin{aligned} \cos^2\alpha + \sin^2\alpha &= 1 \\ \tan\alpha &= \frac{\sin\alpha}{\cos\alpha} \quad \text{if } \alpha \neq \frac{\pi}{2} \pmod{\pi} \\ \cos(\alpha \pm \beta) &= \cos\alpha\cos\beta \mp \sin\alpha\sin\beta \\ \sin(\alpha \pm \beta) &= \sin\alpha\cos\beta \pm \cos\alpha\sin\beta \\ e^{i\phi} &= \cos\phi + i\sin\phi \end{aligned}$$

Mappings

- $A : V \longrightarrow W : v \mapsto w = A(v)$: Generic notation for a mapping (operator, function, etc., cf. figure):
  - A: The label referring to the mapping.
  - V: The *input space* or *domain of the mapping* A.
  - $v \in V$: Fiducial *argument* (input element).
  - W: The *output space* or *codomain of the mapping* A.
  - $w = A(v) \in W$: The *image of* v *under the mapping* A.
  - $A(V) = \{w \in W \mid w = A(v) \text{ for some } v \in V\} \subset W$: The *range of the mapping* A.
  - If $W = A(V)$ then A is called *surjective* or *onto*.
  - If $A(v_1) = A(v_2)$ implies $v_1 = v_2$ (A preserves 'distinctness') then A is called *injective* or *one-to-one*.
  - If A is both injective and surjective it is called *bijective*.
- $\mathscr{L}(V,W) \stackrel{\text{def}}{=} \{A : V \longrightarrow W \mid A \text{ is linear}\}$: The collection of all *linear* operators $V \longrightarrow W$.

Sketch of the prototype $A : V \to W : v \mapsto w = A(v) \in W$ as a 'blackbox' system. Its inner workings are revealed through explicit specification of $A(v) \in W$ as a function of $v \in V$.

Combinatorics

Factorials: $n! \stackrel{\text{def}}{=} n\,(n-1)\,\ldots\,1$. By definition $0! \stackrel{\text{def}}{=} 1$.

Binomial coefficients: $\binom{n}{k} \stackrel{\text{def}}{=} \dfrac{n!}{(n-k)!k!}$.

Newton's *binomium*: $(a+b)^n = \displaystyle\sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k}$.

Multinomial coefficients: $\binom{n}{k_1 \ldots k_\ell} \stackrel{\text{def}}{=} \dfrac{n!}{k_1! \ldots k_\ell!}$ if $k_1 + \ldots + k_\ell = n$, otherwise 0.

Newton's *multinomium*: $(a_1 + \ldots + a_\ell)^n = \displaystyle\sum_{k_1=0}^{n} \ldots \sum_{k_\ell=0}^{n} \binom{n}{k_1 \ldots k_\ell} a_1^{k_1} \ldots a_\ell^{k_\ell}$.

Multi-Indices

An n-dimensional *multi-index* α is an n-tuple of integers $(\alpha_1, \ldots, \alpha_n)$. Its *order* or *norm* is defined as $|\alpha| \equiv \alpha_1 + \ldots + \alpha_n$. Other multi-index conventions are as follows. If $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_1, \ldots, \beta_n)$ are multi-indices of dimension n, $\mathbf{x} = (x_1, \ldots, x_n)$ is a real-valued coordinate n-tuple, and f is a function, then

- $\alpha + \beta \stackrel{\text{def}}{=} (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n)$.
- $\nabla_\alpha f \stackrel{\text{def}}{=} \dfrac{\partial^{\alpha_1 + \ldots + \alpha_n} f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}$.
- $\mathbf{x}^\alpha \stackrel{\text{def}}{=} x_1^{\alpha_1} \ldots x_n^{\alpha_n}$.
- $\alpha! \stackrel{\text{def}}{=} \alpha_1! \ldots \alpha_n!$.
- $\binom{\alpha}{\beta} \stackrel{\text{def}}{=} \displaystyle\prod_{i=1}^{n} \binom{\alpha_i}{\beta_i}$.
- $\displaystyle\sum_{\alpha} \stackrel{\text{def}}{=} \sum_{\alpha_1} \ldots \sum_{\alpha_n}$, etc.

Example. A multivariate analytical function $f \in C^\omega(\Omega)$, $\Omega \subset \mathbb{R}^n$, has a local *Taylor expansion* at a fiducial point $a = (a_1, \ldots, a_n) \in \Omega$. If $x = (x_1, \ldots, x_n) \in \Omega$, we have in multi-index notation

$$f(x) = \sum_\alpha \frac{1}{\alpha!} \nabla_\alpha f(a)\, (x-a)^\alpha \,.$$

The sum runs over all multi-indices $\alpha = (\alpha_1, \ldots, \alpha_n)$ with nonnegative integer entries.

Tensor Indices

A tensor index i in n dimensions is an integer in the range $i \in \{1, \ldots, n\}$. The *Einstein summation convention* applies to pairs of equal indices, one of which must be an *upper index*, the other a *lower index*. If Q is a *mixed tensor* with components Q^i_j, $1 \le i, j \le n$, then

$$Q^i_i \overset{\text{def}}{=} \sum_{i=1}^n Q^i_i \,.$$

Example. A multivariate analytical function $f \in C^\omega(\Omega)$, $\Omega \subset \mathbb{R}^n$, has a local *Taylor expansion* at a fiducial point $a = (a_1, \ldots, a_n) \in \Omega$. If $x = (x_1, \ldots, x_n) \in \Omega$, we have in tensor-index notation

$$f(x) = \sum_{k=0}^\infty \frac{1}{k!} \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n \frac{\partial^k f(a)}{\partial x^{i_1} \ldots \partial x^{i_k}} (x-a)^{i_1} \ldots (x-a)^{i_k} \overset{\text{def}}{=} \sum_{k=0}^\infty \frac{1}{k!} f_{i_1 \ldots i_k}(a) (x-a)^{i_1} \ldots (x-a)^{i_k} \,.$$

The Einstein summation convention has been used on the r.h.s. to suppress the k inner $\sum$-symbols, together with a self-explanatory shorthand for the partial derivative coefficients (note the intentional use of lower indices). Cf. this expression with its multi-index counterpart in the example on page 6.

Condensed Indices

Condensed indices, introduced in the context of quantum field theory [4], are particularly convenient (though not widely known) in image analysis. A condensed index captures both discrete as well as continuous degrees of freedom. Thus for example ϕ^i, using a condensed index $i \equiv (x; m)$ with $x \in \mathbb{R}^n$ and $m \in \mathbb{Z}^k$, corresponds to a k-tuple of functions $x \mapsto \phi_m(x)$ on an n-dimensional continuum.

Condensed indices may simplify notation without loss of generality. The *summation convention* for a condensed index entails a summation over its discrete part as well as an integration over its continuous part. Furthermore, functions of ϕ^i correspond to functionals of $\phi_m(x)$, partial derivatives with respect to ϕ^i to functional derivatives with respect to $\phi_m(x)$, et cetera.

Condensed index convention: Let $i \equiv (x; m)$ and $j \equiv (y; p)$ denote condensed indices, each comprising a continuous label $x, y \in \mathbb{R}^n$ as well as a discrete label $m, p \in \mathbb{Z}^k$, then the following conventions apply:

$$i \overset{\text{def}}{=} (x; m)\,, \quad \phi^i \overset{\text{def}}{=} \phi_m(x)\,, \quad \sum_i \overset{\text{def}}{=} \sum_{m \in \mathbb{Z}^k} \int_{\mathbb{R}^n} dx\,, \quad J(\phi) \overset{\text{def}}{=} J[\phi]\,, \quad \frac{\partial}{\partial \phi^i} \overset{\text{def}}{=} \frac{\delta}{\delta \phi_m(x)}\,, \quad \delta^j_i \overset{\text{def}}{=} \delta^p_m \delta(x-y)\,.$$

Example. Identify $i \equiv y \in \mathbb{R}$, $\alpha \equiv (x; m) \in \mathbb{R} \times \mathbb{Z}^+_0$, $\phi^i \equiv \phi(y) \in V$, $R_{\alpha i} \equiv \frac{d^m}{dx^m} \delta(x-y) \in \mathscr{L}(V, W)$. Here, the indices i and α refer to field components with respect to function spaces V and W, respectively, and

$\mathscr{L}(V,W)$ denotes the vector space of linear transformations $V \longrightarrow W$. Then the linear mapping

$$F_\alpha(\phi) = R_{\alpha i}\,\phi^i$$

corresponds to the m-th order derivative

$$F_m(x)[\phi] = \int_{\mathbb{R}} \frac{d^m}{dx^m}\delta(x-y)\,\phi(y)\,dy = \frac{d^m\phi(x)}{dx^m}\,.$$

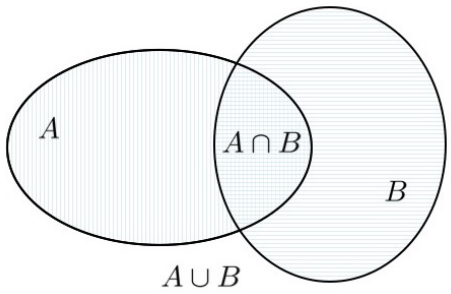
Caveat. The matrix-like condensed index notation for a linear mapping might suggest that linear mappings are always continuous. The example illustrates that this is *not* the case!

Sets

- $x \in A$ means that x is an *element* of the set A. If x is not an element of A we write $x \notin A$.
- $x \in A \cap B$ iff $x \in A$ and $x \in B$ (cf. figure).
- $x \in A \cup B$ iff $x \in A$ or $x \in B$ (cf. figure).
- $X \subset A$ iff $x \in X$ implies $x \in A$. X is called a *subset* of A. Example: $(A \cap B) \subset A$.
- $Y \supset A$ iff $A \subset Y$. Y is called a *superset* of A. Example: $(A \cup B) \supset A$.
- Suppose $X \subset A$, then we write $x \in A\backslash X$ iff $x \in A$ and $x \notin X$.
- $\emptyset$ denotes the empty set with no elements. For any set A, $\emptyset \subset A$. Note that $\emptyset \subset A\backslash\emptyset$, since $A\backslash\emptyset = A$.
- The *characteristic function* or *indicator function on* A, χ_A, is the function defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

- $x \in A \times B$ means that x is of the form $x = (a,b)$, with $a \in A$ and $b \in B$.
- A^k is shorthand for $A \times \ldots \times A$ (k-fold product, $k \in \mathbb{N}$). By definition $A^0 = \emptyset$.



Venn diagram illustrating A, B, $A \cap B$, and $A \cup B$.

Induction Principle

A conjecture $\forall n \in \mathbb{N}: \mathscr{C}(n) = 0$ is logically valid if

- $\mathscr{C}(1) = 0$ (*induction step*), and if
- the *induction hypothesis* $\mathscr{C}(n) \overset{\text{i.h.}}{=} 0$ for any fixed $n \in \mathbb{N}$ implies $\mathscr{C}(n+1) = 0$.

Example. Prove the following conjecture: $\forall n \in \mathbb{N}_0 : \sum_{k=0}^{n} k = \frac{1}{2} n(n+1)$.

Proof. Define $\mathscr{C}(n) \overset{\text{def}}{=} \sum_{k=0}^{n} k - \frac{1}{2} n(n+1)$, then

- $\mathscr{C}(0) = 0$ is obvious, and
- $\mathscr{C}(n+1) \overset{\text{def}}{=} \sum_{k=0}^{n+1} k - \frac{1}{2}(n+1)(n+2) = \sum_{k=0}^{n} k + (n+1) - \frac{1}{2}(n+1)(n+2) = \sum_{k=0}^{n} k - \frac{1}{2} n(n+1) \overset{\text{def}}{=} \mathscr{C}(n) \overset{\text{i.h.}}{=} 0.$

Q.E.D.

1

SIGNALS AND IMAGES: BASICS

1.1. Operational Representations

A (real-valued) n-dimensional digital image is a mapping

$$f : \Omega \longrightarrow \mathbb{R} : (i_1, \ldots, i_n) \mapsto f(i_1, \ldots, i_n)\,, \tag{1.1}$$

in which $\Omega \subset \mathbb{Z}^n$ is a finite grid, usually a regular mesh. The $(n+1)$-tuple $(i_1, \ldots, i_n, f(i_1, \ldots, i_n))$ is called a *pixel* (a shorthand for 'picture element'), with *pixel coordinates* $(i_1, \ldots, i_n)$ and associated *pixel value* $f(i_1, \ldots, i_n)$. Usually—but not always—$n \in \{1, 2, 3, 4\}$, the *dimension* of the image domain Ω, typically a subset of space, time, or spacetime. The notation of Eq. (1.1) can be simplified with the help of the multi-index notation, recall page 5:

$$f : \Omega \longrightarrow \mathbb{R} : \alpha \mapsto f(\alpha)\,, \tag{1.2}$$

with the identification $\alpha = (i_1, \ldots, i_n)$.

A continuum model is often preferred for theoretical convenience:

$$f : \Omega \longrightarrow \mathbb{R} : (x^1, \ldots, x^n) \mapsto f(x^1, \ldots, x^n)\,, \tag{1.3}$$

in which now $\Omega \subset \mathbb{R}^n$ represents a continuum, usually a rectangular block. This is particularly useful for images consisting of many pixels (the typical case). In that case the overall percept of the image is of a continuous rather than discrete nature, so that a continuum representation is appropriate. Often we condense the notation by writing

$$f : \Omega \longrightarrow \mathbb{R} : x \mapsto f(x)\,, \tag{1.4}$$

in which $x = (x_1, \ldots, x_n)$ is a multivariate variable. Unless stated otherwise we shall assume that the image co-domain is (a subset of) $\mathbb{R}$ (for real-valued images) or $\mathbb{C}$ (for complex-valued images).

If the quantization depth, i.e. the number of bits allocated for the storage of a single pixel value, is sufficiently large, it is natural and common to model the image codomain as an interval of $\mathbb{R}$ (or a compact subset of $\mathbb{C}$) instead of a discrete set. In reality the interplay between discretization (i.e. sampling on a discrete grid) and quantization depth determines the 'visual quality' of an image, and thus the appropriateness of a continuum representation, cf. Fig. 1.1 for an example of a black-and-white rendering simulating a grey-tone image through so-called dithering (slick use of only black and white picture elements of spatially varying densities). No need to buy fifty shades of grey ink for your printer!

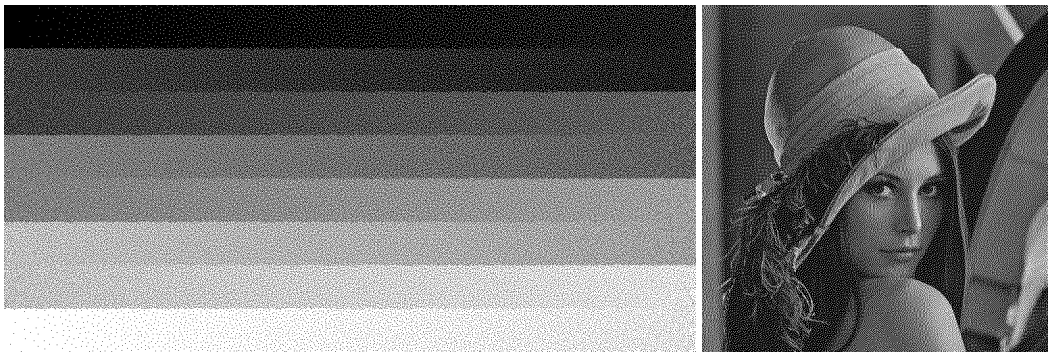


Figure 1.1: Left: Visual illusion of a grey-tone gradient via dithering. Right: Dithering of Lena Söderberg (shot by photographer Dwight Hooker and cropped from the centerfold of a Playboy magazine). This standard test image, known as 'Lena' or 'Lenna', is frequently used by the image processing community, cf. `www.lenna.org`.

Fig. 1.2 shows a 2-dimensional image consisting of 64×64 pixels, which in turn represents a (downsampled) slice obtained from a real 3-dimensional MR brain image. Pixel values have been displayed as grey-tones, with white and black corresponding to the global maximum, respectively minimum pixel value. The digital format of this image may look like this:

$$
\begin{aligned}
f \;=\; & \{\{2,4,2,2,3,2,2,2,3,4,4,3,3,4,4,3,8,19,38,62,85,123,139,155,168,167,174,170,183,165,179,178,\\
& 168,146,100,69,31,11,8,7,6,4,5,7,7,4,4,4,2,4,0,1,3,6,3,5,4,2,3,4,1,4,3,3\},\\
& \vdots\\
& \{4,4,8,8,8,4,5,14,12,61,102,170,121,99,74,98,66,58,97,91,80,34,32,34,29,21,50,66,86,55,46,102,\\
& 68,66,65,95,113,68,65,129,181,186,182,177,120,65,33,18,15,16,21,14,5,6,8,11,4,5,5,6,4,5,3,3\}\}
\end{aligned}
$$

The ellipsis indicates 62 in-between scanlines that have been suppressed.

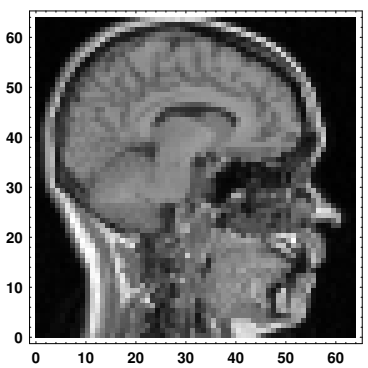


Figure 1.2: Plot of the image f in the previous example.

The effect of discretization on visual appearance can be seen in Figure 1.3. Two degrading factors come into play when the number of grid points (relative to the field of view) becomes too small:

1. A block artifact revealing the nature of the underlying grid.

2. A blur artifact due to insufficient resolution.

The effect of quantization is shown in Figure 1.4. Again, two degrading effects are discernable when quantization becomes too coarse:

1. 'False contouring' revealing the loci of grey-value quantum jumps.

2. Piecewise constant regions revealing the number of bit planes.

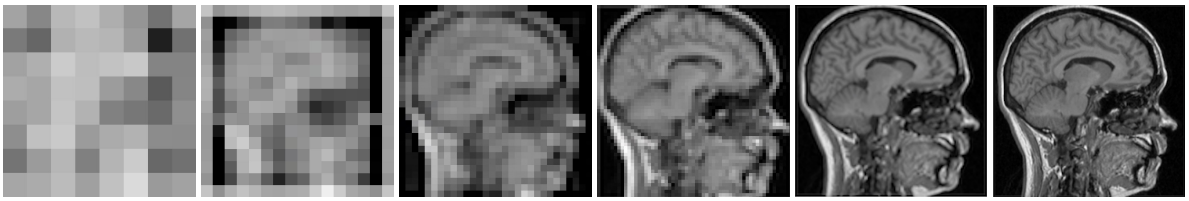


Figure 1.3: From left to right: MR image displayed at a spatial resolution of 8×8, 16×16, 32×32, 64×64, 128×128, 256×256 pixels. In all cases the quantization depth is 1 byte per pixel. As the grid is refined the representation becomes virtually indistinguishable from a continuous one.

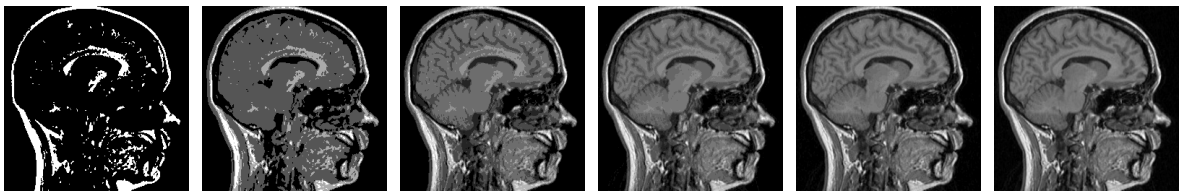


Figure 1.4: From left to right: MR image displayed with quantization depths 1, 2, 3, 4, 5, and 6 bits per pixel. In all cases the discretization grid is 256×256. The 'false contouring' phenomenon on the left is due to insufficient quantization depth. At a quantization depth of 6 bits per pixel this phenomenon is virtually absent.

A 1-dimensional image is normally called a signal. Here is a discrete signal of length $N = 33$:

$$\begin{aligned} s \quad = \quad & \{0,0,0,2,12,51,193,632,1831,4677,10539,20961,36787,56978,77880,93941, \\ & 100000,93941,77880,56978,36787,20961,10539,4677,1831,632,193,51,12,2,0,0,0\}\,. \end{aligned}$$

A signal can be inspected by displaying its (interpolated) graph: Fig. 1.5. Graphical interpolation again expresses our natural inclination to conceive of a signal as a smoothly varying entity defined on a continuum (typically 1-dimensional space or time), which is legitimate in the case of signals consisting of many discrete samples of sufficient quantization depth.

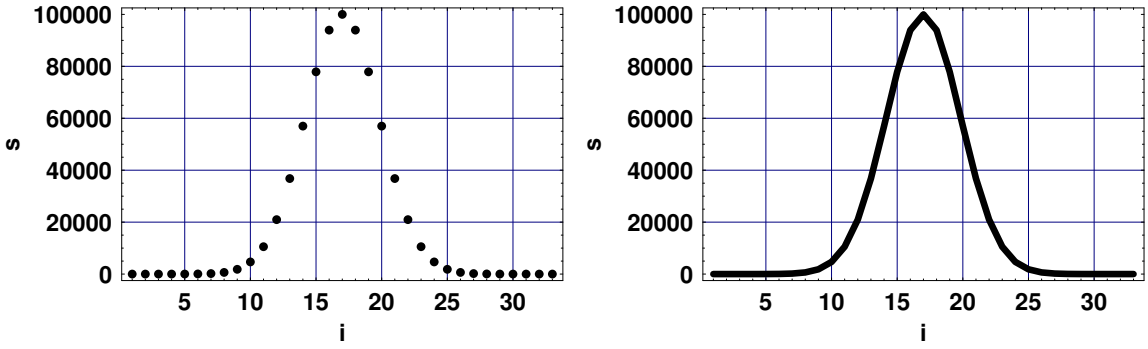


Figure 1.5: Graph of the signal s in the example, without respectively with interpolation.

The display of three-dimensional images ('volume visualization') is nontrivial, cf. Fig. 1.6. Sophisticated algorithms have been developed that exploit various perceptual depth cues, such as shading, motion, and stereo, to create a 3D illusion on a flat screen (sometimes referred to as '$2\frac{1}{2}$D' visualization).

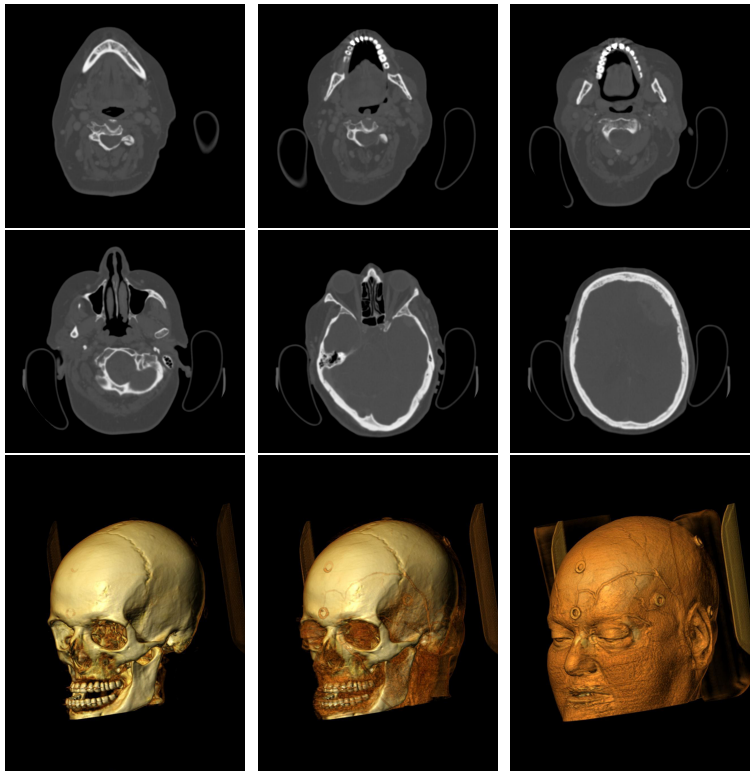


Figure 1.6: Three-dimensional CT (Computed Tomography) image data. Upper two rows: Conventional display slice-wise ('stack of two-dimensional images'). Bottom row: Sophisticated volume visualization emphasizing bone, muscle, respectively skin tissue. This mode of visualization requires prior segmentation (partitioning of the image into meaningful segments) and classification (labelling) of relevant image structures.

Throughout these course notes we will assume that quality factors are such that a continuum representation is appropriate. For this reason we will de-emphasize discretization and quantization aspects after Sections 1.2 and 1.3. Moreover, we will concentrate on general mathematical techniques that are in common use in image processing and analysis.

1.2. From Continuous to Discrete

The procedure of discretizing a function specified on a continuous domain is called *sampling*. The simplest way is to evaluate function values at predefined grid points, a method known as *point evaluation*.

A second discretization method relies on *local averaging*. To this end we average the function over the interior of a grid cell corresponding to a fiducial grid point. Local averaging relies on integration. (Note that dithering could be seen as a kind of 'un-averaging', as it aims to find spatial configurations of black dots with a priori given local pixel averages, the desired grey-tones.) In Chapter 3 we will discuss integration in more detail.

1.3. From Discrete to Continuous

The reverse procedure, defining a continuum representation given a discrete one, is called *interpolation*. A reasonable way to achieve this is to define the value of the interpolated image at an arbitrary sub-pixel location (the base point of interest) as some weighted average of neighbouring pixel values. There exist different interpolation schemes and schemes involving different neighbourhoods and weighing factors.

A simple and popular method is *multilinear interpolation*. To understand how this works we consider *bilinear interpolation*, i.e. multilinear interpolation for the case $n = 2$, cf. Figure 1.7.

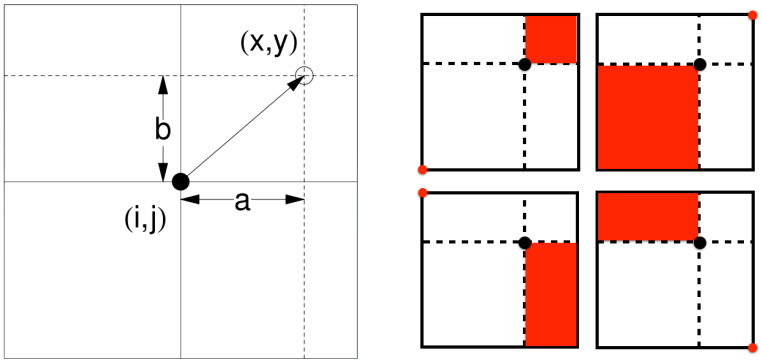


Figure 1.7: Left: The point with integer coordinates (i, j) is a grid point at which the image value $f_d(i, j)$ is known. The point with real-valued coordinates $(x, y) = (i + a, j + b)$, with $0 < a, b < 1$, is an arbitrary point somewhere 'in-between' pixels, to which we would like to assign an interpolated value $f_c(x, y)$. Right: The coloured regions determine the weights by which the indicated pixels (red dots in the corners) contribute to the value of the interpolated point (black dot at the crossing of the dashed lines).

Suppose we wish to evaluate the image value at sub-pixel location $(x, y) = (i+a, j+b)$, in which $(i, j) \in \mathbb{Z}^2$ are the integer coordinates of some grid point, and $0 < a, b < 1$. The interpolated value at this sub-pixel location will be based on those defined at the four nearby grid points (i, j), $(i, j + 1)$, $(i + 1, j)$,and $(i + 1, j + 1)$, the corner points of the rectangular cell in the interior of which the point (x, y) lies. Let $f_c(x, y)$ denote the interpolated function evaluated at (x, y), and $f_d(i, j)$ the discrete image value at the grid point (i, j), then we seek to determine the coefficients of a linear combination of the form

$$f_c(x, y) \stackrel{\text{def}}{=} \alpha\, f_d(i, j) + \beta\, f_d(i + 1, j) + \gamma\, f_d(i, j + 1) + \delta\, f_d(i + 1, j + 1) \quad \text{with } \alpha + \beta + \gamma + \delta = 1.$$

The latter condition guarantees that the interpolated value is a weighted average of the pixel values involved.

Definition 1.3.1. *Let $f_d : \mathbb{Z}^2 \longrightarrow \mathbb{K}$ be a discrete, 2-dimensional image. Then its bilinear interpolation is defined as the function $f_c : \mathbb{R}^2 \longrightarrow \mathbb{K}$ by the following recipe. Let $(x, y) = (i + a, j + b)$ be an arbitrary point in the cell $[i, i + 1] \times [j, j + 1]$, and $0 \leq a, b \leq 1$, then*

$$f_c(x, y) = (1 - a)(1 - b)\, f_d(i, j) + a(1 - b)\, f_d(i + 1, j) + (1 - a)b\, f_d(i, j + 1) + ab\, f_d(i + 1, j + 1)\,.$$

Note that we have included the limiting cases with a and/or b either 0 or 1, and that the coefficients add up to 1. The defining rule that produces the coefficients in this scheme is as follows (cf. the right figure in Fig. 1.7):

- Partition the grid cell comprising the base point of interest into four (2^n in case of multilinear interpolation) rectangles (respectively n-dimensional blocks) by drawing horizontal and vertical lines (respectively n hyperplanes orthogonal to the n coordinate axes) through that base point.
- The weight by which each corner pixel value in this scheme contributes to the requested sub-pixel average corresponds to the relative area (n-dimensional volume) of the subdividing rectangle (block) diagonally opposite to that pixel's grid point.

It should now be obvious how to modify the recipe of Definition 1.3.1 to the cases of *linear interpolation* of a discrete signal and *trilinear interpolation* for establishing a sub-pixel value in a three-dimensional image.

1.4. Histograms and Histogram Transformations

A phenomenological fact of human vision is its limited capability to segregate grey-tones. Fig. 1.4 illustrates that a quantization depth of more than 6 bits per pixel is hardly noticeable. However, many medical imaging modalities produce images with a much higher quantization depth, e.g. 12 bits per pixel is typical for Computed Tomography (CT). There is no way to visually appreciate the rich structure hidden in such an image without interactively manipulating image grey-values (*contrast enhancement*).

An important concept in this context is that of a *grey-value histogram*. This is a density function that expresses the relative occupancy of grey-values in an image. In the continuous case[1],

$$h(u)\, du \stackrel{\text{def}}{=} \text{relative volume of the image domain with grey-levels in the range } [u, u + du). \tag{1.5}$$

Alternatively we can state the definition as follows.

Definition 1.4.1. *Let $u : \mathbb{R}^n \longrightarrow \mathbb{R} : x \mapsto u(x)$ be a real-valued image, then its histogram density function $h_u : \mathbb{R} \longrightarrow \mathbb{R}_0^+ : s \mapsto h(s)$ is defined such that*

$$\int_{u_1}^{u_2} h_u(s)\, ds \stackrel{\text{def}}{=} \textit{relative volume of the image domain with grey-values } u_1 < s < u_2.$$

With this definition the histogram density function is normalized such that

$$\int_{u_-}^{u_+} h_u(s)\, ds = 1\,, \tag{1.6}$$

if u_- and u_+ is the smallest, respectively largest grey-value present in the image. Figure 1.8 shows the (discrete) histograms of the images of Figure 1.4.

[1] In the discrete case the histogram displays the relative number of pixels in grey-value bins $[u, u + \Delta u)$, after one has agreed on the bin-width Δu. In practice one often considers $\mathbb{Z}$-valued images, taking $\Delta u = 1$. Thus e.g. $h(56)$ then counts the number of pixels with value 56, divided by the total number of pixels. Since the total number of pixels is a constant one often refrains from normalizing the histogram.

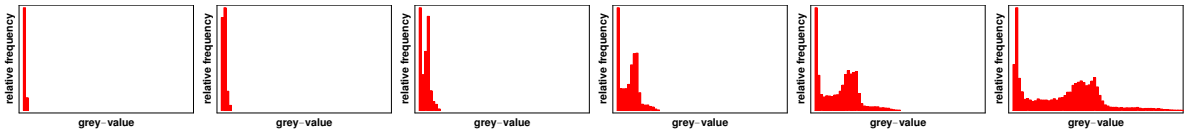


Figure 1.8: Histograms of the images of Figure 1.4. The grey-value interval (horizontal axis) is the same for all graphs. Units on the vertical axes are such that the total area of the bars is unity in each case.

A transformation of type $T : \mathbb{R} \longrightarrow \mathbb{R} : u \mapsto v = T(u)$ can be applied *pointwise* to a real-valued image. That is, if $u : \mathbb{R}^n \longrightarrow \mathbb{R}$ is an n-dimensional image, we may transform its grey-values by setting $v(x) = T(u(x))$ for each $x \in \mathbb{R}^n$, in other words, $v = T \circ u$, which obviously affects the histogram of the image. For this reason T is called a *histogram transformation*. If, in addition, $T \in C^1(\mathbb{R})$ and

$$T'(s) \neq 0 \quad \text{for all } s \in \mathbb{R}, \tag{1.7}$$

then the transformation is one-to-one (i.e. it preserves 'distinctness', recall page 4) and essentially no information is lost. Knowing the transformed image v and the grey-value transformation T, the original image u can be recovered, viz. $u(x) = T^{\text{inv}}(v(x))$ for all $x \in \mathbb{R}^n$, i.e. $u = T^{\text{inv}} \circ v$. The condition of Eq. (1.7) is needed for the inverse transformation to exist on the range of T. In fact, T must be either monotonically increasing ($T'(s) > 0$) or decreasing ($T'(s) < 0$). The latter has the effect that light and dark are interchanged.

It may happen that an image appears too dark or too bright when rendered on a screen (or printer) without modification of grey-values. In that case one would like to brighten or darken grey-levels in a certain range of interest so as to improve contrast. For instance, if we want to enhance contrast near the middle of the image's grey-value range, we could use histogram transformations such as those in Figure 1.9. The effect is illustrated in Figure 1.10, where these transformations are applied to an MRI slice. The price for contrast enhancement in this example is that contrast in very dark and very bright regions is suppressed. In the extreme case contrast stretching becomes *thresholding*, cf. the rightmost graph in Figure 1.9 and the binary result on the right in Fig. 1.10. Note that in this limit information will be lost, since it violates Eq. (1.7).

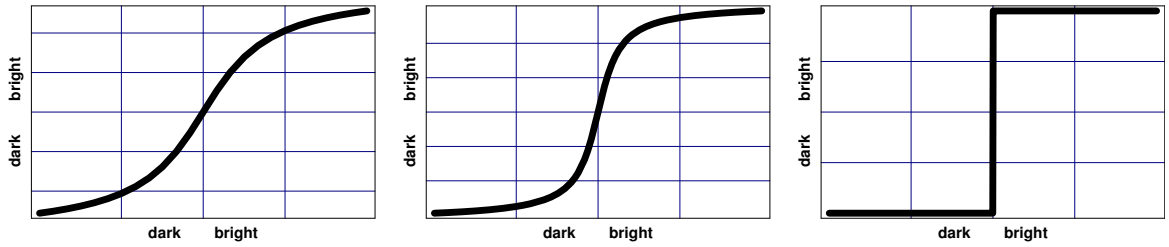


Figure 1.9: Histogram stretching from moderate (left) to extreme (right).

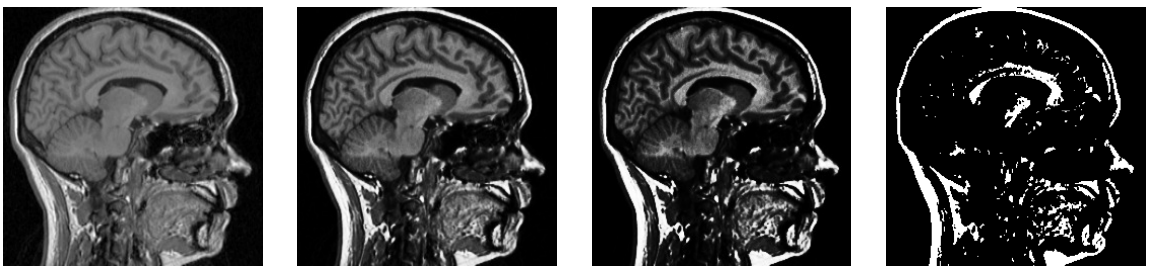


Figure 1.10: Original MRI image (left) and the effect of applying the contrast stretching transformations depicted in Figure 1.9. The price for contrast enhancement in a certain range is contrast loss elsewhere.

Image enhancement refers to any type of manipulation that enhances some (codomain) feature of interest, such as grey-level contrast. Other basic grey-level transformations frequently used for image enhancement include logarithmic, exponential, and power-law transformation. Some manipulations affect image values 'indirectly' by acting within the image domain (such as rotations, translations, scalings, or warpings).

In order to process or analyze images one must first of all furnish suitably chosen classes with appropriate internal structure so as to enable the kind of operations one wishes to perform on these images. Moreover, any definition one comes up with needs to be formulated in terms of *operational concepts* (precise 'recipes' or algorithms). This poses rather strong 'admissibility' constraints on mathematical concepts. Many well-established and cherished 'classical' results fail the test! For instance, if we want to apply derivative operators to analyze the local structure of an image (the first thing a mathematically inclined image engineer should think of), we will have to rethink the entire machinery of infinitesimal calculus bottom-up, since, from an operational point of view, it is founded on 'infinitesimal nonsense' without operational significance. The pivotal '$\lim_{\epsilon \to 0}$' operation is a physically void mental construct.

In the next chapter we review some classical mathematical concepts every image scientist should be familiar with, de-emphasizing operational details. The latter will be addressed in subsequent chapters, notably differentiation and integration in Chapter 3 and Fourier transformation in Chapter 4.

2

BASIC MATHEMATICAL CONCEPTS

2.1. Groups and Semigroups

Image transformations can be characterized as either reversible ('information preserving') or irreversible ('lossy'). For instance, an image rotation can be 'undone' by another (viz. inverse) rotation, but the effect of blurring or downsampling cannot. This leads us to consider the mathematical concepts of a *group* and of a *semigroup*. The former captures the notion of reversibility and some other natural requirements. The latter formalizes an important class of irreversible operations.

Definition 2.1.1. *A group is a collection G together with an internal operation*

$$\circ : G \times G \longrightarrow G : (x, y) \mapsto x \circ y\,, \quad \textit{such that}$$

- *the operation is associative, i.e.* $(x \circ y) \circ z = x \circ (y \circ z)$ *for all* $x, y, z \in G$,
- *there exists an identity element* $e \in G$ *such that* $x \circ e = e \circ x = x$ *for all* $x \in G$, *and*
- *for each* $x \in G$ *there exists an inverse element* x^{inv} *such that* $x^{\text{inv}} \circ x = x \circ x^{\text{inv}} = e$.

The above definition implies several properties not listed explicitly as an axiom, e.g.

- the identity element $e \in G$ is unique,
- $e^{\text{inv}} = e$,
- the inverse element x^{inv} for any given $x \in G$ is unique,
- $(x^{\text{inv}})^{\text{inv}} = x$ for all $x \in G$.

(Prove these, using only axioms from Definition 2.1.1.)

Definition 2.1.2. *A semigroup is a collection G together with an internal operation*

$$\circ : G \times G \longrightarrow G : (x, y) \mapsto x \circ y \,,$$

such that

- *the operation is associative, i.e. $(x \circ y) \circ z = x \circ (y \circ z)$ for all $x, y, z \in G$,*
- *there exists an identity element $e \in G$ such that $x \circ e = e \circ x = x$ for all $x \in G$.*

Definition 2.1.3. *Recall Definitions 2.1.1–2.1.2. If, in addition, $x \circ y = y \circ x$ for all $x, y \in G$, then we call the (semi)group commutative, or abelian.*

If you want to verify whether a set G constitutes a (semi)group relative to some stipulated infix operator $\circ$, the *first* thing to prove is the *closure property*, i.e. the property that taking $x, y \in G$ arbitrarily implies $x \circ y \in G$. (This property is implicit in the notation '$G \times G \longrightarrow G$'.)

2.1.1. Examples and Counterexamples of Groups

Conjecture 2.1.1. *Continuously differentiable bijective histogram transformations on $\Omega \subset \mathbb{R}$ constitute a group*

$$G = \left\{ T \in C^1(\Omega) \mid T'(s) \neq 0 \text{ for all } s \in \Omega \text{ and } T(\Omega) = \Omega \right\} .$$

If the group elements are functions it will be tacitly assumed, unless stated otherwise, that the group operator $\circ$ refers to *function composition*, i.e. if $S, T \in C^1(\Omega)$, then $S \circ T \in C^1(\Omega)$ is the function given by

$$(S \circ T)(s) \stackrel{\text{def}}{=} S(T(s)) \quad \text{for all } s \in \Omega. \tag{2.1}$$

The proof of Conjecture 2.1.1 is as follows.

- Closure: If $S, T \in G$, then $S \circ T \in C^1(\Omega)$. More specifically, we have the chain rule: $(S \circ T)'(s) = S'(T(s))\, T'(s) \neq 0$, and $(S \circ T)(\Omega) = S(T(\Omega)) = S(\Omega) = \Omega$, whence $S \circ T \in G$.
- Associativity: Let $S, T, U \in G$, then $((S \circ T) \circ U)(s) = (S \circ T)(U(s)) = S(T(U(s))) = S((T \circ U)(s)) = (S \circ (T \circ U))(s)$ for all $s \in \Omega$, whence $(S \circ T) \circ U = S \circ (T \circ U)$.
- Identity element: Let $E : \Omega \longrightarrow \Omega : s \mapsto E(s) = s$, then clearly $E \in C^1(\Omega)$ and $E'(s) = 1 \neq 0$, so that $E \in G$. For any $T \in G$ we have $(T \circ E)(s) = T(E(s)) = T(s) = E(T(s)) = (E \circ T)(s)$ for all $s \in \Omega$, so that $T \circ E = E \circ T = T$. Thus the *identity function* $E \in G$ is the identity element.

- Inverse element: Let $T \in G$, then it has been argued that T has an inverse $T^{\text{inv}} \in C^1(\Omega)$. Graphically you can construct the graph of this inverse function by mirroring the graph of T with respect to the graph of the identity element E. From this construction it is also obvious that

$$\left.\frac{dT^{\text{inv}}(t)}{dt}\right|_{t=T(s)} = \left(\frac{dT(s)}{ds}\right)^{-1} \neq 0\,,$$

so that in fact $T^{\text{inv}} \in G$. Alternatively: $1 = E' = (T^{\text{inv}} \circ T)' = (T^{\text{inv}\prime} \circ T)T'$, whence $T^{\text{inv}\prime} \circ T = 1/T'$ is nonzero. Likewise, $1 = E' = (T \circ T^{\text{inv}})' = (T' \circ T^{\text{inv}})T^{\text{inv}\prime}$, whence $T^{\text{inv}\prime} = 1/(T' \circ T^{\text{inv}})$ is nonzero. Finally, from $T(\Omega) = \Omega$ it is immediately obvious that $T^{\text{inv}}(\Omega) = \Omega$.

Remark. Notice that for associativity we did not need any *specific* properties of the function space, other than the requirement that domain and range of the functions be compatible, so as to admit arbitrary compositions.

Definition 2.1.4. *A subgroup $F \subset G$ of a group G is a subset of G which is itself a group.*

The subgroup F inherits its group operation from the group G. The following theorem is known as the *subgroup theorem*.

Theorem 2.1.1. *Let G be a group and $F \subset G$. If for all $f_1, f_2 \in F$ we have $f_1 \circ f_2 \in F$ and $f_1^{\text{inv}} \in F$, then F is itself a group, in other words, a subgroup of G.*

Conjecture 2.1.2. *The set of histogram transformations $G^+ = \left\{T \in C^1(\Omega) \mid T'(s) > 0 \text{ for all } s \in \Omega \text{ and } T(\Omega) = \Omega\right\}$ constitutes a group. Thus $G^+ \subset G$ is a subgroup of the group defined in Conjecture 2.1.1.*

Proof. Either reconsider the steps of the previous proof for G in the context of G^+ by replacing the requirement $T'(s) \neq 0$ by $T'(s) > 0$, or apply the subgroup theorem, Theorem 2.1.1. What happens if we define the set G^- by replacing the condition $T'(s) > 0$ in the definition of G^+ by $T'(s) < 0$? (Answer: G^- does *not* define a group.) The details are left as an exercise.

As another example, consider the *negative transformation*

$$T : \mathbb{R} \longrightarrow \mathbb{R} : s \mapsto T(s) = -s\,. \tag{2.2}$$

It clearly satisfies the requirement of Eq. (1.7), with $T'(s) = -1 \neq 0$ for all $s \in \mathbb{R}$. Figure 2.1 shows the (pixelwise) effect of this transformation on an MRI slice.

Conjecture 2.1.3. *The two-element set of histogram transformations $D = \{E, T\}$, with T as given by Eq. (2.2), and $E : \mathbb{R} \longrightarrow \mathbb{R} : s \mapsto E(s) = s$, i.e. the identity map, constitutes a commutative group. That is, $D \subset G$ is a commutative subgroup of the group defined in Conjecture 2.1.1.*

Since this group, unlike previous examples, has a *finite* number of elements, it is also referred to as a *discrete group* or a *finite group*.

Proof. In the case of a discrete group with n elements, D_n say, it is often helpful to evaluate all possible products $x_i \circ x_j$ explicitly for $x_i, x_j \in D_n$, $i, j = 1, \ldots, n$, and to represent the result as a table ($n \times n$ square matrix), cf. Table 2.1. The proof then follows from the group multiplication table, as follows.

- Closure: In this case one can evaluate all possible products, cf. Table 2.1. (Identify $x_1 \equiv E$, $x_2 \equiv T$.)
- Associativity: It has been argued that this holds for functions of the specified type in general.
- Identity element: From the first column, respectively the first row of Table 2.1 it is clear that $E \in G$ is the identity element, since $x_i \circ E = E \circ x_i = x_i$ for each $i = 1, 2$.

- Inverse element: From the diagonal of Table 2.1 it is clear that each element equals its own inverse, i.e. $x_i^{\text{inv}} = x_i \in G$ for all $i = 1, 2$.

- Commutativity: From the symmetry of Table 2.1 it is clear that $x_i \circ x_j = x_j \circ x_i$ for all $i, j = 1, 2$.

	$j = 1$	$j = 2$
$i = 1$	E	T
$i = 2$	T	E

Table 2.1: Product table for the discrete group of Conjecture 2.1.3. The (i, j)-th element in this table contains the product $x_i \circ x_j$ in which $x_1 = E$ and $x_2 = T$ enumerate all elements of G.

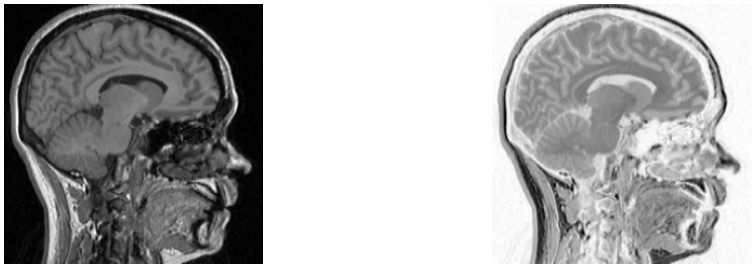


Figure 2.1: MRI image slice $u : \mathbb{R}^2 \longrightarrow \mathbb{R} : (x, y) \mapsto u(x, y)$ and its negative $T \circ u : \mathbb{R}^2 \longrightarrow \mathbb{R} : (x, y) \mapsto (T \circ u)(x, y) = T(u(x, y))$.

Figure 2.2 shows the power-law transformation

$$T_\gamma : [0, 1] \longrightarrow [0, 1] : s \mapsto T_\gamma(s) = s^\gamma \quad (\gamma \in \mathbb{R}^+) \,, \tag{2.3}$$

for eight different parameter values. This type of transformation is known as a *gamma correction*. Figure 2.3 shows the result of gamma correction for several values of γ. Figure 2.4 shows the corresponding histograms.

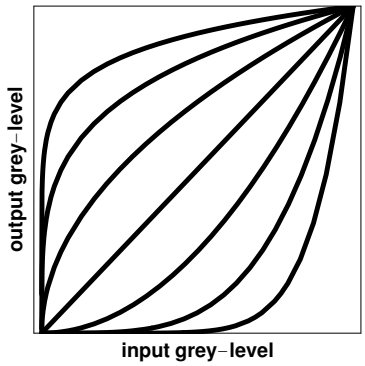


Figure 2.2: Power-law transformations $T_\gamma(s)$ with $s \in [0, 1]$ for $\gamma = 1/8, 1/4, 1/2, 1, 2, 4, 8$. It is left as an exercise to attach the appropriate γ-label to each graph.

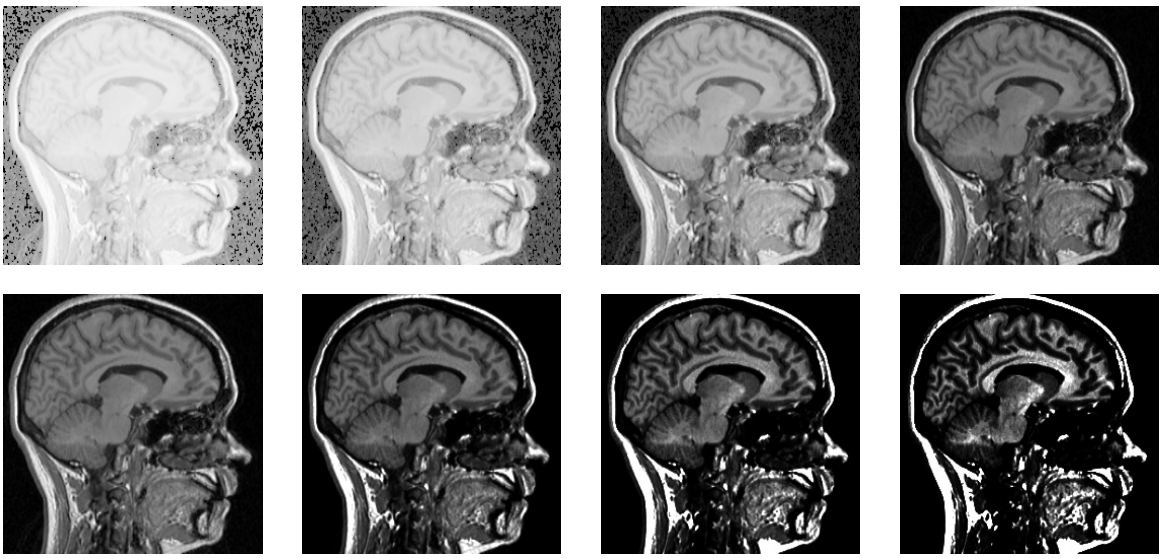


Figure 2.3: Gamma corrections with $\gamma = 1/8, 1/4, 1/2, 1$ (top row), respectively $\gamma = 1, 2, 4, 8$ (bottom row).

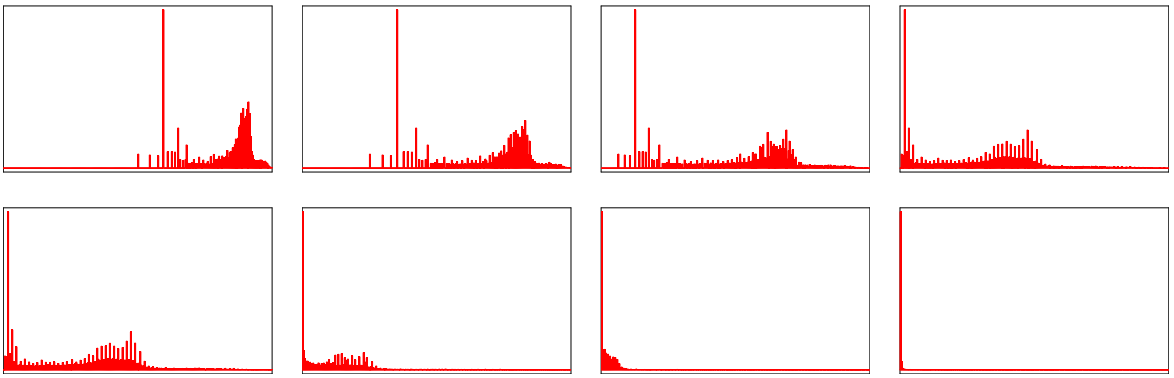


Figure 2.4: Histograms corresponding to the images of Figure 2.3.

Conjecture 2.1.4. *Cf. Eq. (2.3). Gamma corrections,* $\Gamma = \{T_\gamma \mid \gamma \in \mathbb{R}^+\}$*, constitute a commutative group.*

Proof. The following lemma is useful in this context.

Lemma 2.1.1. *Cf. Eq. (2.3). For* $\gamma, \mu \in \mathbb{R}^+$ *we have* $T_\gamma \circ T_\mu = T_{\gamma\mu}$.

Again we have taken function composition as the default internal operation: $(T_\gamma \circ T_\mu)(s) = T_\gamma(T_\mu(s))$ for all $s \in \mathbb{R}$. The proof of the lemma is given in the first item below.

- Closure: Let $\gamma, \mu \in \mathbb{R}^+$ be arbitrary. Then $(T_\gamma \circ T_\mu)(s) = T_\gamma(T_\mu(s)) = (s^\mu)^\gamma = s^{\gamma\,\mu} = T_{\gamma\mu}(s)$ for all $s \in \mathbb{R}$, whence $T_\gamma \circ T_\mu = T_{\gamma\mu} \in \Gamma$, since $\gamma\mu \in \mathbb{R}^+$. This proves closure and in particular Lemma 2.1.1.
- Associativity follows from the fact that function composition is associative by construction, as we have seen previously.
- The identity element is $T_1 \in \Gamma$, i.e. the identity function, for we have, according to Lemma 2.1.1, $T_\gamma \circ T_1 = T_1 \circ T_\gamma = T_\gamma$ for any $\gamma \in \mathbb{R}^+$.

- The inverse element of T_γ is $T_{1/\gamma}$, for we have, according to Lemma 2.1.1, $T_\gamma \circ T_{1/\gamma} = T_{1/\gamma} \circ T_\gamma = T_1$, for any $\gamma \in \mathbb{R}^+$.
- Commutativity also follows with the help of Lemma 2.1.1, since for any $\gamma, \mu \in \mathbb{R}^+$ we have $T_\gamma \circ T_\mu = T_{\gamma\mu} = T_{\mu\gamma} = T_\mu \circ T_\gamma$.

Remark. Note that several group properties have been related to (group) properties of real number multiplication.

2.2. Vector Spaces

2.2.1. Vector Spaces in General

In order to handle signals and images conceptually we need to endow collections of these with some suitable structure. Probably the most important structure one can impose is that of a *linear space* or *vector space*. Recall the definition of a vector space [28, 29]:

Definition 2.2.1. *A vector space V over the field $\mathbb{K}$ is a set of objects that can be added and multiplied by scalars, such that the sum of two elements of V as well as the product of an element in V with a scalar are again elements of V (closure property). Moreover, the following properties are satisfied for any $u, v, w \in V$ and $\lambda, \mu \in \mathbb{K}$:*

- $(u + v) + w = u + (v + w)$,
- *there exists an element $o \in V$ such that $o + u = u$,*
- *there exists an element $-u \in V$ such that $u + (-u) = o$,*
- $u + v = v + u$,
- $\lambda \cdot (u + v) = \lambda \cdot u + \lambda \cdot v$,
- $(\lambda + \mu) \cdot u = \lambda \cdot u + \mu \cdot u$,
- $(\lambda\mu) \cdot u = \lambda \cdot (\mu \cdot u)$,
- $1 \cdot u = u$.

The infix operator $+$ denotes *vector addition*, the infix operator $\cdot$ denotes *scalar multiplication*. Note that a vector space is an extension of an abelian ('additive') group (closure and first four items) with the additional option to scale or magnify elements (last four items), recall Definitions 2.1.1 and 2.1.3.

Caveat. Never confuse vector addition and scalar multiplication with addition and multiplication of ordinary numbers.

If we talk about 'the vector space V' without reference to the field $\mathbb{K}$, the nature of the latter is supposed to be clear from the context (or otherwise immaterial for the discussion). In practice one usually omits the scalar multiplication sign, since no confusion is likely to arise. The above definition implies several properties not listed explicitly, such as

- $u + o = u$,
- $(-1) \cdot u = -u$,

- $0 \cdot u = o$,
- the null vector $o \in V$ is unique,
- $-o = o$,
- the antivector $-u$ for any given $u \in V$ is unique,
- $-(-u) = u$.

(Prove these, using only axioms from Definition 2.2.1.) These properties also justify some sloppiness of notation one frequently encounters, such as writing the null element o simply as 0, which is exactly what we will do henceforth, unless explicitly stated otherwise. It should be clear from the context whether this indicates the null vector in V or the scalar number zero in $\mathbb{K}$.

Caveat. Never confuse the vector $0 \in V$ with the ordinary number $0 \in \mathbb{K}$.

A (linear) *subspace* is a subset of a vector space which is itself a vector space. The criterion to be verified in order to determine whether a subset is indeed a subspace is the following closure property, a.k.a. the *subspace theorem*.

Theorem 2.2.1. *Let V be a vector space over $\mathbb{K}$ and $W \subset V$. If for all $w_1, w_2 \in W$ and all $\lambda, \mu \in \mathbb{K}$ we have $\lambda\, w_1 + \mu\, w_2 \in W$, then W is itself a vector space, in other words, a subspace of V.*

In some textbooks vectors are defined as n-tuples of numbers ('column' or 'row vectors', depending on layout), e.g. $(x, y) \in \mathbb{R}^2$ is considered a (row) vector in $n = 2$ dimensions. Indeed, if we define $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$ and $\lambda\,(x, y) = (\lambda\, x, \lambda\, y)$, then all properties of Definition 2.2.1 are satisfied. (Verify this!) However, n-tuples endowed with these addition and scalar multiplication rules are only specific instances of vectors. Definition 2.2.1 is more general and admits other objects than n-tuples.

Caveat. Never confuse a vector with a set of numbers (a 'row' or 'column vector').

Signals and images can be regarded as elements of a vector space. To this end one must state how addition and scalar multiplication are to be performed such that the defining vector space axioms are satisfied.

For discrete signals and images (sharing a common domain of definition) the default vector addition and scalar multiplication rules are defined in terms of entry-wise 'ordinary' (i.e. numerical) addition and multiplication, as follows. Let $\alpha \in \mathbb{Z}^n$ label any grid point on an n-dimensional grid, and let $s, t : \mathbb{Z}^n \longrightarrow \mathbb{K}$ and $\lambda \in \mathbb{K}$, then

$$(s + t)(\alpha) = s(\alpha) + t(\alpha)\,, \tag{2.4}$$
$$(\lambda \cdot s)(\alpha) = \lambda s(\alpha)\,. \tag{2.5}$$

Likewise, for continuous signals and images vector addition and scalar multiplication are defined in terms of point-wise 'ordinary' counterparts. Let $x \in \mathbb{R}^n$ denote any point in an n-dimensional continuous domain, and let $f, g : \mathbb{R}^n \longrightarrow \mathbb{K}$ and $\lambda \in \mathbb{K}$, then

$$(f + g)(x) = f(x) + g(x)\,, \tag{2.6}$$
$$(\lambda \cdot f)(x) = \lambda f(x)\,. \tag{2.7}$$

A vector space of functions is often called a *function space*. Notice the different meanings of summation and multiplication symbols on left and right hand sides in Eqs. (2.4–2.7). You should verify yourself that all properties listed in Definition 2.2.1 are indeed satisfied in these cases.

Definition 2.2.2. *The span of a set of vectors $\{v_1, \ldots, v_N\} \subset V$ is the vector space consisting of all linear combinations ('superpositions') of those vectors:* $\operatorname{span}\{v_1, \ldots, v_N\} \stackrel{\text{def}}{=} \{\lambda_1\, v_1 + \ldots + \lambda_N\, v_N \mid \lambda_1, \ldots, \lambda_N \in \mathbb{K}\}$.

Theorem 2.2.2. *The span of a set of vectors constitutes a vector space.*

The proof is left as an exercise.

Definition 2.2.3. *A set of N vectors $\{v_1, \ldots, v_N\} \subset V$ is called a basis of V if*

- $\operatorname{span}\{v_1, \ldots, v_N\} = V$, *and*
- $\lambda_1 \cdot v_1 + \ldots + \lambda_N \cdot v_N = 0$ *if and only if* $\lambda_1 = \ldots = \lambda_N = 0$.

The first property says that you can construct any element in V as a suitable linear combination of basis vectors. The second property says that a basis is not redundant in the sense that the elements are linearly independent.

Definitions 2.2.2 and 2.2.3 can be extended to include the formal case $N = \infty$, provided suitable conditions are imposed on the amplitudes $\{\lambda_i\}_{i=1}^{\infty}$. The details are beyond the scope of these course notes, but important examples of such ∞-dimensional vector spaces are the function classes listed in the Preliminaries on page 2. You should verify that these indeed constitute vector spaces. (Hint: Theorem 2.2.1 is applicable in several cases.)

2.2.2. Inner Product Spaces

Vector spaces can be endowed with additional structure. An inner product specifies, in a precise sense, the 'similarity' of two vectors, accounting for both magnitude and relative direction. (The notion of 'magnitude' is captured by a so-called *norm induced by the inner product*, cf. Section 2.2.3. The notion of 'relative direction' can be expressed in terms an angle, cf. Eq. (2.18) below.)

Definition 2.2.4. *Let V be a vector space over $\mathbb{C}$. A complex inner product is a nondegenerate positive definite hermitean sesquilinear mapping $\langle \,|\, \rangle : V \times V \longrightarrow \mathbb{C}$ satisfying the following properties. For all $u, v, w \in V$ and $\lambda, \mu \in \mathbb{C}$ we have*

- $\langle \lambda u + \mu v | w \rangle = \lambda^* \langle u|w\rangle + \mu^* \langle v|w\rangle$ *('semi-linearity' in first argument),*
- $\langle u | \lambda v + \mu w \rangle = \lambda \langle u|v\rangle + \mu \langle u|w\rangle$ *(linearity in second argument),*
- $\langle u|v\rangle = \langle v|u\rangle^*$ *(conjugate symmetry),*
- $\langle u|u\rangle > 0$ *for all* $u \neq 0$ *(positivity and nondegeneracy).*

The term 'sesquilinear' refers to the first two properties and indicates that the inner product is 'almost' bilinear. The inner product would be bilinear if there were no conjugation symbol in the first rule. This is of course the case if we consider a real inner product on a real vector space.

Definition 2.2.5. *Let V be a vector space over $\mathbb{R}$. A real inner product is a nondegenerate positive definite symmetric bilinear mapping $\langle \,|\, \rangle : V \times V \longrightarrow \mathbb{R}$.*

It is left as an exercise to restate the defining rules for a real inner product on a real vector space, Definition 2.2.5, in terms of explicit formulas similar to those in Definition 2.2.4.

Thus an inner product can be seen as a 'slot machine' with two empty slots. You can drop a vector into each slot and the machine produces a number for you. Even if you don't know the precise inner workings of that machine its input-output relation must satisfy Definition 2.2.4 or Definition 2.2.5.

Example 2.2.1. Let $V = \mathbb{C}^2$ with the usual complex vector space structure and let $v = (v_1, v_2)$ and $w = (w_1, w_2)$ be two row vectors in V. Then the *standard inner product* is given by

$$\langle v|w\rangle = v_1^* \, w_1 + v_2^* \, w_2 \,. \tag{2.8}$$

This rule also applies to column vectors. The standard inner product for row or column vectors is often written as $v \cdot w$ instead of $\langle v|w\rangle$.

You should verify yourself that the standard inner product in this example is indeed a complex inner product. The example can be generalized to discrete signals and images.

Example 2.2.2. Let r and s be discrete signals of equal length N. A complex inner product can then be defined component-wise as follows:

$$\langle r|s\rangle = \sum_{i=1}^{N} r(i)^* \, s(i) \,. \tag{2.9}$$

Likewise for a real inner product (no conjugation symbol).

You should again verify that this is indeed an inner product according to Definition 2.2.4.

Example 2.2.3. For analogue signals defined on the unit interval $[0, 1]$, say, we define the standard inner product as follows:

$$\langle r|s\rangle = \int_0^1 r(t)^* \, s(t) \, dt \,. \tag{2.10}$$

Once again, you should check yourself that this is indeed an inner product according to Definition 2.2.4.

Example 2.2.4. For n-dimensional discrete images f, g on a finite regular grid $\Omega \subset \mathbb{Z}^n$ we define the standard inner product as

$$\langle f|g\rangle = \sum_{\alpha\in\Omega} f(\alpha)^* \, g(\alpha) \,, \tag{2.11}$$

in which α is a multi-index coordinate of the grid points.

Example 2.2.5. Finally, for functions on a continuous domain $\Omega \subset \mathbb{R}^n$ the standard inner product is defined as follows:

$$\langle f|g\rangle = \int_\Omega f(x)^* \, g(x) \, dx \,. \tag{2.12}$$

Of course, the analogue functions in Examples 2.2.3 and 2.2.5 must such that the integrals converge. This also holds for the discrete series in Examples 2.2.2 and 2.2.4 in case $N = \infty$ or if Ω is not a finite grid.

Caveat. The standard or 'usual' inner products given by Eqs. (2.9–2.12) are merely instances of inner products. The defining properties of Definition 2.2.4 admit many non-standard choices.

Example 2.2.6. Consider images of the type $f, g : \mathbb{R}^n \longrightarrow \mathbb{K}$, and let $m : \mathbb{R}^n \longrightarrow \mathbb{R}^+$ be a fixed, positive definite function, such that

$$\langle f|g\rangle_m = \int_{\mathbb{R}^n} f(x)^* \, g(x) \, m(x) \, dx \tag{2.13}$$

is well-defined.

You should verify yourself that this is an inner product (and, as you go, establish why the function m needs to be strictly positive). The function m acts as a 'weight' function and is called a 'measure'. Eq. (2.13) is clearly not the standard inner product, unless $m(x) = 1$ for all $x \in \mathbb{R}^n$.

Example 2.2.7. As a second example of a non-standard inner product, consider the function space $H^1(\mathbb{R})$, consisting of functions $f : \mathbb{R} \longrightarrow \mathbb{K}$ for which both f itself as well as its derivative f' are elements of $L^2(\mathbb{R})$. Then you may verify that

$$\langle f|g\rangle_{H^1} = \int_{-\infty}^{\infty} \left(f(x)^* \, g(x) + f'(x)^* \, g'(x)\right) \, dx \tag{2.14}$$

is an inner product. (We assume that the integral is well-defined, v.i.)

You should verify yourself why we cannot omit the first (zeroth order) term in the integrand.

Assumption 2.2.1. *Whenever we talk about inner products of signals or images we shall mean the standard inner products defined according to Eqs. (2.9–2.12), unless stated otherwise.*

In an inner product space any basis can be orthogonalized and, if desired, normalized.

Definition 2.2.6. *Recall Definition 2.2.3. Let V be an inner product space. A set of N vectors $\{e_1, \ldots, e_N\} \subset V$ is an orthogonal basis of V if it is a basis and $\langle e_i | e_j \rangle = c_i \, \delta_{ij}$, for some $c_i \in \mathbb{R}^+$. A set of N vectors $\{e_1, \ldots, e_N\} \subset V$ is an orthonormal basis of V if it is a basis and $\langle e_i | e_j \rangle = \delta_{ij}$.*

There exist systematic procedures for generating an orthonormal basis from an arbitrary one. A popular one is the *Gram-Schmidt orthogonalization process*. This will be discussed in Section 2.3, notably Theorem 2.3.4.

With the help of an orthonormal basis one can represent any vector uniquely in terms of a tuple of numbers, or coefficients (relative to that basis). This is in fact true—by definition—for any basis, but orthonormality greatly simplifies things as the following theorem shows.

Theorem 2.2.3. *Let V be an inner product space, with orthonormal basis $\{e_1, \ldots, e_N\} \subset V$, and $v \in V$ any vector. Then there exists a unique decomposition*

$$v = \sum_{i=1}^{N} c_i \, e_i \, .$$

The N-tuple of coefficients c_i are given by $c_i = \langle e_i | v \rangle$.

Proof. The first part is true by definition (otherwise $\{e_1, \ldots, e_N\} \subset V$ would not constitute a basis). To determine the coefficients, consider

$$\langle e_i | v \rangle = \langle e_i | \sum_{j=1}^{N} c_j \, e_j \rangle = \sum_{j=1}^{N} c_j \, \langle e_i | e_j \rangle = \sum_{j=1}^{N} c_j \, \delta_{ij} = c_i \, .$$

This completes the proof. □

Theorem 2.2.3 can be generalized to so-called *Hilbert spaces*, in which there exists a countable, but not necessarily finite orthonormal basis ($N = \infty$). This subject is beyond the scope of these course notes.

2.2.3. Normed Spaces and Metric Spaces

Instead of an inner product it sometimes suffices to provide a vector space with a *norm*. A norm specifies, in a precise sense, how 'big' a vector is.

Definition 2.2.7. *Let V be a vector space over $\mathbb{K}$. A norm is a nondegenerate positive definite mapping $\| \; \| : V \longrightarrow \mathbb{R}$ such that for all $v, w \in V$, $\lambda \in \mathbb{K}$,*

- $\|v\| \geq 0$ *and* $\|v\| = 0$ *if and only if* $v = 0$,
- $\|\lambda \, v\| = |\lambda| \, \|v\|$,
- $\|v + w\| \leq \|v\| + \|w\|$.

A vector space equipped with a norm is called a *normed vector space*, or simply a *normed space*. The last property in Definition 2.2.7 is known as the *triangle inequality*.

Closely related to a norm is the concept of a *distance*. A distance concept allows us to specify, in a precise sense, 'how far apart' two vectors are.

Definition 2.2.8. *Let V be a vector space over $\mathbb{K}$. A distance on V is a nondegenerate positive definite symmetric mapping $d : V \times V \longrightarrow \mathbb{R}$ such that for all $u, v, w \in V$, $\lambda \in \mathbb{K}$,*

- $d(v, w) \geq 0$ *and* $d(v, w) = 0$ *if and only if* $v = w$,
- $d(v, w) = d(w, v)$,
- $d(v, w) \leq d(v, u) + d(u, w)$.

A vector space with a distance concept is referred to as a *metric space*. A distance function treats its arguments essentially as *points* in an abstract space. Unlike with a norm, its definition therefore does not necessitate a vector space structure; no reference to vector addition or scalar multiplication is actually made in the axioms above. The concept of distance is therefore more generally applicable to spaces other than vector spaces, although vector spaces are often 'attached' to each point on such a space ('base manifold') in the form of (local) 'tangent spaces', collectively making up a 'tangent bundle' that 'sits over' that space.

We can construct a distance function once we are given a norm as follows:

$$d(v, w) = \|w - v\| . \tag{2.15}$$

Thus every normed space is a metric space. Not every metric space is a normed space, however. You should verify yourself that Eq. (2.15) is indeed a distance.

Furthermore, given an inner product we can derive a norm and therefore a distance, viz.

$$\|v\| = \sqrt{\langle v|v\rangle} , \tag{2.16}$$

i.e. the length of the vector v, whence

$$d(v, w) = \sqrt{\langle w - v|w - v\rangle} . \tag{2.17}$$

Thus an inner product space is both a normed space as well as a metric space. It is left as an exercise to show that Eqs. (2.16–2.17) are good definitions, in the sense that they comply with Definitions 2.2.7–2.2.8.

Caveat. In physics it is often taken for granted that a metric space is an inner product space, in which the inner product induces the norm and distance functions. For this reason an inner product is often referred to as a metric in physics.

In analogy with the geometrical interpretation of the usual inner product in $\mathbb{R}^2$ and $\mathbb{R}^3$ one defines the abstract *angle* $\alpha \in [0, 2\pi)$ between two (nonzero) vectors in general via the formula

$$\langle v|w\rangle = \|v\| \, \|w\| \, \cos\alpha , \tag{2.18}$$

in which the norm is the one induced by the inner product according to Eq. (2.16). In this way you may e.g. compute the angle between two functions. As an instructive example you may want to compute the angles between the function f and the respective functions g_1, g_2, g_3, given by $f(x) = e^x$, $g_0(x) = 1$, $g_1(x) = 1 + x$, respectively $g_2(x) = 1 + x + \frac{1}{2}x^2$, with domain $[0, 1]$, relative to the standard inner product.

Definition 2.2.9. *Let V be an n-dimensional normed space with norm $\| \| : V \longrightarrow \mathbb{R}$, and $a \in V$, $r \in \mathbb{R}^+$.*

- *The sphere with midpoint a and radius r is the set* $\partial\Omega(a,r)=\{v\in V \mid \|v-a\|=r\}$.
- *The open ball with midpoint a and radius r is the set* $\Omega(a,r)=\{v\in V \mid \|v-a\|<r\}$.
- *The closed ball with midpoint a and radius r is the set* $\overline{\Omega}(a,r)=\{v\in V \mid \|v-a\|\leq r\}=\Omega(a,r)\cup\partial\Omega(a,r)$.

A sphere in $n=2$ dimensions is usually called a circle, whereas a ball is usually called a disk. In $n=1$ a sphere degenerates to a pair of points, and a ball becomes an interval.

Example 2.2.8. If (x,y) and (u,v) are two points in $\mathbb{R}^2$, then you may check yourself that

- $d((x,y),(u,v))=\sqrt{(u-x)^2+(v-y)^2}$, and
- $d((x,y),(u,v))=|u-x|+|v-y|$

both define instances of distance functions on $\mathbb{R}^2$ by verification of the defining properties in Definition 2.2.8. A sphere in the sense of the first ('Euclidean') distance measure—our 'usual' notion of distance—has the shape of a circle. However, relative to the second ('Manhatten') distance measure a sphere is actually a square!

An important class of norms in image analysis is given by the so-called *p-norms*, $p\geq 1$, defined as follows.

Definition 2.2.10. *Let $f:\Omega\subset\mathbb{Z}^n\longrightarrow\mathbb{K}$ be a discrete n-dimensional image defined on the n-dimensional regular grid $\Omega\subset\mathbb{Z}^n$. Then its p-norm is defined as*

$$\|f\|_p=\left[\sum_{\alpha}|f(\alpha)|^p\right]^{1/p}\qquad p\geq 1\,,$$

in which $\alpha\in\Omega$ is the multi-index coordinate of a grid point. For $p=\infty$ we define the sup-norm as

$$\|f\|_\infty=\sup_{\alpha\in\Omega}|f(\alpha)|\,,$$

where the supremum is taken over all pixels $\alpha\in\Omega$. For images $f:\Omega\subset\mathbb{R}^n\longrightarrow\mathbb{K}$ on a continuous domain we define these norms as

$$\|f\|_p=\left[\int_\Omega|f(x)|^p\,dx\right]^{1/p},$$

respectively

$$\|f\|_\infty=\operatorname{ess\,sup}_{x\in\Omega}|f(x)|\,.$$

The most important instances in practice are those for $p=1$, $p=2$, and $p=\infty$.

Example 2.2.9. Let $s=\{0,i,2i,-6,2,1,0\}$ and $f=\{\{0,1,0\},\{1,-4,1\},\{0,1,0\}\}$. Then you may verify (no calculator required) that $\|s\|_p$ and $\|f\|_p$ for $p=1,2,\infty$ are as follows:

$$\|s\|_1=12\,,\quad \|s\|_2=\sqrt{46}\approx 6.78\,,\quad \|s\|_\infty=6\,,\quad \|f\|_1=8\,,\quad \|f\|_2=2\sqrt{5}\approx 4.47\,,\quad \|f\|_\infty=4\,. \tag{2.19}$$

Recall that in the discrete case the grid may even consist of infinitely many grid points, provided the discrete functions considered are admissible in the sense that their norm defining series converge. In the continuous case the functions must be such that their norm defining integrals exist. The space of functions with domain $\Omega\subset\mathbb{R}^n$ for which the p-norm is well-defined is denoted by $L^p(\Omega)$.

In fact, the normed space $L^p(\Omega)$ is an instance of a so-called *Banach space*, i.e. it is *complete*. This means that every Cauchy sequence in $L^p(\Omega)$—i.e. a sequence $\{f_n\}_{n\in\mathbb{N}}$ such that $\|f_{n+m}-f_n\|\longrightarrow 0$ as $n\longrightarrow\infty$ for every $m\geq 0$—converges to an element in $L^p(\Omega)$. Banach spaces are beyond the scope of these course notes.

We have the following important result, known as the *Hölder inequality*.

Theorem 2.2.4 (Hölder Inequality)**.** *Let* $1 \leq p, q \leq \infty$ *such that* $1/p + 1/q = 1$*, and let* $f \in L^p(\Omega)$ *and* $g \in L^q(\Omega)$*, then*

$$\|fg\|_1 \leq \|f\|_p \, \|g\|_q \, .$$

The pair (p, q) in Theorem 2.2.4 is referred to as a *Hölder conjugate pair*, and the entries p and q are said to be eachother's *Hölder conjugates*. By formal definition, ∞ ('not-a-number') is admitted as the Hölder conjugate of 1, vice versa, by defining $1/\infty = 0$. Note that 2 is the only number that is the Hölder conjugate of itself. A special case is therefore the so-called *Schwartz inequality*.

Theorem 2.2.5 (Schwartz Inequality)**.** *Let* $f, g \in L^2(\Omega)$*, then*

$$\|fg\|_1 \leq \|f\|_2 \, \|g\|_2 \, .$$

In view of its reciprocal role in the Hölder inequality, the Hölder index value $p = 2$ can be seen as lying 'halfway' in-between the instances $p = 1$ and $p = \infty$.

Example 2.2.10. To familiarize yourself with these theorems you may want to verify the following two claims:

1. The Hölder inequality can be used to show that for an image $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{K}$ with a finite domain volume, $V = \mathrm{Vol}\,\Omega = \int_\Omega dx$ say, we have $\|f\|_1 \leq V \, \|f\|_\infty$, provided $\|f\|_\infty$ exists.
2. The Schwartz inequality can be invoked to prove the existence of the integral

$$\langle f|g\rangle_{H^1} = \int_{-\infty}^{\infty} \left(f(x)^* \, g(x) + f'(x)^* \, g'(x)\right) \, dx \, ,$$

   recall Eq. (2.14).

The *inclusion theorem* tells us how all L^p-type function spaces are nested, at least for compact domains.

Theorem 2.2.6 (Inclusion Theorem)**.** *For compact* $\Omega \in \mathbb{R}^n$ *we have*

$$L^q(\Omega) \subset L^p(\Omega) \quad \textit{whenever} \quad 1 \leq p \leq q \, .$$

That is, $L^1(\Omega)$ is the largest, $L^\infty(\Omega)$ the smallest of all L^p-type spaces. Theorem 2.2.6 holds for compact domains only. A compact space is a normed space in which all points lie within some fixed distance of each other. A closed and bounded subset of $\mathbb{R}^n$ (equipped with its standard norm) provides an example of a compact space.

2.3. Linear Operators

2.3.1. Linear Operators in General

Definition 2.3.1. *A linear operator* $A : V \longrightarrow W$ *between the vector spaces* V *and* W *over* $\mathbb{K}$ *is a mapping that satisfies the following property:* $A(\lambda \, v + \mu \, w) = \lambda \, A(v) + \mu \, A(w)$ *for all* $\lambda, \mu \in \mathbb{K}$ *and* $v, w \in V$*.*

In other words, linear combinations of vectors are mapped to corresponding linear combinations of their images.

Linear operators are often referred to simply as 'operators'. A special case arises if $W = \mathbb{C}$ or $\mathbb{R}$. In that case one often refers to the linear operator as a *linear functional* (or simply functional), or as a *covector*. We may interpret a linear functional or covector as a machine with one empty slot into which we can drop a vector so as to produce a number in a linear fashion.

Example 2.3.1. Verify the following claims yourself.

- Let V be an inner product space, and $a \in V$ fixed. Then $L_a : V \longrightarrow \mathbb{K} : v \mapsto L_a(v) = \langle a|v\rangle$ is a covector.
- Let $f \in L^1((0,1))$ fixed. Then $L_f : L^\infty((0,1)) \longrightarrow \mathbb{K} : \phi \mapsto L_f(\phi) = \int_0^1 f(x)\,\phi(x)\,dx$ is a linear functional. (Hint: Recall Theorem 2.2.4 and first show that the integral is well-defined.)
- Let $x \in \mathbb{R}$ be a fixed point. Then the 'point evaluation map' $\delta_x : C(\mathbb{R}) \longrightarrow \mathbb{K} : \phi \mapsto \delta_x(\phi) = \phi(x)$, is a linear functional, known as the *Dirac point distribution*.

Definition 2.3.2. *The collection of all linear operators $V \longrightarrow W$ is denoted by $\mathscr{L}(V,W)$:*

$$\mathscr{L}(V,W) \stackrel{\text{def}}{=} \{A : V \longrightarrow W \mid A \text{ is linear}\}\ .$$

For $A, B \in \mathscr{L}(V,W)$ and $\lambda, \mu \in \mathbb{K}$ we define the linear combination $\lambda\,A + \mu\,B \in \mathscr{L}(V,W)$ as follows:

$$(\lambda\,A + \mu\,B)(v) \stackrel{\text{def}}{=} \lambda\,A(v) + \mu\,B(v) \quad \textit{for all } v \in V.$$

Theorem 2.3.1. *Recall Definition 2.3.2: $\mathscr{L}(V,W)$ is a vector space.*

The proof of Theorem 2.3.1 is left as an exercise.

In the special case that $V = W$ we may concatenate linear transformations, i.e. if $A, B \in \mathscr{L}(V,V)$, then $A \circ B$ is the linear transformation obtained by first applying B, then A. Note that this combined transformation is generally different from $B \circ A$.

Caveat. Linear transformations do not commute in general.

More generally, if $A \in \mathscr{L}(V,X)$, $B \in \mathscr{L}(X,W)$, then $B \circ A \in \mathscr{L}(V,W)$, but $A \circ B$ does not exist, unless the range of B happens to be a subspace of V.

Definition 2.3.3. *If $A, B \in \mathscr{L}(V,V)$, then their commutator is defined as $[A,B] \stackrel{\text{def}}{=} A \circ B - B \circ A$.*

In particular we may repeatedly apply a linear transformation $A \in \mathscr{L}(V,V)$ so as to obtain $A \circ \ldots \circ A$, k times, say. One usually omits the concatenation symbol $\circ$ and simply writes AB, BA, A^k, et cetera.

Example 2.3.2. You should verify yourself that the following claims hold.

- If $A, B \in \mathscr{L}(V,V)$, then both AB as well as BA are linear transformations.
- Let $A \in \mathscr{L}(V,V)$, $I \in \mathscr{L}(V,V)$ the identity, and $0 \in \mathscr{L}(V,V)$ the null map. Then $[A,A] = [A,I] = 0$.
- For all $A, B \in \mathscr{L}(V,V)$ we have $[A,B] = -[B,A]$.
- The mapping $\mathscr{D} : C^{k+1}(\mathbb{R}) \longrightarrow C^k(\mathbb{R}) : s \mapsto \dot{s}$, in which $\dot{s}$ denotes the derivative of s, is a linear operator.
- Let $f \in L^1(\mathbb{R})$ and $\phi \in C^\infty(\mathbb{R}) \cap L^\infty(\mathbb{R})$, then the map

$$F : C^\infty(\mathbb{R}) \cap L^\infty(\mathbb{R}) \longrightarrow \mathbb{R} : \phi \mapsto F(\phi) = (f\,|\phi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(x)\,\phi(x)\,dx$$

  is a well-defined linear map. (Why is this *not* an inner product? Recall Eq. (2.10).)

Definition 2.3.4. *Let V and W be inner products spaces with inner products denoted by $\langle\ |\ \rangle_V$, respectively $\langle\ |\ \rangle_W$, and $A \in \mathscr{L}(V,W)$. Then the adjoint operator $A^\dagger \in \mathscr{L}(W,V)$ is defined as the linear operator for which $\langle A^\dagger w|v\rangle_V = \langle w|Av\rangle_W$ for all $v \in V, w \in W$.*

Theorem 2.3.2. *Let $A \in \mathscr{L}(V,W)$. Then $A^\dagger \in \mathscr{L}(W,V)$. Furthermore, if $\lambda, \mu \in \mathbb{K}$, then we have $(\lambda\, A + \mu\, B)^\dagger = \lambda^*\, A^\dagger + \mu^*\, B^\dagger$, $I^\dagger = I$, $(AB)^\dagger = B^\dagger A^\dagger$, and, if A is invertible, $(A^{\text{inv}})^\dagger = A^{\dagger\,\text{inv}}$.*

Proof. The proof is left as an exercise to the reader.

Definition 2.3.5. *Let $A \in \mathscr{L}(V,V)$. If $A^\dagger = A$, then A is called self-adjoint.*

Definition 2.3.6. *Let $A \in \mathscr{L}(V,V)$. If $A^{\text{inv}} = A^\dagger$, then A is called unitary.*

In the context of vector spaces over $\mathbb{R}$ instead of $\mathbb{C}$ one usually writes A^{T} instead of $A^\dagger$, and refers to it as the *transposed operator*. If $A^{\text{T}} = A$, then A is called *symmetric*, and if $A^{\text{inv}} = A^{\text{T}}$, then it is called *orthogonal*.

2.3.2. Matrix Representations of Linear Operators

In the finite-dimensional case linear operators $A \in \mathscr{L}(V,W)$, with V and W inner product spaces, are most conveniently characterized if we are in possession of an orthonormal basis for V and W, $\{e_1,\ldots,e_M\} \subset V$, respectively $\{f_1,\ldots,f_N\} \subset W$, say. In that case we have the unique decompositions $v = \sum_{i=1}^M v_i\, e_i \in V$, respectively $w = \sum_{a=1}^N w_a\, f_a \in W$. Applying the defining rules for a linear operator, Definition 2.3.1, yields

$$A(v) = A(\sum_{i=1}^M v_i\, e_i) = \sum_{i=1}^M v_i\, A(e_i)\,. \tag{2.20}$$

The vectors $A(e_i) \in W$, $i = 1,\ldots,M$, are unique linear combinations of the basis vectors f_a, $a = 1,\ldots,N$, so

$$A(e_i) = \sum_{a=1}^N A_{ai}\, f_a\,. \tag{2.21}$$

We have $A_{ai} = \langle f_a | A(e_i)\rangle_W = \langle A^\dagger(f_a) | e_i\rangle_V$. (Verify this yourself.) Thus if we set $\tilde{v} = A(v)$ for simplicity, then combining the previous results we observe that

$$\tilde{v} \overset{\text{def}}{=} A(v) \overset{\text{def}}{=} \sum_{a=1}^N \tilde{v}_a\, f_a\,, \tag{2.22}$$

in which the components $\tilde{v}_a$ have been defined as

$$\tilde{v}_a = \sum_{i=1}^M A_{ai}\, v_i \quad a = 1,\ldots,N\,. \tag{2.23}$$

Observation 2.3.1. *The components A_{ai} of the A-image of the i-th basis vector $e_i \in \{e_1,\ldots,e_M\}$ of V relative to the basis $\{f_1,\ldots,f_N\}$ of W define an $N \times M$ matrix*

$$\mathbf{A} = \begin{bmatrix} A_{11} & \ldots & A_{1M} \\ & \vdots & \\ A_{N1} & \ldots & A_{NM} \end{bmatrix}.$$

If we collect the components of $v \in V$ and $\tilde{v} \in W$ relative to their respective standard basis into column vectors $\mathbf{v}$ and $\tilde{\mathbf{v}}$ we can use *matrix notation* to rewrite Eq. (2.23):

$$\begin{bmatrix} \tilde{v}_1 \\ \vdots \\ \tilde{v}_N \end{bmatrix} = \begin{bmatrix} A_{11} & \ldots & A_{1M} \\ & \vdots & \\ A_{N1} & \ldots & A_{NM} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_M \end{bmatrix} \qquad \text{or} \qquad \tilde{\mathbf{v}} = \mathbf{A}\,\mathbf{v}\,. \tag{2.24}$$

That is, A_{ai} is the entry in the a-th row and i-th column of the matrix $\mathbf{A}$. We will refer to Eq. (2.23) as the *index notation* corresponding to Eq. (2.24).

Example 2.3.3. Verify the following the matrix representation of a linear derivative operator relative to a basis:

- Let $V = \operatorname{span}\{s_0, s_1, s_2\}$ with $s_0 : [0,1] \to \mathbb{R} : x \mapsto s_0(x) = 1$, $s_1 : [0,1] \to \mathbb{R} : x \mapsto s_1(x) = \sqrt{3}(2x-1)$ and $s_2 : [0,1] \to \mathbb{R} : x \mapsto s_2(x) = \sqrt{5}(6x^2 - 6x + 1)$, endowed with an inner product as given by Eq. (2.10). Let $\mathscr{D} : V \longrightarrow V : s \mapsto \dot{s}$ be the derivative map on V (note that V is closed under $\mathscr{D}$). Then $\{s_0, s_1, s_2\}$ constitutes an orthonormal basis of V. The matrix representation $\mathbf{D}$ of $\mathscr{D}$ relative to this basis is given by

$$\mathbf{D} = \begin{bmatrix} 0 & 2\sqrt{3} & 0 \\ 0 & 0 & 2\sqrt{15} \\ 0 & 0 & 0 \end{bmatrix}.$$

Theorem 2.3.3. *Suppose* $\mathbf{A}$ *is the* $N \times M$ *matrix corresponding to a linear transformation* $A \in \mathscr{L}(V, W)$ *and* $\mathbf{A}^\dagger$ *is the matrix corresponding to the adjoint linear transformation* $A^\dagger \in \mathscr{L}(W, V)$*. If the entries of* $\mathbf{A}$ *are given by* A_{ai}*,* $a = 1, \ldots, N$*,* $i = 1, \ldots, M$*, then the adjoint matrix* $\mathbf{A}^\dagger$ *is the* $M \times N$ *matrix with entries* $A^\dagger_{ia} = A^*_{ai}$*:*

$$\mathbf{A} = \begin{bmatrix} A_{11} & \ldots & A_{1M} \\ & \vdots & \\ A_{N1} & \ldots & A_{NM} \end{bmatrix} \quad \Longleftrightarrow \quad \mathbf{A}^\dagger = \begin{bmatrix} A^*_{11} & \ldots & A^*_{N1} \\ & \vdots & \\ A^*_{1M} & \ldots & A^*_{NM} \end{bmatrix}.$$

In particular, if V is a real vector space, then A^{T} has matrix representation $\mathbf{A}^{\mathrm{T}}$ with $A^{\mathrm{T}}_{ia} = A_{ai}$.

Remark. Remember that a matrix representation of a linear transformation always requires a pair of bases relative to which the original and mapped vectors are decomposed. Such a choice of bases for V and W is implicit in Theorem 2.3.3. It is left as an exercise to prove Theorem 2.3.3.

Caveat. Never confuse a linear transformation with its matrix representation.

Eq. (2.21) leads to the following, important interpretation of the columns in a matrix.

Result 2.3.1. *The* i*-th column of a matrix* $\mathbf{A}$ *associated with a linear map* $A : V \longrightarrow W$ *contains the coefficients of the* A*-images of the* i*-th basis vector of* V *relative to the basis of* W*.*

The coordinates of the i-th basis vector e_i relative to the basis $\{e_1, \ldots, e_M\} \subset V$ are all 0 except for a 1 at i-th position. Written in explicit matrix notation, Eq. (2.21) therefore becomes

$$\begin{bmatrix} A_{11} & \ldots & A_{1i} & \ldots & A_{1M} \\ & & \vdots & & \\ A_{N1} & \ldots & A_{Ni} & \ldots & A_{NM} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} A_{1i} \\ \vdots \\ A_{Ni} \end{bmatrix},$$

which indeed reproduces the coefficients of $A(e_i) = \sum_{a=1}^{N} A_{ai}\, f_a$ relative to the basis $\{f_1, \ldots, f_N\} \subset W$.

2.3.3. Projections

Projection maps constitute an important class of linear operators. Their defining property is *idempotency*.

Definition 2.3.7. *Recall Definition 2.3.1. Let V be a vector space and $W \subset V$ a subspace. A projection $P : V \longrightarrow W$ is a linear idempotent mapping, i.e. $P^2 = P$.*

If V is an inner product space, then one often considers *orthogonal projections*.

Definition 2.3.8. *Recall Definition 2.3.7. Let V be a vector space and $W \subset V$ a subspace. An orthogonal projection $P : V \longrightarrow W$ is a self-adjoint linear idempotent mapping, i.e. $P^2 = P$ and $P^\dagger = P$.*

We illustrate their importance by introducing a systematic procedure for obtaining an orthonormal basis for an inner product space given an arbitrary basis.

Theorem 2.3.4 (Gram-Schmidt)**.** *Let V be an inner product space with basis $\{v_1, \ldots, v_N\} \subset V$. Then an orthonormal basis $\{e_1, \ldots, e_N\} \subset V$ can be constructed as follows. Let*

$$e'_i = v_i - \sum_{k=1}^{i-1} \frac{\langle e'_k | v_i \rangle}{\langle e'_k | e'_k \rangle} \, e'_k \,,$$

for some index $i = 1, \ldots, N$. Then $\{e'_1, \ldots, e'_N\} \subset V$ is an orthogonal basis of V. Consequently, renormalizing

$$e_i = \frac{1}{\sqrt{\langle e'_i | e'_i \rangle}} \, e'_i \quad (i = 1, \ldots, N) \,,$$

will render the set $\{e_1, \ldots, e_N\} \subset V$ orthonormal.

Proof. Suppose $\{e'_1, \ldots, e'_{i-1}\}$ is orthogonal, i.e. $\langle e'_k | e'_j \rangle = c_k \delta_{kj}$ for $j, k = 1, \ldots, i-1$ and $c_k \in \mathbb{R}$. Then

$$\langle e'_i | e'_j \rangle = \langle v_i - \sum_{k=1}^{i-1} \frac{\langle e'_k | v_i \rangle}{\langle e'_k | e'_k \rangle} \, e'_k | e'_j \rangle \overset{*}{=} \langle v_i | e'_j \rangle - \sum_{k=1}^{i-1} \left(\frac{\langle e'_k | v_i \rangle}{\langle e'_k | e'_k \rangle} \right)^* \langle e'_k | e'_j \rangle \overset{\star}{=} \langle v_i | e'_j \rangle - \langle e'_j | v_i \rangle^* \overset{*}{=} 0 \,.$$

The identities marked by $*$ rely on inner product axioms, the one marked by $\star$ uses in addition the orthogonality assumption above. Note that c_k is strictly real (why cannot it have an imaginary component?).

Remark. $P_w(v) = \dfrac{\langle w | v \rangle}{\langle w | w \rangle} \, w$ is referred to as the orthogonal projection of $v \in V$ onto $\operatorname{span}\{w\}$.

Theorem 2.3.5. *Let V be a vector space and $W = \operatorname{span}\{w_1, \ldots, w_M\} \subset V$. Let $\{e_1, \ldots, e_N\}$ be a basis of V and $w_j = \sum_{i=1}^N a_{ij} \, e_i$, $j = 1, \ldots, M$. Furthermore, let $\mathbf{A}$ be the matrix of coefficients*

$$\mathbf{A} = \begin{bmatrix} a_{11} & \ldots & a_{1M} \\ & \vdots & \\ a_{N1} & \ldots & a_{NM} \end{bmatrix} .$$

Then the orthogonal projection $P : V \longrightarrow W$ has matrix representation

$$\mathbf{P} = \mathbf{A} \big(\mathbf{A}^\dagger \mathbf{A} \big)^{\text{inv}} \mathbf{A}^\dagger \,.$$

Although the matrix $\mathbf{A}$ depends on the choice of the vectors w_j, the matrix $\mathbf{P}$ actually depends only on $\operatorname{span}\{w_1, \ldots, w_M\}$ and not on the particular choices of w_j. To see this, consider any independent linear combination, say $\widetilde{w}_j = \sum_{k=1}^M V_{kj} w_k$. Note that $\operatorname{span}\{\widetilde{w}_1, \ldots, \widetilde{w}_M\} = \operatorname{span}\{w_1, \ldots, w_M\}$. Proceeding as outlined in Theorem 2.3.5 we consider the coefficients of the new spanning vectors $\widetilde{w}_j$ relative to the basis $\{e_1, \ldots, e_N\}$. These follow

from the definition: $\widetilde{w}_j = \sum_{k=1}^{M} V_{kj} w_k = \sum_{k=1}^{M} V_{kj} \sum_{i=1}^{N} a_{ik} e_i = \sum_{i=1}^{N} \widetilde{a}_{ij} e_i$, so that $\widetilde{a}_{ij} = \sum_{k=1}^{M} a_{ik} V_{kj}$. Collecting these coefficients in a matrix we get $\widetilde{\mathbf{A}} = \mathbf{AV}$, where $\mathbf{V}$ is the invertible $M \times M$ matrix

$$\mathbf{V} = \begin{bmatrix} V_{11} & \dots & V_{1M} \\ & \vdots & \\ V_{N1} & \dots & V_{NM} \end{bmatrix}.$$

Thus according to Theorem 2.3.5 the matrix representation of the corresponding orthogonal projection equals

$$\widetilde{\mathbf{P}} = \widetilde{\mathbf{A}} \Big(\widetilde{\mathbf{A}}^\dagger \widetilde{\mathbf{A}} \Big)^{\text{inv}} \widetilde{\mathbf{A}}^\dagger = \mathbf{AV} \Big((\mathbf{AV})^\dagger \mathbf{AV} \Big)^{\text{inv}} (\mathbf{AV})^\dagger .$$

Substituting $(\mathbf{AV})^\dagger = \mathbf{V}^\dagger \mathbf{A}^\dagger$ and $\Big((\mathbf{AV})^\dagger \mathbf{AV} \Big)^{\text{inv}} = \mathbf{V}^{\text{inv}} (\mathbf{A}^\dagger \mathbf{A})^{\text{inv}} \mathbf{V}^{\dagger^{\text{inv}}}$, we obtain

$$\widetilde{\mathbf{P}} = \mathbf{A}(\mathbf{V}\mathbf{V}^{\text{inv}})(\mathbf{A}^\dagger \mathbf{A})^{\text{inv}} (\mathbf{V}^{\dagger^{\text{inv}}} \mathbf{V}^\dagger) \mathbf{A}^\dagger = \mathbf{A} \big(\mathbf{A}^\dagger \mathbf{A} \big)^{\text{inv}} \mathbf{A}^\dagger = \mathbf{P} .$$

Example 2.3.4. Consider $W = \operatorname{span}\{w\}$ for a single vector $w \in V$, say $w = \sum_{i=1}^{N} w_i e_i$. We indicate the column vector of components of a vector v relative to the given basis by $\mathbf{v}$. Then

$$\mathbf{A} = \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} \quad \text{and} \quad \mathbf{A}^\dagger = [w_1 \dots w_N]$$

and

$$\big(\mathbf{A}^\dagger \mathbf{A} \big)^{\text{inv}} = \left\{ [w_1 \dots w_N] \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} \right\}^{\text{inv}} = \frac{1}{\langle w | w \rangle} ,$$

so that

$$\mathbf{Pv} = \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} \frac{1}{\langle w | w \rangle} [w_1 \dots w_N] \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix} = \frac{\langle w | v \rangle}{\langle w | w \rangle} \mathbf{w} ,$$

in which

$$\mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} .$$

This is the column representation equivalent of $P_w(v)$ presented in the remark on page 33.

Example 2.3.5. Recall Theorem 2.3.5. Verify yourself that projection matrices corresponding to orthogonal projections relative to any given basis are symmetric and idempotent. Furthermore, using the definition of the projection map P_w on page 33, likewise verify that if $\{e_1, \dots, e_N\} \subset V$ is an orthonormal basis of V, then $\sum_{i=1}^{N} P_{e_i} = I$, i.e. the identity map on V.

2.4. Algebras

It is often desirable to combine signals and images in other ways than by vector addition or scalar multiplication. For instance we would like to be able to (somehow) *multiply* them. This calls for an *algebraic structure*.

Definition 2.4.1. *An algebra $\mathscr{A}$ over the field $\mathbb{K}$ is a set endowed with two internal operations and an external operation, usually called 'addition', 'multiplication' (or 'concatenation') and 'scalar multiplication', respectively. Addition and scalar multiplication obey the usual vector space axioms. The requirements for multiplication are as follows. Denoting the infix multiplication operator by $\circ$, we have, for all $a, b, c \in \mathscr{A}$:*

$$\begin{aligned} a \circ b &\in \mathscr{A} \\ (a \circ b) \circ c &= a \circ (b \circ c)\,, \\ a \circ (b + c) &= a \circ b + a \circ c\,, \\ (a + b) \circ c &= a \circ c + b \circ c\,. \end{aligned}$$

In other words, $\circ$ is closed ('internal'), associative, and distributive with respect to addition. Moreover, scalar multiplication must be such that for all $a, b \in \mathscr{A}$ and $\lambda \in \mathbb{K}$,

$$\lambda(a \circ b) = (\lambda a) \circ b = a \circ (\lambda b)\,.$$

An algebra does not necessarily have an identity element $e \in \mathscr{A}$ such that $e \circ a = a \circ e = a$ for all $a \in \mathscr{A}$. If it does, it is called an algebra with identity. If in addition every nonzero element $a \in \mathscr{A}$ has an inverse $a^{-1} \in \mathscr{A}$, meaning $a \circ a^{-1} = a^{-1} \circ a = e$, one calls it a regular algebra. A singular algebra is one in which we cannot invert all nonzero elements. Multiplication is generally not required to commute. If $a \circ b = b \circ a$ for all $a, b \in \mathscr{A}$, we have a commutative algebra.

Two concrete instances of image algebras will turn out important for future use, the *multiplication algebra* and the *convolution algebra*. For simplicity we will take the image domains to be all of $\mathbb{Z}^n$, respectively $\mathbb{R}^n$.

Example 2.4.1. Let V be a vector space of discrete images, and $f, g \in V$. Then multiplication is defined by

$$(f \times g)(\alpha) = f(\alpha)\, g(\alpha) \quad (\alpha \in \mathbb{Z}^n)\,.$$

Similarly, if V is a vector space of images on a continuous domain, and $f, g \in V$, then

$$(f \times g)(x) = f(x)\, g(x) \quad (x \in \mathbb{R}^n)\,.$$

It is common to write fg instead of $f \times g$, et cetera. This algebraic product is usually referred to simply as 'the product' of f and g.

Example 2.4.2. Let V be a vector space of discrete images, and $f, g \in V$. Then convolution is defined by

$$(f * g)(\alpha) = \sum_{\beta \in \mathbb{Z}^n} f(\beta)\, g(\alpha - \beta) \quad (\alpha \in \mathbb{Z}^n)\,.$$

Here we have assumed that $f(\alpha) = g(\alpha) = 0$ whenever the index exceeds the bounds of the underlying grid so as to make the expression on the right hand side well-defined. Similarly, if V is a vector space of images on a continuous domain, and $f, g \in V$, we define

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y)\, g(x - y)\, dy \quad (x \in \mathbb{R}^n)\,.$$

Again, if the function domain $\Omega \subset \mathbb{R}^n$ is a proper subset of $\mathbb{R}^n$ then we set $f(x) = g(x) = 0$ whenever $x \notin \Omega$ so as to make the right hand side well-defined.

You should verify that both multiplication and convolution of images fulfill the basic algebraic requirements of Definition 2.4.1, and that they are commutative.

3

DIFFERENTIATION AND INTEGRATION

3.1. Integration

3.1.1. Definition via Riemann Sums

Recall the definition of a definite integral from standard calculus. We consider the one-dimensional case first, say

$$\int_a^b f(x)\,dx\,,$$

in which $f \in C([a,b])$ is a real-valued function. Results can be extended to $f \in L^1([a,b])$.

Definition 3.1.1. *Subdivide the interval $[a,b]$ into N intervals of equal length by choosing $N+1$ sample points x_k, $k = 0,\ldots,N$, with $x_0 = a$ and $x_N = b$, such that $\Delta x = x_k - x_{k-1} = (b-a)/N$ independent of k. The set of equidistant points $\{x_k = a + k\,(b-a)/N \mid k = 0,\ldots,N\}$ is called a partition of the interval $[a,b]$, and the subinterval width $x_k - x_{k-1} = \Delta x$ the norm of the partition. In each subinterval choose an arbitrary sample point $x_k^* \in [x_{k-1}, x_k]$, $k = 1,\ldots,N$. In Figure 3.1 the situation has been sketched for the case in which $x_k^* = (x_k - x_{k-1})/2$, i.e. the midpoint of the k-th subinterval. We now define*

$$\int_a^b f(x)\,dx \stackrel{\text{def}}{=} \lim_{\Delta x \downarrow 0} \sum_{k=1}^{N} f(x_k^*)\,\Delta x\,.$$

Because Δx is a constant (in this case) we may put it in front of the summation sign[1]*. Note that in the limit $\Delta x \downarrow 0$ we have $N \longrightarrow \infty$. Without this limit the right hand side of the defining equation in Definition 3.1.1 is called the Riemann sum associated with the partition $\{x_k\}_{k=0,\ldots,N}$ and the representative sample points $\{x_k^*\}_{k=1,\ldots,N}$.*

For a discrete signal defined on a regular grid we approximate the integral by identifying the partition with the sampling grid and by discarding the limit. One often tacitly assumes that $\Delta x = 1$.

Definition 3.1.2. *If x_k^* is the point in the interval $[x_{k-1}, x_k]$ at which f takes its maximum (minimum), the associated Riemann sum is called the upper Riemann sum (respectively lower Riemann sum).*

Upper and lower Riemann sums provide useful bounds for the exact integral. The conjecture is that in the limit of vanishing Δx both coincide, and are equal to the limit of any other Riemann sum regardless of the partition.

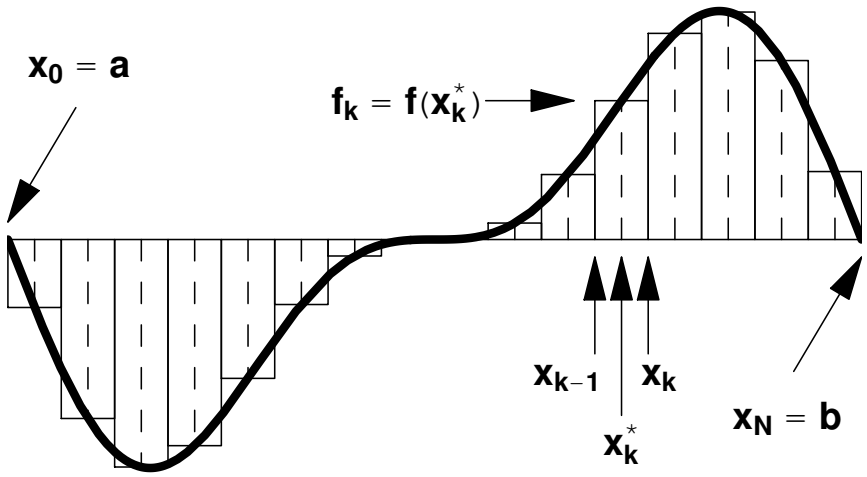


Figure 3.1: Approxiomation scheme for one-dimensional definite integral by a Riemann sum.

Thus, as Figure 3.1 illustrates, the Riemann sum approximation is based on summing areas of narrow rectangles.

Definition 3.1.3. *One may instead consider the areas enclosed by the curves*

$$\gamma_k = \{(x_{k-1}, 0), (x_{k-1}, f(x_{k-1})), (x_k, f(x_k)), (x_k, 0), (x_{k-1}, 0)\} \qquad k = 1, \ldots, N\,,$$

so as to obtain the so-called trapezoidal approximation

$$\int_a^b f(x)\,dx \approx \sum_{k=1}^{N} \frac{f(x_{k-1}) + f(x_k)}{2}\,\Delta x\,.$$

You may want to sketch the situation in a plot similar to Figure 3.1 and explain this result as well as its name.

Definition 3.1.1 can be generalized to multiple dimensions. For simplicity we first consider integrals of the type

$$\int_{a_n}^{b_n} \ldots \int_{a_1}^{b_1} f(x^1, \ldots, x^n)\,dx^1 \ldots dx^n\,,$$

in which $f : [a_1, b_1] \times \ldots \times [a_n, b_n] \longrightarrow \mathbb{R}$ is an image defined on a rectangular domain. Again we assume $f \in C([a_1, b_1] \times \ldots \times [a_n, b_n])$, although results can be extended to $f \in L^1([a_1, b_1] \times \ldots \times [a_n, b_n])$.

[1] One can generalize the argument to non-equidistant points. The norm of the partition is then the longest subinterval width.

Definition 3.1.4. *For each $i = 1, \ldots, n$, subdivide the interval $[a_i, b_i]$ into N_i intervals of equal length Δx^i by choosing $N_i + 1$ sample points $x^i_{k_i}$, $k_i = 0, \ldots, N_i$, with $x^i_0 = a_i$ and $x^i_N = b_i$, such that $\Delta x^i = x^i_{k_i} - x^i_{k_i - 1} = (b_i - a_i)/N_i$ independent of k_i. Choose an arbitrary sample point $x^{i\,*}_{k_i} \in [x^i_{k_i - 1}, x^i_{k_i}]$. Then we define*

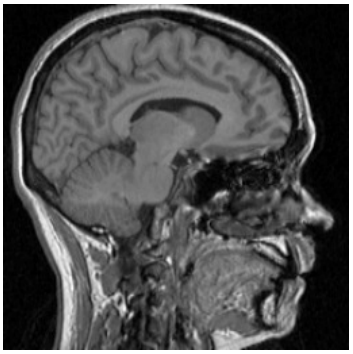
$$\int_{a_n}^{b_n} \ldots \int_{a_1}^{b_1} f(x^1, \ldots, x^n)\, dx^1 \ldots dx^n \stackrel{\text{def}}{=} \lim_{\Delta V \downarrow 0} \sum_{k_1=1}^{N_1} \ldots \sum_{k_n=1}^{N_n} f(x^{1\,*}_{k_1}, \ldots, x^{n\,*}_{k_n})\, \Delta V\,,$$

in which $\Delta V = \Delta x^1 \ldots \Delta x^n$.

Because the volume element ΔV is a constant for a regular grid we may put it in front of the summation sign.

Definition 3.1.5. *Recall Definition 3.1.4. Then upper Riemann sums and lower Riemann sums are obtained by choosing the sample points $(x^{1\,*}_{k_1}, \ldots, x^{n\,*}_{k_n})$ such that $f(x^{1\,*}_{k_1}, \ldots, x^{n\,*}_{k_n})$ is the maximum, respectively minimum of f on the corresponding blocks $[x^1_{k_1 - 1}, x^1_{k_1}] \times \ldots \times [x^n_{k_n - 1}, x^n_{k_n}]$.*

Figure 3.2 shows an approximation of the integral of a two-dimensional MRI image in terms of a Riemann sum with 64 terms. (Note that each term is itself the outcome of a Riemann sum!) This approximation can be improved by refining the Riemann sum to a sum over the original pixel mesh (65536 terms, one for each pixel), but obviously no further refinement is possible in practice. The sum of all numbers listed in Figure 3.2 equals the Riemann sum evaluated over the original 256×256-grid. (Why?)



3	50	85	90	73	48	13	3
35	80	101	105	108	95	64	3
61	104	110	93	98	110	90	23
74	103	90	113	83	71	56	17
61	88	110	94	57	48	74	41
21	135	87	84	95	89	94	27
4	103	105	69	101	113	77	8
6	95	80	58	89	116	71	4

Figure 3.2: Approximation scheme for two-dimensional definite integral by a Riemann sum. Left: 256×256 MRI image. Right: A coarse Riemann sum for an 8×8 grid can be obtained as the sum of all pixel values of a correspondingly downsampled version of the image. In this case each pixel shown on the right is the average over a square of 32×32 pixels in the original image on the left.

Definition 3.1.6. *Recall Definition 3.1.4. If the integration domain is not rectangular, but instead occupies some arbitrarily shaped, simply connected (and sufficiently decent) volume $\Omega \subset \mathbb{R}^n$, then Definition 3.1.4 applies after multiplication of the integrand by the indicator function on $\Omega \subset \mathbb{R}^n$:*

$$\chi_\Omega(x^1, \ldots, x^n) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } (x^1, \ldots, x^n) \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

That is,

$$\int_\Omega f(x)\, dx \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} f(x)\, \chi_\Omega(x)\, dx\,.$$

If Ω is a finite region, then the sum will effectively contain a finite number of terms. If not (e.g. if $\Omega = \mathbb{R}^n$), then additional restrictions apply to the type of functions that are allowed in the integrand (e.g. $f \in L^1(\mathbb{R}^n)$).

The integrals we have encountered previously and will elaborate on further are mostly of the type covered by Definition 3.1.4 with $n = 2$ or $n = 3$, and occasionally with $n = 1$ (e.g. signals, or 'maximum intensity

projection' of images) or $n = 4$ (movies of volume data). Other types of integrals are also of interest in image analysis, such as line integrals and surface integrals of scalar and vector images. In this context the famous integration theorems, such as Stokes' Theorem and related ones (Green, Gauss), come into play. Suffice it to say that the Riemann sum approximation scheme discussed here provides the basic insight needed in order to appreciate integrals of general type.

3.1.2. Convolution and Correlation

The *convolution product* (recall the definition on page 35) and the *correlation product* are bilinear operations closely related to the inner product defined for images in Section 2.2.2. However, as opposed to the sesquilinear inner product, these are defined to be bilinear also in the complex case:

Definition 3.1.7. *Let V be a vector space of discrete images, and $f, g \in V$. The discrete correlation product of f and g is the discrete image $f \star g$ defined by*

$$(f \star g)(\alpha) = \sum_{\beta \in \mathbb{Z}^n} f(\beta)\, g(\beta - \alpha) \quad (\alpha \in \mathbb{Z}^n)\,.$$

The discrete convolution product of f and g is the discrete image $f * g$ defined by

$$(f * g)(\alpha) = \sum_{\beta \in \mathbb{Z}^n} f(\beta)\, g(\alpha - \beta) \quad (\alpha \in \mathbb{Z}^n)\,.$$

Here we assume that $f(\alpha) = g(\alpha) = 0$ whenever the multi-index exceeds the bounds of the actual grid so as to make the right hand sides well-defined.

Definition 3.1.8. *Let V be a vector space of images on a continuous domain, and $f, g \in V$. Then we define the continuous correlation product as the image $f \star g$ defined by*

$$(f \star g)(x) = \int_{\mathbb{R}^n} f(y)\, g(y - x)\, dy \quad (x \in \mathbb{R}^n)\,.$$

Finally, the continuous convolution product is the image $f * g$ defined by

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y)\, g(x - y)\, dy \quad (x \in \mathbb{R}^n)\,.$$

If the function domain $\Omega \subset \mathbb{R}^n$ is a proper subset of $\mathbb{R}^n$ then we set $f(x) = g(x) = 0$ whenever $x \notin \Omega$ so as to make the right hand sides well-defined.

Thus the difference between correlation and convolution lies in a relative minus sign. The implication of this is that the correlation product fails to generate an algebra, as opposed to the convolution product. Also, the convolution product is commutative, the correlation product is not. You should verify these claims yourself by proving the following claims: $(f * g) * h = f * (g * h)$, $(f \star g) \star h = f \star (g * h)$, $f * g = g * f$, and, for all $x \in \mathbb{R}^n$, $(f \star g)(x) = (g \star f)(-x)$.

The similarity between convolution and correlation is formalized by the following observation:

Observation 3.1.1. *Let $g : \mathbb{R}^n \longrightarrow \mathbb{K}$ be a suitably defined function, and $\check{g}(x) = g(-x)$ for all $x \in \mathbb{R}^n$. Then*

$$(f \star g)(x) = (f * \check{g})(x) \quad \textit{and} \quad (g \star f)(x) = (f * \check{g})(-x)\,.$$

Caveat. In the image literature different definitions coexist for correlation and convolution.

In a convolution or correlation product of two functions, one typically represents an image, the other a 'filter', i.e. a suitably constructed function for the purpose of image processing. In practice one has to deal with the 'boundary problem' that arises as soon as points outside the actual image/filter domain are referenced.

Example 3.1.1. Figure 3.3 shows a symmetric, 3×3 filter g. In filter plots such as this it is tacitly understood that the origin of the coordinate system is *in the middle* of the image. Figure 3.4 and 3.5 illustrate how correlation is performed between an 8×8 image f and this filter. The symmetry in each direction causes convolution and correlation to coincide. The question marks indicate boundary pixels at which correlation (or convolution) cannot be calculated without additional assumptions.

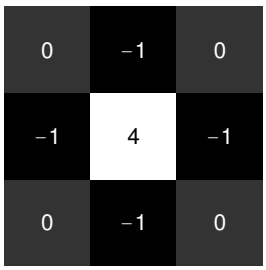


Figure 3.3: Symmetric 3×3 filter mask. The coordinate origin is in the middle.

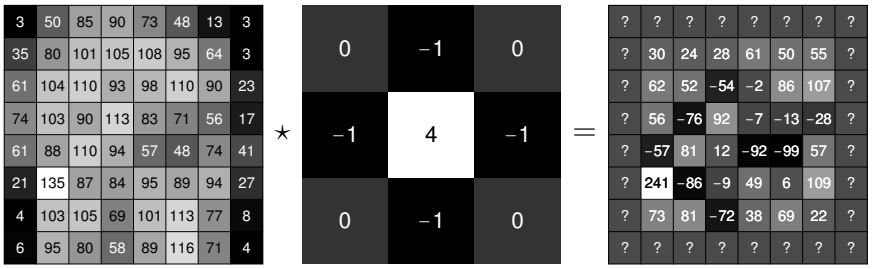


Figure 3.4: Principle of discrete correlation. From left to right: Discrete 8×8 image labelled with pixel values, 3×3 filter mask, and correlation product.

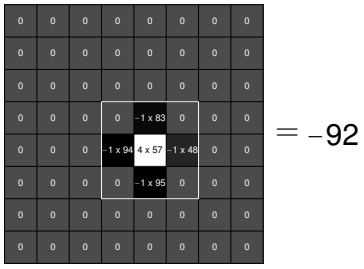


Figure 3.5: Product of image and filter mask, in this example centered at the pixel $(5, 4)$ relative to the lower left corner, with value 57, cf. Figure 3.4. In this image it can be seen how the correlation product at pixel $(5, 4)$ is evaluated, viz. $-1 \times 83 - 1 \times 94 + 4 \times 57 - 1 \times 48 - 1 \times 95 = -92$.

If, unlike Figure 3.4, the filter is not symmetric in all directions, then convolution and correlation will in general not be identical. Consider Figure 3.6 as an example. Figures 3.7 and 3.8 illustrate the correlation and convolution products, with the boundary effect explicitly indicated, and the way of constructing their respective values at the same pixel of interest as before.

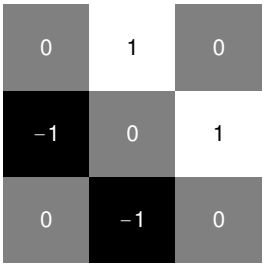


Figure 3.6: Non-symmetric 3×3 filter mask. Again, the coordinate origin is in the middle.

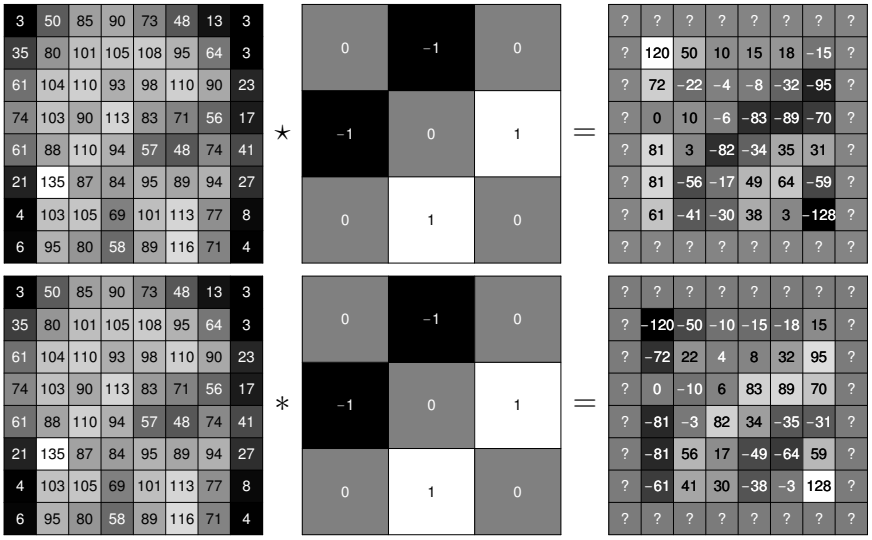


Figure 3.7: Correlation versus convolution. Top row, from left to right: Discrete 8×8 image labelled with pixel values, 3×3 correlation filter mask, and correlation product. Bottom row, from left to right: Same as first row, but now for convolution instead of correlation.

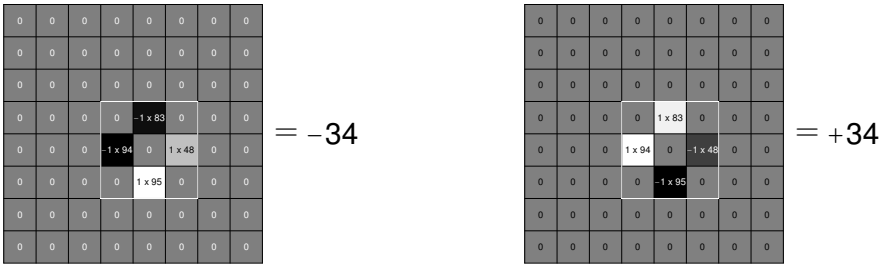


Figure 3.8: Product of image and filter mask, again centred at pixel $(5, 4)$ with value 57, cf. Figure 3.7. The left image illustrates the case of correlation. It shows how the correlation product at the pixel of interest is evaluated, viz. $-1 \times 83 - 1 \times 94 + 1 \times 48 + 1 \times 95 = -34$. The right image illustrates the case of convolution. It shows how the convolution product at pixel $(5, 4)$ is evaluated, viz. $1 \times 83 + 1 \times 94 - 1 \times 48 - 1 \times 95 = 34$. Notice the mirroring of the filter mask.

3.2. Classical Differentiation

An important tool in image analysis besides integration is differentiation. We distinguish between *classical differentiation*, pertaining to infinitesimal limiting procedures, and *distributional differentiation*. The decisive issue that will force us to replace classical by distributional differentiation is *well-posedness* [15]. Well-posedness and continuity are close-knit. In essence they boil down to the requirement that small errors in input ('noise') should produce small errors in output. More precisely, *smaller* errors in input should produce *smaller* errors in output, whereby the latter should vanish in the limit whereby the former is made to vanish.

3.2.1. Exact and Numerical Definitions of Classical Differentiation

The primordial example of classical differentiation is the following.

Definition 3.2.1. *Let $f \in C^1(\mathbb{R})$ be a real-valued function. If the limit*

$$f'(x) \stackrel{\text{def}}{=} \lim_{h \longrightarrow 0} \frac{f(x+h) - f(x)}{h}$$

exists, then $f'(x)$ is called the first order derivative of f at x, and f is said to be differentiable at x in classical sense, or simply differentiable at x. A differentiable function is a function that is everywhere differentiable.

To obtain the k-th order derivative of f at x, notation $f^{(k)}(x)$, apply the above procedure k times in succession.

Definition 3.2.2. *The class of k-fold continuously differentiable functions with domain $\mathbb{R}$ is indicated by $C^k(\mathbb{R})$.*

We have the inclusion relation $C^\omega(\mathbb{R}) \subset C^\infty(\mathbb{R}) \subset C^k(\mathbb{R}) \subset C^m(\mathbb{R})$ for all $m = 0, \ldots, k$.

Example 3.2.1. The infinitesimal limiting procedure for obtaining the second order derivative $f''(x)$, assuming $f \in C^2(\mathbb{R})$, is as follows. Set $g(x) = f'(x)$, thus

$$g(x) = \lim_{h \longrightarrow 0} \frac{f(x+h) - f(x)}{h}\,.$$

Then $f''(x) = g'(x)$, i.e.

$$f''(x) = \lim_{h \longrightarrow 0} \frac{g(x+h) - g(x)}{h} = \lim_{h \longrightarrow 0} \lim_{k \longrightarrow 0} \frac{f(x+h+k) - f(x+h) - f(x+k) + f(x)}{kh}\,.$$

We may simplify this by setting $k = h$:

$$f''(x) = \lim_{h \longrightarrow 0} \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2}\,.$$

Alternatively, we may choose to set $k = -h$, in which case we obtain instead

$$f''(x) = \lim_{h \longrightarrow 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}\,.$$

In numerical schemes (dropping the limiting procedure) the latter 'central difference quotient' is often preferred in view of its symmetry w.r.t. the central point x.

The example shows that the asymmetry w.r.t. the point x in the difference quotient in Definition 3.2.1—i.e., again, discarding the limit—tends to propagate to higher orders. A way to obtain central differences directly is to define

$$f'(x) \stackrel{\text{def}}{=} \lim_{h \longrightarrow 0} \frac{f(x+\frac{h}{2}) - f(x-\frac{h}{2})}{h} = \lim_{h \longrightarrow 0} \frac{f(x+h) - f(x-h)}{2h} \tag{3.1}$$

instead of Definition 3.2.1. The central difference quotient on the right allows us to identify h with the grid constant in a straightforward numerical computation. Offsets by fractions of h, such as in the middle, can be numerically handled after interpolation. You may verify that linear interpolation in this case would make both central difference quotients coincide.

For multivariate functions *partial derivatives* may be defined in a similar fashion. If $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is such a function, then in order to take a partial derivative w.r.t. x^μ, say, freeze all function arguments except x^μ and proceed as for the one-dimensional case. In other words, if $g : \mathbb{R} \longrightarrow \mathbb{R}$ is the one-dimensional function defined by $g(y) = f(x^1, \ldots, x^{\mu-1}, y, x^{\mu+1}, \ldots, x^n)$ for fixed $x^1, \ldots, x^{\mu-1}, x^{\mu+1}, \ldots, x^n$, then

$$\frac{\partial f}{\partial x^\mu}(x^1, \ldots, x^n) \stackrel{\text{def}}{=} \frac{dg}{dy}(y)\Big|_{y\,=\,x^\mu} .$$

The class of functions for which this is possible for each $\mu = 1, \ldots, n$ and for which the derivatives are continuous is indicated by $C^1(\mathbb{R}^n)$. The class $C^k(\mathbb{R}^n)$ is defined as the class of functions for which all k-th order partial derivatives exist and are continuous. We have $C^\omega(\mathbb{R}^n) \subset C^\infty(\mathbb{R}^n) \subset C^k(\mathbb{R}^n) \subset C^m(\mathbb{R}^n)$ for all $m = 0, \ldots, k$.

Example 3.2.2. The first order partial derivatives are often collectively denoted using the *gradient*:

$$\nabla f = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\, e_i \,,$$

relative to the basis $\{e_1, \ldots, e_n\}$.

Example 3.2.3. The *Laplacean* is defined as

$$\Delta f = \operatorname{div} \nabla f \,.$$

The *divergence* of a function $g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, with components $g^i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $i = 1, \ldots, n$, is the function

$$\operatorname{div} g = \sum_{i=1}^{n} \frac{\partial g^i}{\partial x^i} \,,$$

so that

$$\Delta f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x^i \partial x^i} \,.$$

Example 3.2.4. The infinitesimal limiting procedure for obtaining the Laplacean of a 2-dimensional image $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ using central differences reads as follows:

$$\Delta f(x, y) = \lim_{h \longrightarrow 0} \frac{f(x+h, y) + f(x, y+h) - 4f(x, y) + f(x-h, y) + f(x, y-h)}{h^2} \,.$$

You should extend this yourself to the n-dimensional case. Another useful exercise is to derive the central difference infinitesimal limiting procedure for the action of the biharmonic operator, $\Delta^2 f(x) \stackrel{\text{def}}{=} \Delta(\Delta f(x))$.

The simplest and most commonly used discretization scheme is obtained by dropping the operationally void $\lim$-operator in front of a suitably chosen discrete difference quotient, taking h in the order of the grid spacing. It is important to realize that difference quotients that converge to the same limit may not be equivalent without the infinitesimal $\lim$-procedure. Central differences are often preferred for symmetry reasons.

Example 3.2.5. How to discretize a first order derivative, $f'(x)$? Dropping the $\lim$-operator in Definition 3.2.1 and assuming that the grid constant equals 1 suggest the following definition:

$$f'(x) \approx f(x+1) - f(x) \,,$$

in which we have set $h \equiv 1$, the grid constant. If we wish to avoid the spatial bias we may instead apply Eq. (3.1), but now we have a dilemma. If we set $h \equiv 1$ while dropping the limiting procedure we get

$$f'(x) \approx f(x + \frac{1}{2}) - f(x - \frac{1}{2}) \,,$$

which requires interpolation of each term, recall Section 1.3. Alternatively we may use the central difference scheme obviating the need for interpolation by choosing $h \equiv 2$:

$$f'(x) \approx \frac{f(x+1) - f(x-1)}{2} \,.$$

Note that the central value $f(x)$ at x now plays no role. You should convince yourself that both options are in fact equivalent if linear interpolation is used in the former case.

Figures 3.9 and 3.10 show the first order partial derivative in horizontal direction on an MR image.

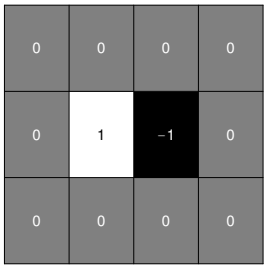


Figure 3.9: Discrete correlation mask for computing first order derivative in horizontal direction. The coordinate origin is in the middle.

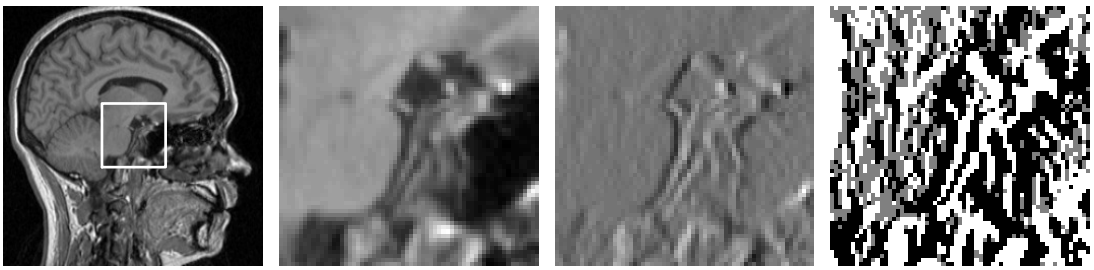


Figure 3.10: Discrete first order derivative in horizontal direction computed with the correlation mask of Figure 3.9. From left to right: Original image with region of interest (ROI), enlargement of the ROI, discrete derivative in horizontal direction of ROI (middle grey corresponds to value zero, black means negative, white means positive), and its sign.

Discrete differentiation actually boils down to discrete correlation (or convolution). For instance, compare the central difference scheme for the Laplacean $\Delta f(x, y)$ in two dimensions with Figures 3.3–3.5. It is left as an exercise to find filter masks g_1, g_2, g_3, g_4, etc. that allow you to rewrite the discrete central difference schemes for $f'(x)$, $f'(x)$, $f''(x)$, $f''(x)$, etc. in terms of correlations, $(f \star g_1)(x)$, $(f \star g_2)(x)$, $(f \star g_3)(x)$, $(f \star g_4)(x)$, and so forth.

3.2.2. Properties of Classical Differentiation

Returning to the continuous domain we can list a number of properties of differential operators. For simplicity we consider $\mathbb{K}$-valued functions of the type $f \in C^\infty(\mathbb{R}^n)$, and use the shorthand notation

$$\partial_i f(x) \stackrel{\text{def}}{=} \frac{\partial f}{\partial x^i}(x) \quad (i = 1, \ldots, n)\,.$$

Observation 3.2.1. *The classical differential operator $\partial_i : C^\infty(\mathbb{R}^n) \longrightarrow C^\infty(\mathbb{R}^n) : f \mapsto \partial_i f$ satisfies the following properties:*

- *∂_i is a linear operator, recall Section 2.3:*

$$\partial_i(\lambda\, f + \mu\, g) = \lambda\, \partial_i f + \mu\, \partial_i g \quad \textit{for all } \lambda, \mu \in \mathbb{K},\ f, g \in C^\infty(\mathbb{R}^n),$$

- *∂_i satisfies the so-called Leibniz' product rule:*

$$\partial_i(f\, g) = \partial_i f\, g + f\, \partial_i g \quad \textit{for all } f, g \in C^\infty(\mathbb{R}^n).$$

Another important property concerns the behavior of partial derivative operators under minor perturbations. An unpleasant surprise awaits us, as the following example illustrates.

Example 3.2.6. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be given by $f(x) = \epsilon \sin(\omega x)$ with $0 < \epsilon \ll 1$ and $\omega \gg \epsilon^{-1}$. This represents a 'negligibly small signal', meaning $\|f\|_\infty = \epsilon \ll 1$. Since $f'(x) = \epsilon\, \omega \cos(\omega x)$, it follows that $\|f'\|_\infty = \epsilon\, \omega \gg 1$.

The example shows that *a negligible amount of additive noise in an input signal randomizes the output under classical differentiation*. Rephrased in mathematical terms [15]:

Observation 3.2.2. *Classical differentiation $\partial_i : C^\infty(\mathbb{R}^n) \longrightarrow C^\infty(\mathbb{R}^n) : f \mapsto \partial_i f$ is ill-posed relative to the $\|\,.\,\|_\infty$-norm.*

The example is not idiosyncratic, for in Chapter 4 we will see that virtually any function of practical interest can be decomposed into a sum of sinusoids of various frequencies. In fact, when we represent an image conceptually as a function f, we implicitly assume that this function has a finite measurement tolerance, so that small perturbations of f—e.g. obtained by repeating the acquisition protocol under similar circumstances—should be reckoned with. Therefore we *must* take into consideration the effect of noise whenever we process the image. Thus when we subject an image f to some operator A, we should always make sure that $A(f + \delta f)$ is close to $A(f)$ for a sufficiently small perturbation δf in some appropriate norm. That is, we should consider only *well-posed* operators. Ill-posedness and well-posedness require us to specify norms for input and output elements, for it is only relative to such norms that it makes sense to talk about 'small' perturbations.

Conclusion 3.2.1. *Ill-posedness forces us to reject classical differentiation as a viable tool for analyzing images.*

3.2.3. Towards a Well-Posed Alternative for Classical Differentiation

In order to arrive at a well-posed alternative for differential operators, the following examples are instructive.

Example 3.2.7. Let $A : C^0([0,1]) \longrightarrow \mathbb{K}$ be the linear integral operator given by

$$A(f) = \int_0^1 f(x)\, dx\,.$$

By default, let the input space $C^0([0,1])$ be endowed with the $\| \, . \, \|_\infty$-norm (sup-norm, cf. Section 2.2.3), and the output space $\mathbb{K}$ with the natural $| \, . \, |$-norm (absolute value). The following estimate reveals the well-posedness of A relative to these norms:

$$|A(f)| = |\int_0^1 f(x)\,dx| \leq \int_0^1 |f(x)|\,dx \leq \max_{x\in[0,1]} |f(x)| \int_0^1 dx = \max_{x\in[0,1]} |f(x)| = \|f\|_\infty \,.$$

(Verify each step in this sequence.) This shows that if we let the input function f tend to zero (as a function), in the sense that $\|f\|_\infty \longrightarrow 0$ (i.e. as a number), then also the output $A(f)$ will tend to zero, in the absolute sense of $|A(f)| \longrightarrow 0$, so that we may conclude that the operator A is well-posed relative to the stated norms. The input space may be extended to $L^\infty([0,1])$.

In fact, this example can be generalized, as the following one shows.

Example 3.2.8. Again, let $A : C^0([0,1]) \longrightarrow \mathbb{K}$ be the linear integral operator given by

$$A(f) = \int_0^1 f(x)\,dx \,.$$

Let the input space $C^0([0,1])$ now be endowed with the $\| \, . \, \|_p$-norm (p-norm, cf. Section 2.2.3), keeping the output norm on $\mathbb{K}$ is as before. The following estimate reveals the well-posedness of A relative to these norms:

$$|A(f)| = |\int_0^1 f(x)\,dx| \leq \int_0^1 |f(x)|\,dx \overset{*}{=} \|f\|_1 \overset{\star}{\leq} \|f\|_p \, \|1\|_q = \|f\|_p \,.$$

The equality marked by $*$ makes use of the inclusion theorem, Theorem 2.2.6, which states that $f \in L^p([0,1])$ implies $f \in L^1([0,1])$. The inequality marked by $\star$ uses the Hölder inequality, Theorem 2.2.4, with $q = p/(p–1)$, by interpreting the integrand f as the product $1 \cdot f$ with constant function $1 : C^0([0,1]) \longrightarrow \mathbb{K} : x \mapsto 1$. Note that for each $q \geq 1$ we have $\|1\|_q = 1$ in this example. The case $p = \infty$ is admissible and reproduces the previous example. The case $p = 1$, or $q = \infty$, is admissible as well. The input space may be extended to $L^p([0,1])$. As a result it is clear that if $\|f\|_p \longrightarrow 0$, then also $|A(f)| \longrightarrow 0$, proving well-posedness of A.

This example may be generalized even further.

Example 3.2.9. Let $A : L^p([0,1]) \longrightarrow \mathbb{K}$, with $p \geq 1$, be the linear integral operator given by

$$A(f) = \int_0^1 f(x)\,\phi(x)\,dx \,,$$

with fixed $\phi \in L^q([0,1])$, and q satisfying the Hölder condition $1/p + 1/q = 1$. Then, again with the help of the Hölder inequality,

$$|A(f)| = |\int_0^1 f(x)\,\phi(x)\,dx| \leq \int_0^1 |f(x)\,\phi(x)|\,dx = \|f\,\phi\|_1 \leq \|f\|_p \, \|\phi\|_q \,.$$

Again, if $\|f\|_p \longrightarrow 0$, then also $|A(f)| \longrightarrow 0$, once again proving well-posedness of A.

The above examples have in common that they are based on integral operators. As a final example we consider a case that does not rely on integration.

Example 3.2.10. Let $A : C^0([0,1]) \longrightarrow \mathbb{K}$ be the linear *point evaluation operator* given by

$$A(f) = f(a) \quad \text{for some } a \in [0,1] \,.$$

If $\|f\|_\infty \longrightarrow 0$, then $|A(f)| \longrightarrow 0$, since

$$|A(f)| = |f(a)| \leq \max_{x\in[0,1]} |f(x)| = \|f\|_\infty \,.$$

We conclude that A is well-posed relative to the $(\| \, . \, \|_\infty, | \, . \, |)$-norms.

These examples make clear that the specification of domain, codomain, and associated norms, is important. For if, in the last example, we would have replaced the input space by $L^1([0,1])$, then the limit $\|f\|_1 \longrightarrow 0$ does not permit us to say anything about convergence of $|A(f)|$. In fact, you should be able to provide a counterexample of well-posedness by constructing a sequence of functions f_n, $n \in \mathbb{N}$, such that $\lim_{n\longrightarrow\infty} \|f_n\|_1 = 0$ while $|A(f_n)| = |f_1(a)| \neq 0$ remains fixed for all $n \in \mathbb{N}$.

3.3. Distributional Differentiation

The core of ill-posedness of classical differentiation lies in its defining, physically void infinitesimal limiting procedure. This is a conceptual problem in the continuous domain, but it induces related problems in the discrete setting, such as lack of robustness against noise and strong dependency on the arbitrariness of digitization. One should not want to implement ill-posed operators. In general, ill-posed problems are handled with the help of some *regularization* procedure, rendering the regularized problem well-posed. One should appreciate, though, that a 'regularized problem' is a *different* problem!

In the context of differentiation in particular, the problem of ill-posedness has been solved elegantly in the 1950's by the mathematician Laurent Schwartz [9, 24, 25], collecting various results established during the first half of the twentieth century. Curiously it took several decades before his basically simple idea was appreciated by the image processing community. We will give an introduction to his *theory of tempered distributions* below. This theory has a clear interpretation in the context of image processing. (Mild adaptations of this theory exist, but all are based on the same idea.)

Definition 3.3.1. *The class $\mathscr{S}(\mathbb{R}^n)$ of smooth test functions of rapid decay, $\phi : \mathbb{R}^n \longrightarrow \mathbb{K}$, a.k.a. Schwartz functions, is defined as follows:*

$$\phi \in \mathscr{S}(\mathbb{R}^n) \quad \text{iff} \quad \phi \in C^\infty(\mathbb{R}^n) \quad \text{and} \quad \sup_{x\in\mathbb{R}^n} |x^\alpha \nabla_\beta \phi(x)| < \infty\,,$$

for all multi-indices α and β.

Thus a Schwartz function is an infinitely differentiable function which is 'essentially compact', meaning that it decays 'rapidly' towards infinity. In fact, rapid decay applies to any of its derivatives and means that decay of function values towards zero is faster than any polynomial factor would be able to counteract. Instead of 'polynomial factor' we may read 'function of polynomial growth'.

Definition 3.3.2. *An almost everywhere defined function $f : \mathbb{R}^n \longrightarrow \mathbb{K}$ is called a function of polynomial growth, notation $f \in \mathscr{P}(\mathbb{R}^n)$, if there exist constants $c > 0$ and $m \geq 0$ such that $|f(x)| \leq c\left(1 + \|x\|^2\right)^m$ for almost all $x \in \mathbb{R}^n$.*

In other words, a function of polynomial growth need not be itself a polynomial, but it is bounded from below and from above by a polynomial, except possibly at a set of points of measure zero.

Example 3.3.1. Consider the following functions:

a. $\sin x$,

b. $\dfrac{1}{1+x^2}$,

c. $\exp(-|x|)$,

d. $\exp(-x^2)$.

Of these, only the last one belongs to $\mathscr{S}(\mathbb{R})$. The first one does not, because it does not tend to zero as $x \longrightarrow \pm\infty$. The second one does not, because its decay is not 'sufficiently fast', which can be seen by multiplying it with a factor x^2, which leads to a product function that does not decay towards zero as $x \longrightarrow \pm\infty$. The third one does not, because although its decay is sufficiently fast (recall $\lim_{x\longrightarrow\pm\infty} x^m \exp(-|x|) = 0$ for all $m \in \mathbb{N}$), it is not smooth, since it has no derivative at the origin. The last one fulfills all conditions of Definition 3.3.1.

The class of Schwartz functions constitutes a convolution algebra, recall Section 2.4.

Definition 3.3.3. *The convolution product of two test functions $\phi, \psi \in \mathscr{S}(\mathbb{R}^n)$ is defined by*

$$(\phi * \psi)(x) \stackrel{\text{def}}{=} \int \phi(y)\, \psi(x-y)\, dy\,.$$

Theorem 3.3.1. *$\mathscr{S}(\mathbb{R}^n)$ constitutes a convolution algebra, i.e. for all $\phi, \psi \in \mathscr{S}(\mathbb{R}^n)$ we have $\phi * \psi \in \mathscr{S}(\mathbb{R}^n)$.*

We now turn to the so-called *topological dual* of $\mathscr{S}(\mathbb{R}^n)$, notation $\mathscr{S}'(\mathbb{R}^n)$. To this end recall the definition of a linear functional introduced in Section 2.3.

Definition 3.3.4. *The topological dual $\mathscr{S}'(\mathbb{R}^n)$ of $\mathscr{S}(\mathbb{R}^n)$ is the class of linear functionals $T : \mathscr{S}(\mathbb{R}^n) \longrightarrow \mathbb{K}$, which are continuous[2]. The linear functionals $T \in \mathscr{S}'(\mathbb{R}^n)$ are called tempered distributions.*

Thus $\mathscr{S}'(\mathbb{R}^n)$ is also referred to as the *class of tempered distributions*. Without proof we state the following result, which explains the attribute 'tempered'.

Theorem 3.3.2. *If $T \in \mathscr{S}'(\mathbb{R}^n)$, then there exist a constant $c > 0$ and multi-indices α, β such that*

$$|T\,[\phi]\,| \leq c \sup_{x\in\mathbb{R}^n} |x^\alpha \nabla_\beta \phi(x)|\,.$$

Two tempered distributions $T_1, T_2 \in \mathscr{S}'(\mathbb{R}^n)$ are equal if and only if $T_1[\phi] = T_2[\phi]$ for all $\phi \in \mathscr{S}(\mathbb{R}^n)$, i.e. if they yield the same response when exposed to an arbitrary collection of test functions. A tempered distribution $T \in \mathscr{S}'(\mathbb{R}^n)$ is called *positive* if $T[\phi] > 0$ for all positive test functions $\phi \in \mathscr{S}(\mathbb{R}^n)$. Of special interest are the so-called *regular tempered distributions*.

Definition 3.3.5. *The regular tempered distribution $T_f \in \mathscr{S}'(\mathbb{R}^n)$ associated with a function of polynomial growth, $f \in \mathscr{P}(\mathbb{R}^n)$, is the tempered distribution*

$$T_f : \mathscr{S}(\mathbb{R}^n) \longrightarrow \mathbb{R} : \phi \mapsto \int_{\mathbb{R}^n} f(x)\, \phi(x)\, dx\,.$$

In other words, a regular tempered distribution is associated with a 'function under the integral'. One may verify that a regular tempered distribution as defined by Definition 3.3.5 is indeed a tempered distribution according to Theorem 3.3.2. A regular tempered distribution T_f is often whimsically identified with the function f, even though this correspondence is not without ambiguity.

Example 3.3.2. Consider the tempered distribution $T : \mathscr{S}(\mathbb{R}) \longrightarrow \mathbb{R}$ defined by

$$T[\phi] = \langle\phi\rangle_{[a,b]} \stackrel{\text{def}}{=} \frac{1}{b-a}\int_a^b \phi(x)\, dx\,,$$

i.e. the average value of the test function ϕ on a finite interval $[a, b] \subset \mathbb{R}$. This is in fact a regular tempered distribution, since it can be written as $T[\phi] = T_f[\phi]$, cf. Definition 3.3.5, with

$$f(x) \stackrel{\text{def}}{=} \frac{1}{b-a}\,\chi_{[a,b]}(x)\,,$$

in which $\chi_{[a,b]}$ is the indicator function on $[a, b]$.

[2]The requirement of continuity is necessary in the context of ∞-dimensional vector spaces. Linear functionals on finite-dimensional vector spaces are always continuous.

Thus a function f of polynomial growth uniquely determines a regular tempered distribution T_f. However, the next example shows that a regular tempered distribution T_f does not uniquely determine the function f.

Example 3.3.3. Consider the regular tempered distributions $T_f, T_g \in \mathscr{S}'(\mathbb{R}^n)$, with $f, g \in \mathscr{P}(\mathbb{R}^n)$ such that

$$f(x) = g(x) \quad \text{for all } x \neq 0, \text{ and } f(0) \neq g(0).$$

Then although $f \neq g$ we have

$$T_f[\phi] \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} f(x)\,\phi(x)\,dx = \int_{\mathbb{R}^n} g(x)\,\phi(x)\,dx \stackrel{\text{def}}{=} T_g[\phi] \quad \text{for all } \phi \in \mathscr{S}(\mathbb{R}^n),$$

since the value of f or g at a single point does not affect the integral. Consequently we must conclude that $T_f = T_g \in \mathscr{S}'(\mathbb{R}^n)$.

One can show that if $f, g \in \mathscr{P}(\mathbb{R}^n)$ are such that $f \neq g$, but $T_f = T_g$, then the associated functions f and g must be equal *almost everywhere*, meaning $f(x) = g(x)$ for all $x \in \mathbb{R}^n$ *except* possibly for a set of exceptions $a \in \mathbb{R}^n$ that has no effect on the defining integrals. A subset of such points is called a *set of measure zero*. This shows that it is more natural to interpret the label f as an *equivalence class of functions* rather than a single function. This allows us to remove ambiguity by redefining what we mean by saying that two functions f and g are equal. Instead of defining $f = g$ if and only if $f(x) = g(x)$ for *all* $x \in \mathbb{R}^n$ we say that f and g are *equivalent* (or, members of the same equivalence class), notation $f \sim g$ if and only if $f(x) = g(x)$ for *almost all* $x \in \mathbb{R}^n$.

The following example shows two functions that differ on a set of measure zero, and are thus to be regarded equivalent.

Example 3.3.4. Consider the functions $f_1, f_2 : \mathbb{R} \longrightarrow \mathbb{R}$, defined as follows:

$$f_1(x) = 0 \quad \text{for all } x \in \mathbb{R}, \quad \text{respectively} \quad f_2(x) = \begin{cases} 0 & \text{if } x \in \mathbb{R}\backslash\mathbb{Q} \\ 1 & \text{if } x \in \mathbb{Q}. \end{cases}$$

Then $f_1 \sim f_2$ because $\int_{-\infty}^{\infty} (f_2(x) - f_1(x))\,\phi(x)\,dx = 0$ for all $\phi \in \mathscr{S}(\mathbb{R})$. This is because the rational numbers form a set of measure zero (no proof here). In classical sense the functions f_1 and f_2 are quite different, e.g. f_1 is smooth, whereas f_2 is everywhere discontinuous, and yet they coincide in distributional sense. In this case the null function f_1 would be the most parsimonious representative of the equivalence class.

Definition 3.3.6. *It is customary to write $f = g$ instead of $f \sim g$ for two equivalent functions of polynomial growth, $f, g \in \mathscr{P}(\mathbb{R}^n)$, associated with the regular tempered distributions $T_f, T_g \in \mathscr{S}'(\mathbb{R}^n)$.*

Remember, though, that outside the context of distributions, functions are identified if and only if they agree in *all* their arguments. In a way the strength of distribution theory lies in the many-to-one relationship between functions and regular tempered distributions, because the latter ignores, so to speak, unobservable details. Functions, on the other hand, need to be specified at an infinitesimal level of detail, which has no physical significance.

The space of tempered distributions is strictly larger than the space of regular tempered distributions, as the following lemma shows.

Lemma 3.3.1. *Let $\phi \in \mathscr{S}(\mathbb{R}^n)$, and $a \in \mathbb{R}^n$ any given point. Then*

$$\delta_a : \mathscr{S}(\mathbb{R}^n) \longrightarrow \mathbb{K} \;:\; \phi \mapsto \phi(a)$$

defines a tempered distribution, called the Dirac point distribution at point $a \in \mathbb{R}^n$.

Lemma 3.3.2. *Cf. Lemma 3.3.1 and Definition 3.3.5. The Dirac point distribution $\delta_a \in \mathscr{S}'(\mathbb{R}^n)$ is not a regular tempered distribution.*

Proof. Suppose there exists a function f such that for all $\phi \in \mathscr{S}(\mathbb{R}^n)$,

$$\phi(a) = \int_{\mathbb{R}^n} f(x)\,\phi(x)\,dx\,.$$

In particular, consider a test function that vanishes identically on the open ball $B(a,\varepsilon) = \{x \in \mathbb{R}^n \mid \|x-a\| < \varepsilon\}$, but is otherwise arbitrary, for arbitrarily small $\varepsilon > 0$, in which case we have

$$0 = \int_{\mathbb{R}^n \setminus B(a,\varepsilon)} f(x)\,\phi(x)\,dx\,,$$

showing that f must have its essential support confined to the region $B(a,\varepsilon)$. But ε is arbitrary, implying supp $f = \{a\}$, whence supp $(f\,\phi) = \{a\}$, contradicting a non-zero value of the first integral above.

Definition 3.3.7. *In spite of Lemma 3.3.2 one tends to abuse the notation of Definition 3.3.5, writing any tempered distribution T in the form T_f, even if no such function $f \in \mathscr{P}(\mathbb{R}^n)$ exists. In particular, the Dirac point distribution δ_a is then associated with the 'function under the integral' $\delta_a(x) = \delta(x-a)$,*

$$\delta_a[\phi] \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} \delta_a(x)\,\phi(x)\,dx \stackrel{\text{def}}{=} \phi(a)\,.$$

In turn this encourages the sloppiness of writing tempered distributions as if they were ordinary functions in all cases, i.e. whether they are regular or not. The Dirac point distribution at the origin $a = 0$ is often simply written as δ, or as $\delta(x)$ in function form. Notice, however, that $\delta(x) = 0$ for all $x \neq 0$, and that it is meaningless to ask for the value $\delta(0)$.

Definition 3.3.8. *Regular tempered distributions corresponding to functions $f, g, h, \ldots$ will henceforth be denoted by corresponding capitals $F \stackrel{\text{def}}{=} T_f, G \stackrel{\text{def}}{=} T_g, H \stackrel{\text{def}}{=} T_h, \ldots$*

So small f maps points and capital F maps test functions.

Remark. One often declines to make a notational distinction between a function and its associated distribution. Thus if we use the symbol f to denote either, it should be clear from the context whether it denotes a proper function (prototype: $f : \mathbb{R}^n \longrightarrow \mathbb{K}$, or $f \in \mathscr{P}(\mathbb{R}^n)$), or a distribution (prototype: $f : \mathscr{S}(\mathbb{R}^n) \longrightarrow \mathbb{K}$, or $f \in \mathscr{S}'(\mathbb{R}^n)$).

The core of distribution theory is that we can define *arbitrary derivatives* of tempered distributions in a *well-posed* way.

Definition 3.3.9. *For any nonnegative multi-index α the derivative $\nabla_\alpha F \in \mathscr{S}'(\mathbb{R}^n)$ of a tempered distribution $F \in \mathscr{S}'(\mathbb{R}^n)$ is the tempered distribution defined by*

$$\nabla_\alpha F : \mathscr{S}(\mathbb{R}^n) \longrightarrow \mathbb{K} \;:\; \phi \mapsto (\nabla_\alpha F)\,[\phi] \stackrel{\text{def}}{=} F[\nabla_\alpha^\dagger \phi]\,,$$

in which the transposed partial differential operator is given by $\nabla_\alpha^\dagger = (-1)^{|\alpha|}\nabla_\alpha$.

The reason for the minus signs is that in the subspace of regular tempered distributions T_f for which f is sufficiently smooth, the above definition boils down to the classical definition of differentiation. This can be verified by repeated partial integrations using the integral expression on the right hand side of Definition 3.3.5 and the fact that f is smooth and of polynomial growth. For let $f \in C^\infty(\mathbb{R}^n) \cap \mathscr{P}(\mathbb{R}^n)$, then on the one hand we have

$$\nabla_\alpha F[\phi] \stackrel{\text{def}}{=} F[\nabla_\alpha^\dagger \phi] = (-1)^{|\alpha|} \int_{\mathbb{R}^n} f(x)\,\nabla_\alpha \phi(x)\,dx\,,$$

by Definition 3.3.9. On the other hand, since f is smooth, we may consider the regular tempered distribution $T_{\nabla_\alpha f}$ associated with the function $\nabla_\alpha f$, i.e.

$$\int_{\mathbb{R}^n} \nabla_\alpha f(x)\,\phi(x)\,dx \stackrel{\text{p.i.}}{=} (-1)^{|\alpha|} \int_{\mathbb{R}^n} f(x)\,\nabla_\alpha \phi(x)\,dx\,,$$

where the right hand side has been obtained from the left hand side by repeated partial integrations. *Note that all boundary terms that show up in this process vanish as a result of the rapid decay of the test function.* Thus each time we transfer a partial derivative from f to ϕ in the integrand we merely obtain a minus-sign. Comparing the last two results we observe that the regular tempered distribution associated with the function $\nabla_\alpha f$ is indeed precisely $\nabla_\alpha F$. This explains why Definition 3.3.9 is the natural way to define $\nabla_\alpha F$. Of course, most functions of interest are not differentiable, which is why the definition does not make use of derivatives of f, but uses a detour via the test function space. After all, the test functions are infinitely differentiable by construction.

In particular, we can now differentiate our Dirac point distribution.

Result 3.3.1. *Recall Definition 3.3.9. Let $\phi \in \mathscr{S}(\mathbb{R}^n)$, $a \in \mathbb{R}^n$ any given point, and α a multi-index. Then*

$$\nabla_\alpha \delta_a : \mathscr{S}(\mathbb{R}^n) \longrightarrow \mathbb{K} \ : \ \phi \mapsto \nabla_\alpha^\dagger \phi(a) = (-1)^{|\alpha|} \nabla_\alpha \phi(a) \, .$$

It can be shown that a distribution whose support is a single point is a finite linear combination of derivatives of the Dirac distribution at that point.

Theorem 3.3.3. *If $F \in \mathscr{S}'(\mathbb{R}^n)$ is a general point distribution, i.e. if* supp $F = \{a\}$ *for some point $a \in \mathbb{R}^n$, then there exists an order $m \in \mathbb{Z}_0^+$ such that*

$$F = \sum_{|\alpha| \leq m} c_\alpha \nabla_\alpha \delta_a \, .$$

A special case is $m = 0$ (such point distributions are also referred to as *point measures*), in which case we end up a with single 'spike' δ_a at point a with a certain amplitude c_0. Linear combinations of such spikes, one for each grid point, may be used to model a discrete image on a continuous domain. This model is actually close to common practice since raw images are stored by specifying the amplitudes associated with each grid point in a file (the 'raw data'). The grid itself can thus be identified with the linear combination of tempered distributions δ_a at grid points $a \in \mathbb{Z}^n$ (the information of which is typically stored in the 'header' accompanying the raw data).

We have the following inclusions:

$$\mathscr{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n) \subset \mathscr{S}'(\mathbb{R}^n) \, . \tag{3.2}$$

In other words, $\mathscr{S}'(\mathbb{R}^n)$ is a very large space, so in practice we may always model an image as a tempered distribution. The test functions, $\mathscr{S}(\mathbb{R}^n)$, on the other hand, are quite restrictive. Because of their spatial confinement and regularity properties they are suitable for representing filters. For future purposes we therefore decide upon the following model once and for all.

Definition 3.3.10. *A raw image f is a tempered distribution, a filter ϕ is a smooth function of rapid decay:*

$$f \in \mathscr{S}'(\mathbb{R}^n) \quad \textit{and} \quad \phi \in \mathscr{S}(\mathbb{R}^n) \, .$$

Here we have identified a function f with its tempered distribution F, recall the remark on page 51.

For simplicity, Latin symbols $f, g, h, \ldots$ (or $F, G, H, \ldots$, cf. a previous remark) will henceforth be used to denote images, whereas Greek symbols $\phi, \chi, \psi, \ldots$ will be reserved for filters.

It is possible to define convolutions and correlations of distributions, most easily with the help of the somewhat sloppy function notation for the distributions.

Definition 3.3.11. *The convolution product of a tempered distribution $f \in \mathscr{S}'(\mathbb{R}^n)$ with a test function $\phi \in \mathscr{S}(\mathbb{R}^n)$ is defined by*

$$(f * \phi)(x) \stackrel{\text{def}}{=} \int f(y) \, \phi(x - y) \, dy \, .$$

And we have indeed a kind of 'closure' property under convolution on $\mathscr{S}'(\mathbb{R}^n)$.

Theorem 3.3.4. *For all $f \in \mathscr{S}'(\mathbb{R}^n)$, $\phi \in \mathscr{S}(\mathbb{R}^n)$, we have $f * \phi \in \mathscr{S}'(\mathbb{R}^n)$.*

Definition 3.3.12. *The correlation product of a tempered distribution $f \in \mathscr{S}'(\mathbb{R}^n)$ with a test function $\phi \in \mathscr{S}(\mathbb{R}^n)$ is defined as*

$$(f \star \phi)(x) \stackrel{\text{def}}{=} \int f(y)\, \phi(y - x)\, dy\,.$$

Theorem 3.3.5. *For all $f \in \mathscr{S}'(\mathbb{R}^n)$, $\phi \in \mathscr{S}(\mathbb{R}^n)$, we have $f \star \phi \in \mathscr{S}'(\mathbb{R}^n)$.*

In other words, if you filter an image by a (convolution or correlation) filter, the result is again an image. This is by no means trivial, but follows from our choice of function spaces. These type of image in/image out procedures are the subject of the field of *image processing*.

Convolution and correlation (Theorems 3.3.4 and 3.3.5) produces functions in $C^\infty(\mathbb{R}^n) \cap \mathscr{P}(\mathbb{R}^n) \subset \mathscr{S}'(\mathbb{R}^n)$, regardless of the (lack of) regularity of the input image f. The rule of thumb is that a convolution or correlation product inherits the nicest regularity properties from each of its operands, thus in the case at hand ∞-differentiability from the filter.

We may concatenate convolutions or correlations. The former adheres to the usual algebraic associativity rule. In the latter case we have the following result.

Result 3.3.2. *If $f \in \mathscr{S}'(\mathbb{R}^n)$ and $\phi, \psi \in \mathscr{S}(\mathbb{R}^n)$, then $(f \star \phi) \star \psi = f \star (\phi * \psi)$.*

Once again a convolution product emerges as the effective correlation filter corresponding to consecutive correlations.

A general theorem useful in the context of convolution algebras follows from the Hölder inequality (Theorem 2.2.4 on Page 29).

Theorem 3.3.6 (Young Inequality)**.** *Let $1 \leq p, q \leq \infty$ with $1/p + 1/q \geq 1$, and let $f \in L^p(\Omega)$, $\phi \in L^q(\Omega)$, then*

$$\|f * \phi\|_r \leq \|f\|_p \|\phi\|_q\,,$$

where $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$.

See e.g. Wheeden and Zygmund [31] for a proof. For further details on distribution theory, cf. [1, 3, 5, 9, 10, 11, 24, 25, 30].

3.4. Scale Space Theory

3.4.1. Retrospective

In distributional sense a derivative of a function is only defined *operationally* by virtue of a class of suitable test functions ('duality'). This shift of paradigm has profound implications for image analysis:

- **Flexibility:** Test functions and their derivatives can be seen as *filters* for linear image processing (including point sampling and differentiation). Moreover, these filters are not chained to a particular grid, but can be specified prior to discretization. Thus differentiation and discretization have been decoupled unlike with numerical differentiation in classical sense.

- **Robustness:** Distributions are infinitely differentiable in a well-posed sense, with no no artificial restriction to finite order or to smooth functions, which are physical non-entities. After all, differences between various C^k classes are 'infinitesimal', and cannot be revealed by finite resolution measurements. Thus a C^k-classification makes no sense in a theory of empirically obtained image functions. If an image is differentiable in operationally well-defined terms, it must be infinitely differentiable.

Example 3.4.1. Suppose $f : \mathbb{R} \longrightarrow \mathbb{R}$ is the discontinuous function defined by $f(x) = \operatorname{sign}(x)$. We cannot differentiate this function in classical sense, nor even consider its zeroth order value at $x = 0$. This is peculiar, as the discontinuity pertains to a single point, i.e. a set of measure zero, which should not cause any operational restrictions 'in real life'. Indeed, there is no problem in distributional sense.

You should find the distributional derivative of the function f in this example as an instructive exercise.

In conclusion, in the context of regular tempered distributions we must *never assign significance to the value of a function f at any isolated point x* if this function is intended to represent a distribution. Since we have agreed on modeling images as tempered distributions this means that it is meaningless to ask for the value of an image at a point! Indeed, in practice one can only ask for a pixel value of a discrete image, which represents some physical observable integrated over a finite spatial volume element rather than some infinitesimal point entity. The structure of an image at sub-pixel scales is beyond our reach. Filters, on the other hand, are completely under our control. Therefore we have chosen to model them as smooth functions of rapid decay for convenience. This is also why we can take their (pen-and-paper) derivatives in classical sense. Since we are in possession of their analytical expressions, the act of (classical) differentiation is *exact*, i.e. no approximations are involved, so the problem of ill-posedness does not apply. In particular it does make sense to evaluate a test function at a single point, as opposed to a distribution.

3.4.2. Scale Space

For practical purposes we would like to restrict the class of admissible filters to a proper subspace of $\mathscr{S}(\mathbb{R}^n)$. However, if we would do this in an ad hoc fashion we are likely to violate the closure properties of Theorem 3.3.1, Page 49, and Theorems 3.3.4 and 3.3.5, Page 53. Combined with Result 3.3.2, Page 53, it is clear that we must look for a proper subset of $\mathscr{S}(\mathbb{R}^n)$ which is closed under convolution, i.e. if $\Phi \subset \mathscr{S}(\mathbb{R}^n)$ is such a subset, then we must require that $\phi * \psi \in \Phi$ whenever $\phi, \psi \in \Phi$.

As a starting point, notice that the *class of nonnegative Schwartz functions*, is closed under convolution:

$$\mathscr{S}^+(\mathbb{R}^n) = \{\phi \in \mathscr{S}(\mathbb{R}^n) \mid \phi \geq 0 \text{ and } \phi \not\equiv 0\} .$$

Without loss of generality we may restrict ourselves to filters $\phi \in \mathscr{S}^+(\mathbb{R}^n)$, which have their 'centre of gravity' at the origin, i.e.

$$\int_{\mathbb{R}^n} x^\mu \, \phi(x) \, dx = 0 \quad (\mu = 1, \ldots, n), \tag{3.3}$$

and which are normalized to unity in the sense that

$$\int_{\mathbb{R}^n} \phi(x) \, dx = 1 . \tag{3.4}$$

Eq. (3.3) ensures that the filter is confined to the immediate neighborhood of the origin, so that it may act as a linear probing device for extracting local samples. Eq. (3.4) allows us to interpret the filter as a local averaging device, so that no artificial attenuation or amplification of signal will occur. A filter $\phi \in \mathscr{S}^+(\mathbb{R}^n)$ with these two properties may be aptly called a *zeroth order filter*, and a sample of the form $(f \star \phi)(a)$ a *zeroth order sample* of the image (at point $a \in \mathbb{R}^n$).

The class $\mathscr{S}^+(\mathbb{R}^n)$ is still huge, so we need to further restrict the class of admissible zeroth order filters without violating closure.

A typical thing that happens when you convolve two filters ϕ and ψ is that the profile of $\phi * \psi$ changes relative to that of each of the input filters (even if $\psi = \phi$), and that it becomes broader. As an instructive example, you may want to consider the filter[3] $\phi_0 : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $\phi_0(x) = 1$ if $x \in [-\frac{1}{2}, \frac{1}{2}]$, and $\phi_0(x) = 0$ otherwise, compute $\phi_1 = \phi_0 * \phi_0$, and $\phi_2 = \phi_1 * \phi_0$, and sketch their graphs.

Thus as soon as we admit some arbitrary filter profile we are likely to end up with an avalanche of profiles, which is what we intended to avoid! This suggests that we should focus our attention on a single template $\phi \in \mathscr{S}^+(\mathbb{R}^n)$, subject to the requirements of Eqs. (3.3–3.4), which retains its profile under *autoconvolution*. That is to say, although the extent and amplitude of the generating filter may vary, its shape must be uniquely determined. This *self-similarity constraint* is a very powerful constraint. In fact we immediately obtain the following result.

Theorem 3.4.1. *The only isotropic filters within $\mathscr{S}^+(\mathbb{R}^n)$ that retain their shape under autoconvolution are the normalized Gaussians:*

$$\phi_\sigma(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}^{\,n}} \exp\left[-\frac{1}{2}\frac{\|x\|^2}{\sigma^2}\right] \quad \textit{with } \sigma > 0.$$

In fact we have

$$\phi_\sigma * \phi_\tau = \phi_{\sqrt{\sigma^2+\tau^2}}\,.$$

You should be able to verify that Eqs. (3.3–3.4) are satisfied for all $\sigma > 0$. (Hint: First show that this does not depend on $\sigma > 0$, so that we may take $\sigma = 1$. Subsequently set $n = 2$ and integrate using polar coordinates. Finally generalize the result to arbitrary $n \in \mathbb{N}$.) Cf. Figure 3.11 for an illustration. The parameter $\sigma \in \mathbb{R}^+$ determines the width of the Gaussian and is called the *isotropic scale parameter*.

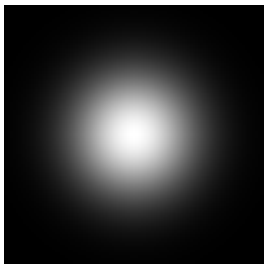


Figure 3.11: Gaussian filter in two dimensions.

One often encounters the normalized Gaussians using a different scale parametrization, viz.

$$\psi_t(x) = \frac{1}{\sqrt{4\pi\,t}^{\,n}} \exp\left[-\frac{\|x\|^2}{4t}\right] \quad \text{with } t > 0. \tag{3.5}$$

The autoconvolution property of Theorem 3.4.1 then takes the form

$$\psi_t * \psi_u = \psi_{t+u}\,. \tag{3.6}$$

Using this parametrization, the Gaussians $\psi_t(x)$ satisfy the following *isotropic heat equation*:

$$\frac{\partial u}{\partial t} = \Delta u\,. \tag{3.7}$$

Note that the scale parameters of Theorem 3.4.1 and Eq. (3.5) are related by $2t = \sigma^2$.

[3]Note that $\phi_0 \notin \mathscr{S}^+(\mathbb{R})$, but this does not affect the essence of the argument, which applies to virtually any filter set.

Theorem 3.4.1 (or Eq. (3.5)) yields the smallest filter class contained in $\mathscr{S}^+(\mathbb{R}^n)$ that is closed under convolution. In particular there exists no parameter-free filter in $\mathscr{S}(\mathbb{R}^n)$ that is invariant under convolution, since convolution always broadens the width of a function. The corresponding Gaussians depend on a single *isotropic scale parameter* ($\sigma \in \mathbb{R}^+$, or $t \in \mathbb{R}^+$), which controls their widths.

For the sake of completeness we state the result for a non-isotropic Gaussian (summation convention applies, recall Page 6). If

$$\phi_\Lambda(x) = \frac{1}{\sqrt{2\pi}^n \sqrt{\det \Lambda}} \exp\left[-\frac{1}{2}\, x^\mu\, \Lambda^{-1}_{\mu\nu}\, x^\nu\right], \tag{3.8}$$

in which Λ is a symmetric positive definite matrix. We have the following closure property:

$$\phi_\Lambda * \phi_{\mathrm{M}} = \phi_{\Lambda+\mathrm{M}}\,. \tag{3.9}$$

In the isotropic case we have $\Lambda = \sigma^2\,\mathrm{I}$, in which I is the $n \times n$ identity matrix. We may need Eq. (3.8) to correct for grid anisotropies. For instance, some tomographic acquisition protocols lead to differences between in-slice ('x–y') and slice-to-slice ('z') resolutions. Eq. (3.8) is also of interest in the context of spatiotemporal images (movies), since space and time must have their own physical scale parameter. Assuming isotropic space we then have, in three spatial dimensions,

$$\Lambda = \begin{bmatrix} \sigma & 0 & 0 & 0 \\ 0 & \sigma & 0 & 0 \\ 0 & 0 & \sigma & 0 \\ 0 & 0 & 0 & \tau \end{bmatrix} \quad (\sigma, \tau \in \mathbb{R}^+).$$

We have arrived at a special instance of Schwartz theory, referred to in the image literature as *scale space theory* [6, 7, 13, 14, 17, 18, 19, 20, 21, 26, 27]. In scale space theory the class of admissible filters is given by Theorem 3.4.1 (or, more generally, Eq. (3.8)), together with its derivatives.

Definition 3.4.1. *The Gaussian family is the class of all normalized Gaussians, cf. Theorem 3.4.1, together with all its derivatives. The vector space spanned by the Gaussian family will be denoted by $\mathscr{G}(\mathbb{R}^n)$. The vector space spanned by the class of zeroth order Gaussians proper will be denoted by $\mathscr{G}^+(\mathbb{R}^n)$.*

We have $\mathscr{G}(\mathbb{R}^n) \subset \mathscr{S}(\mathbb{R}^n)$ and $\mathscr{G}^+(\mathbb{R}^n) \subset \mathscr{S}^+(\mathbb{R}^n)$.

Consequently, an image is a linear functional from the linear space of filters spanned by the Gaussian family into the real numbers. Informally speaking, an image is 'what you are looking at' using linear measurement probes defined by (a linear combination of) integration filters from the Gaussian family. Figures 3.12–3.13 show the correlation product $f \star \phi_\sigma$ of a raw image f with a normalized Gaussian ϕ_σ for increasing σ.

Definition 3.4.2. *The extended function $u : \mathbb{R}^n \times \mathbb{R}^+ \longrightarrow \mathbb{R}$ given by $u(x,\sigma) = (f \star \phi_\sigma)(x)$, or equivalently, $u(x,\sigma) = (f * \phi_\sigma)(x)$, is called the scale space representation of f. The domain of definition of u is called the scale space domain.*

Thus a scale space representation of a raw image is a representation intended to capture that image at a continuum of scales. From the figure it can be seen that *scale* corresponds to *inverse resolution*. Increasing scale amounts to lowering resolution. Compare this figure to Figure 1.3, Page 11 (in which resolution increases from left to right). Lowering resolution in scale space sense proceeds in a more 'graceful' manner, and does not induce spurious block artifacts.

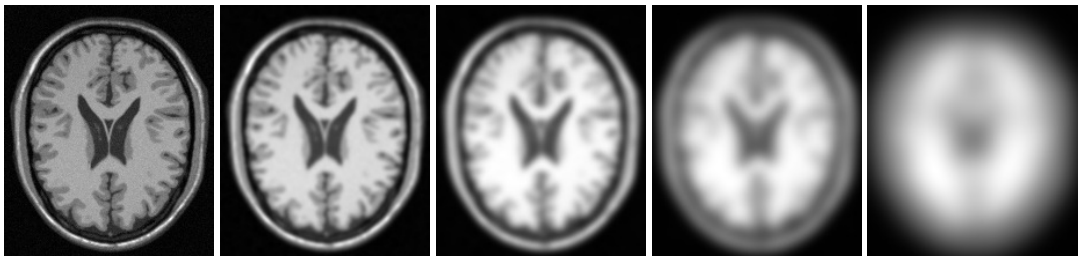


Figure 3.12: Zeroth order scale space representation of a two-dimensional slice from an MRI image. From left to right: $\sigma = 1, 2, 4, 8, 16$ pixels. Notice how the image gradually simplifies. In reality any (physically reasonable) scale level can be computed.

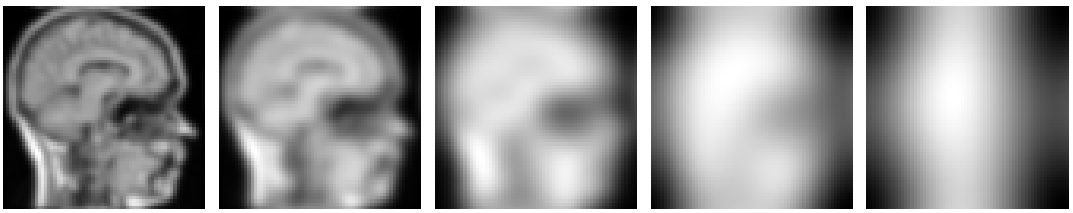


Figure 3.13: Zeroth order scale space representation of a two-dimensional slice from an MRI image. From left to right: $\sigma = 1, 2, 4, 8, 16$ pixels.

Using the fact that the scale space filter $\psi(x,t) = \phi_{\sigma=\sqrt{2t}}(x)$ satisfies the heat equation, recall Eq. (3.7), one can prove the following theorem. (Chapter 4 will provide a technique to prove that the unique solution is indeed given by the scale space representation u of f.)

Theorem 3.4.2. *Let $u : \mathbb{R}^n \times \mathbb{R}^+ \longrightarrow \mathbb{R}$ be the scale space representation of the raw image $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ according to Definition 3.4.2, with $\sigma = \sqrt{2t}$. Then u satisfies the following initial value problem:*

$$\left\{ \begin{array}{rcl} \dfrac{\partial u}{\partial t} & = & \Delta u \\ u(x,0) & = & f(x)\,. \end{array} \right. \tag{3.10}$$

A scale space representation is a zeroth order multi-resolution representation of the raw image data, i.e. it is essentially a family of 'copies' of the raw image, 'degraded' by the blur induced by the finite measurement scale. It can be extended to higher orders by differentiation.

Definition 3.4.3. *Recall Definition 3.4.2. For nonnegative multi-index α the function $\nabla_\alpha u : \mathbb{R}^n \times \mathbb{R}^+ \longrightarrow \mathbb{R}$ given by $\nabla_\alpha u(x,\sigma) = (-1)^{|\alpha|}\,(f \star \nabla_\alpha \phi_\sigma)(x)$, or equivalently, $\nabla_\alpha u(x,\sigma) = (f * \nabla_\alpha \phi_\sigma)(x)$, is called the scale space derivative of f at scale $\sigma > 0$.*

Notice the minus sign in the correlation formula for odd orders, which stems directly from Definition 3.3.9, and notice the absence of a minus sign in the convolution formula.

The autoconvolution property stated in Theorem 3.4.1, or Eq. (3.6), can be generalized as well.

Theorem 3.4.3. *Let $\phi_s, \phi_t \in \mathscr{G}(\mathbb{R}^n)$, and α, β two n-dimensional nonnegative multi-indices. Using the scale parametrization of Eq. (3.6) we have*

$$\nabla_\alpha \phi_s * \nabla_\beta \phi_t = \nabla_{\alpha+\beta} \phi_{s+t}\,.$$

This theorem can be extended to the anisotropic case by replacing the scalar parameters s and t by appropriate matrices, recall Eq. (3.9).

Example 3.4.2. Figures 3.14 and 3.15 illustrate the filter profiles of the Gaussian family in two dimensions up to third order at a fixed scale. Convolving an image with these filters produces corresponding derivatives of the image at the scale of interest, cf. Figure 3.16. Figures 3.12 and 3.13 show a few scale levels of the zeroth order image. Moving 'upward' in scale space reveals a gradual simplification of image structure, not unlike cartographic generalization applied to maps of increasing scales.

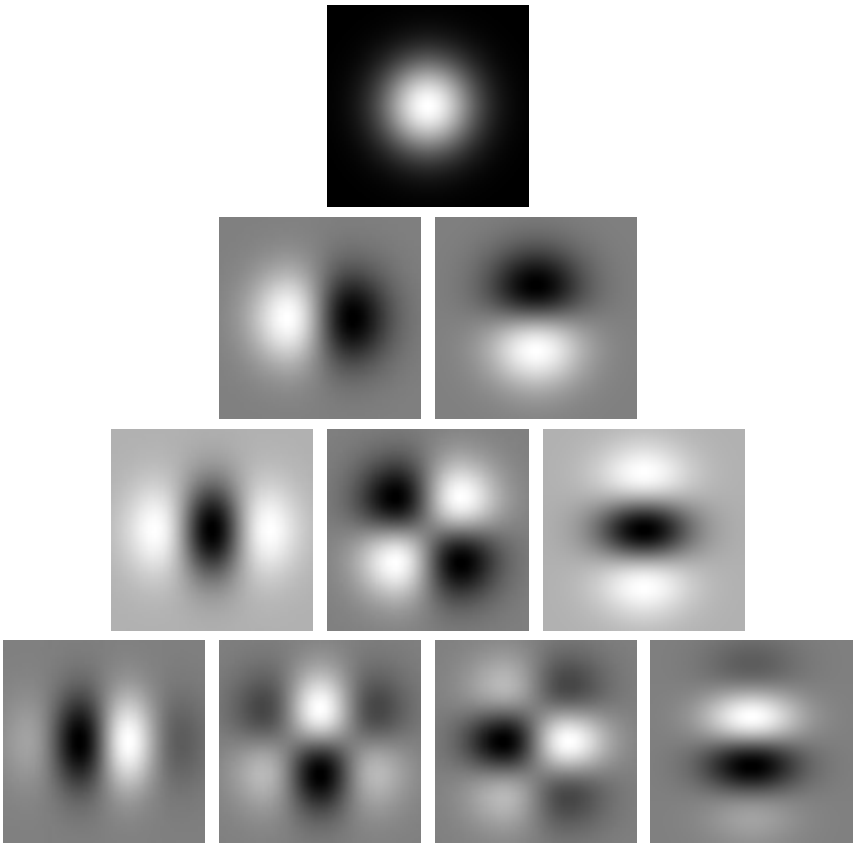
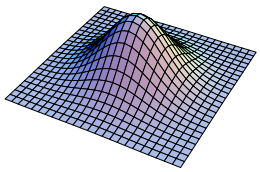


Figure 3.14: Members of the Gaussian family in two dimensions up to third order in Cartesian representation. Order increases from top to bottom. Members of fixed order are displayed from left to right. Scale is the same in all cases. Grey-values are scaled to full range. Top row: 0-th order Gaussian, $\phi(x, y)$. Second row: 1-st order Gaussian derivatives, $\phi_x(x, y)$, respectively $\phi_y(x, y)$. Third row: 2-nd order Gaussian derivatives, $\phi_{xx}(x, y)$, $\phi_{xy}(x, y)$, respectively $\phi_{yy}(x, y)$. Fourth row: 3-rd order Gaussian derivatives, $\phi_{xxx}(x, y)$, $\phi_{xxy}(x, y)$, $\phi_{xyy}(x, y)$, respectively $\phi_{yyy}(x, y)$.



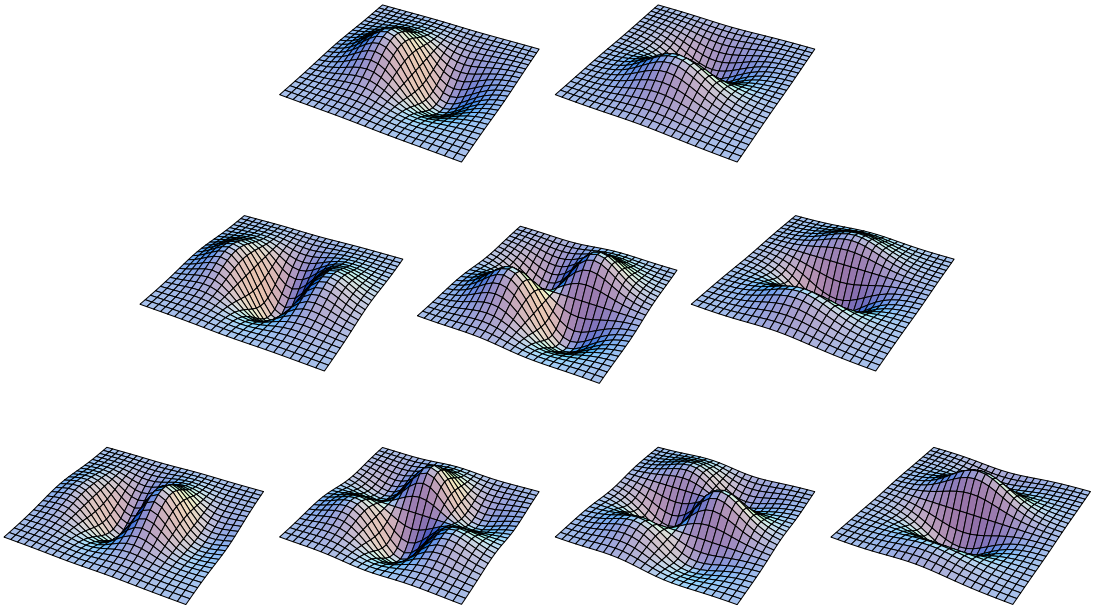


Figure 3.15: Same as Figure 3.14, now with the filters displayed by their graphs.

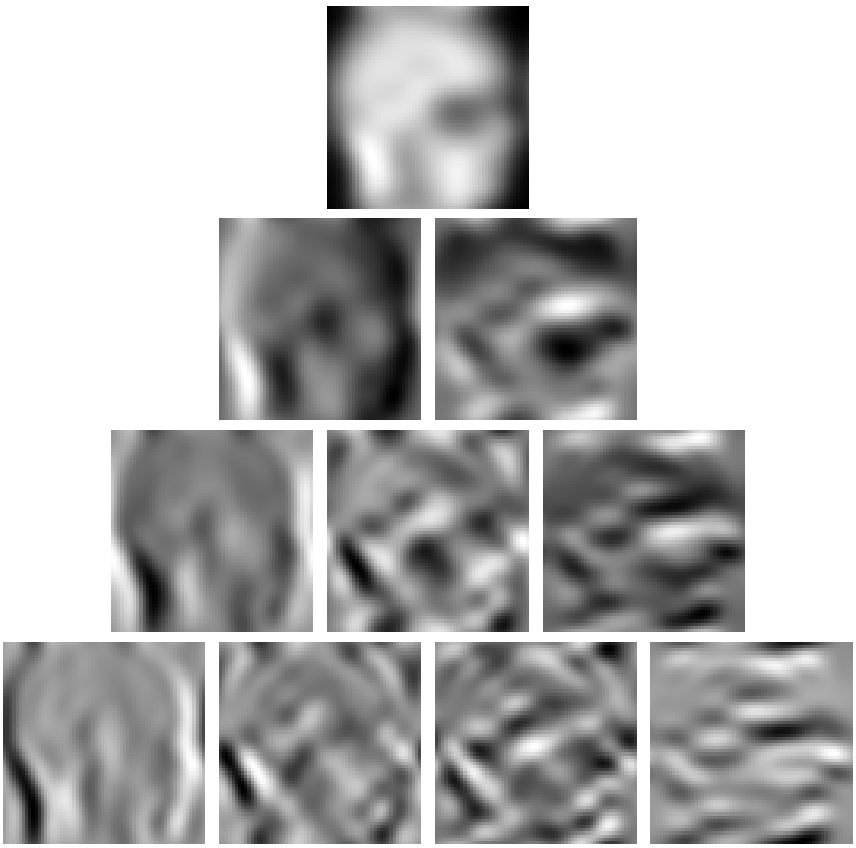


Figure 3.16: Derivatives according to the Gaussian scale space paradigm of the 256×256 MRI image of Figure 1.3, Page 11, using the same ordering as in Figure 3.14 and 3.15. Scale in this example is $\sigma = 4$ pixels.

Example 3.4.3. Figure 3.17 shows a test image corresponding to the following 'noise free' function:

$$f(x,y) = \begin{cases} 1 & \text{if } x^2 + y^2 \leq r^2, \\ 0 & \text{otherwise.} \end{cases}$$

The actual discrete image was obtained by choosing a suitable radius $r > 0$, sampling the above function on a regular, square grid, and adding a substantial amount of pixel-uncorrelated noise using a uniform probability distribution over the range $[-2, 2]$. Figure 3.18 shows a few scale levels from its scale space representation. Notice how the original 'noise free' disk structure gradually emerges as scale increases. This is due to the fact that the additive noise is a microscopic, pixel-scale phenomenon, albeit substantial in amplitude, whereas the disk is macroscopic by construction.

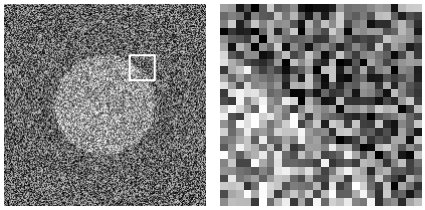


Figure 3.17: Test image with region of interest and magnification of the latter. The bounding box marking the region of interest on the left is not part of the image.

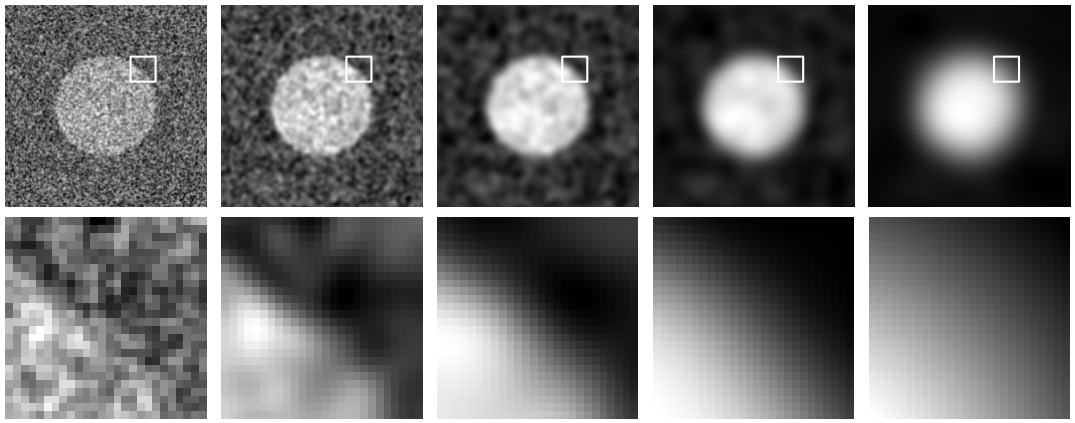


Figure 3.18: Zeroth order scale space representation of the test image of Figure 3.17. From left to right: $\sigma = 1, 2, 4, 8, 16$ pixels. The top row shows the global image, the bottom row is a magnification of the region of interest. Notice how the noise gradually disappears in a trade-off with resolution.

Example 3.4.4. Figure 3.19 shows the scale space representation of the test image's gradient magnitude, i.e. the scale space representation of the hypothetical zero-scale function

$$\|\nabla f(x,y)\| = \sqrt{f_x(x,y)^2 + f_y(x,y)^2}\,.$$

Of course, this is just a formal expression, because f is not a differentiable function, and differentiation in this classical sense has no operational significance. Figure 3.19 actually shows

$$\|\nabla u(x,y,\sigma)\| = \sqrt{u_x(x,y,\sigma)^2 + u_y(x,y,\sigma)^2}\,,$$

for various $\sigma > 0$. Again, as scale increases the fine-scale grey-level transitions induced by noise are de-emphasized, and the coarse-scale boundary of the original disk gradually emerges.

Note that $\|\nabla u\|$ in this example is *not* the scale space representation of $\|\nabla f\|$ in the sense of Definition 3.4.2. (Why not?)

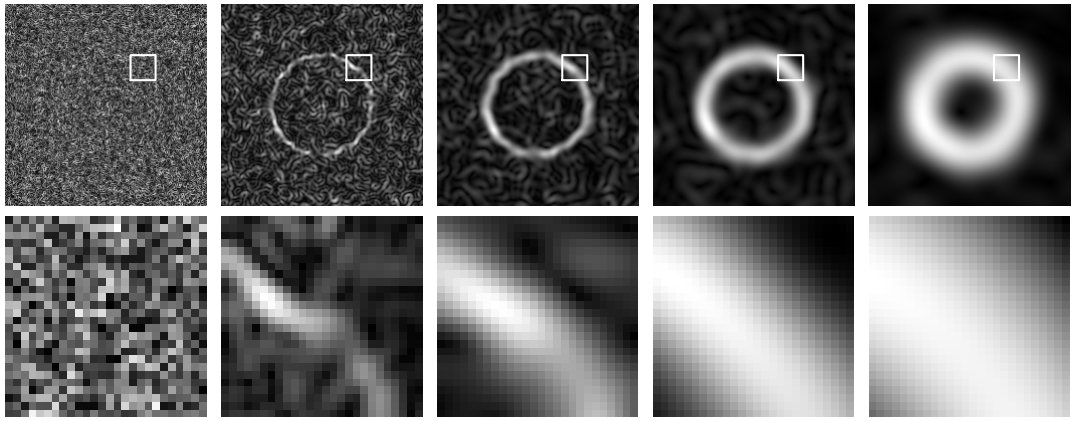


Figure 3.19: Scale space representation of the gradient magnitude of the test image of Figure 3.17. From left to right: $\sigma = 1, 2, 4, 8, 16$ pixels. The top row shows the global image, the bottom row is a magnification of the region of interest. Notice how the grey-level transition at the boundary of the original disk becomes more and more pronounced as scale increases.

3.4.3. Scale and the Correspondence Principle

It makes sense to have a free scale parameter for differentiation (including zeroth order). After all, what an image locally looks like depends on the level of detail you are interested in, in other words, on resolution. If we use pixel size as our unit of length, then choosing $\sigma = 1$ amounts to resolving differential structure roughly at pixel level. This may be undesirable in view of the arbitrariness of the image's rendering. What you basically highlight at this scale are the image's grid details and rendering artifacts, which are typically heavily affected by noise. You should interpret 'noise' generically as any more or less random, or at least unpredictable, grey-value fluctuation caused by the act of discretization, quantization, and measurement uncertainties.

You should resist the temptation to actually take the limit $\sigma \longrightarrow 0$. Nevertheless, the following lemma is of conceptual importance, as it expresses a *correspondence principle*, relating classical differentiation to distributional differentiation according to the Gaussian scale space paradigm.

Lemma 3.4.1. *Let $u : \mathbb{R}^n \times \mathbb{R}^+ \longrightarrow \mathbb{R}$ be the scale space representation of a raw image $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ according to Definition 3.4.2, then $u \in C^\infty(\mathbb{R}^n \times \mathbb{R}^+)$. If in addition $f \in C^\infty(\mathbb{R}^n)$, then*

$$\nabla_\alpha f(x) = \lim_{\sigma \longrightarrow 0} \nabla_\alpha u(x, \sigma) \,,$$

for all multi-indices α of order $|\alpha| \in \mathbb{N}$.

In other words, vanishing scale takes us back to classical differentiation. However, the smoothness assumption on f is as unphysical as the zero-scale limit itself (recall that images are in general tempered distributions, which need not be differentiable or even continuous), and worse, the 'classical limit' in Lemma 3.4.1 would return the misery of ill-posedness that we set out to avoid.

From the above it is clear that the case that really interests us in practice is $\sigma > 1$. Note that there can be no a priori preferred scale. In the absence of prior knowledge about image structure, any scale is as good as any other. This forces us to treat all scales on equal foot, a principle known as *scale invariance*. Of course in practice scale invariance holds only to the extent that we confine ourselves to physical scales actually captured by the image.

It remains an outstanding challenge to exploit scale space theory in image analysis. Nevertheless, many more or less ad hoc applications as well as feasibility studies do appear promising. The incorporation of finite scale and the principle of scale invariance are a 'sine qua non' in the development of robust, grid independent image processing and analysis applications.

4

FOURIER TRANSFORMATION

4.1. Background

One of the most prominent techniques for studying images is Fourier analysis, named after Joseph Fourier, who introduced the basic idea (in a footnote…) in his 'Théorie Analytique de la Chaleur', originally published in 1822 [8]. Fourier analysis deals with two equivalent function representations, known as 'spatial' (or 'temporal', depending on physical interpretation) and 'frequency' (or 'Fourier') representation. The mapping from spatial to frequency representation and its inverse are known as the *Fourier transform* and the *inverse Fourier transform*, respectively.

For our purposes it suffices to consider the so-called *continuous*[1] *Fourier transform* on $\mathscr{S}(\mathbb{R}^n)$, the class of filters, respectively on its dual, $\mathscr{S}'(\mathbb{R}^n)$, the class of images. Since in practice images are discrete, we will also need the *discrete Fourier transform*. This type of transform effectively applies to any multi-dimensional digital function, whether interpreted as an image or as a filter.

All Fourier transforms exploit the fact that any function can be decomposed into a linear combination (a sum or an integral) of sinusoidal functions. The function spaces of interest are all (finite- or infinite-dimensional) vector spaces, and in essence Fourier transformation boils down to a change of basis, recall Section 2.2. We won't elaborate on this linear algebraic interpretation in order to avoid complications inherent to infinite-dimensional vector spaces, but instead concentrate on technique and interpretation.

[1] The attribute 'continuous' pertains to the continuum nature of the spatial or temporal domain of definition. The transformation itself is *always* continuous by construction.

4.2. Mathematical Preliminaries

Before we can state the actual transforms let us recapitulate some basic trigonometry. The pivotal functions are the *trigonometric* or *sinusoidal functions*, $\sin$ and $\cos$. Since these are analytical functions they can be expressed—indeed, defined if you wish—in terms of their convergent Taylor series:

$$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1} \quad (x \in \mathbb{R}), \tag{4.1}$$

$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k} \quad (x \in \mathbb{R}). \tag{4.2}$$

Both functions are bounded and oscillate between ± 1 with a 2π-period. In this respect they seem to bear no relation to the equally popular *exponential function*, $\exp$, which, unlike the sinusoidal functions, has no upper bound, is globally invertible, and is characterized by a rapid, monotonic increase. Since it is also analytical it can again be defined in terms of its Taylor expansion:

$$\exp x = \sum_{k=0}^{\infty} \frac{1}{k!} x^k \quad (x \in \mathbb{R}). \tag{4.3}$$

An important, perhaps not immediately obvious property of the exponential function is its separability:

$$\exp(x+y) = \exp x \, \exp y \quad (x, y \in \mathbb{R}). \tag{4.4}$$

Despite qualitative differences, the trigonometric functions and the exponential function are intimately related. This becomes apparent once we extend the latter definition to the complex plane[2]:

$$\exp z = \sum_{k=0}^{\infty} \frac{1}{k!} z^k \quad (z \in \mathbb{C}). \tag{4.5}$$

The separability property continues to hold for complex arguments. If we substitute a purely imaginary value $z = i\theta$, say, with $\theta \in \mathbb{R}$ and $i^2 = -1$, we can split the series expansion into real and imaginary parts, in which we recognize the sinusoidal expansions, viz.

$$\exp(i\theta) = \cos\theta + i \sin\theta. \tag{4.6}$$

(Convince yourself that this is indeed the case.) Alternatively we may write (check this):

$$\cos\theta = \frac{\exp(i\theta) + \exp(-i\theta)}{2}, \tag{4.7}$$

$$\sin\theta = \frac{\exp(i\theta) - \exp(-i\theta)}{2i}. \tag{4.8}$$

The complex exponential representation, $\exp(i\theta)$, is often preferred over its counterpart in terms of the real-valued pair ($\cos\theta$,$\sin\theta$), due to the nice properties of the $\exp$-function. For instance, Eq. (4.6), applied to $k\theta$ for $k \in \mathbb{N}$ instead of θ, combined with the property

$$\exp(kz) = (\exp(z))^k, \tag{4.9}$$

and Newton's binomium, allows you to express $\cos(k\theta)$ and $\sin(k\theta)$ in terms of $\cos\theta$ and $\sin\theta$.

Thus the sinusoidal functions show up as the real and imaginary parts of the exponential function on the imaginary axis. More generally, if $z = x + iy$, with $x, y \in \mathbb{R}$, then we obtain

$$\exp(x+iy) = \exp x \, (\cos y + i \sin y).$$

[2]Although this is not at all obvious from the defining series, we will take it for granted that $\exp z = e^z$, in which $e \approx 2.71828$ is a famous constant, and we will use both notations interchangeably.

4.3. The Fourier Transform on a Continuous Domain

4.3.1. Some Conceptual Issues

Two function spaces are of special interest: $\mathscr{S}(\mathbb{R}^n)$, the class of *admissible filters*, and $\mathscr{S}'(\mathbb{R}^n)$, the class of *admissible images*. Both spaces are closed under Fourier transformation. This means that a Fourier transformed filter is again an admissible filter, and a Fourier transformed image is again an admissible image. Most function spaces encountered in the image literature are inconsistent in this respect, with the notable exception of the space $L^2(\mathbb{R}^n)$, which is isomorphic to its own dual (i.e. may simultaneously model filters and images), and is likewise closed under Fourier transformation.

In fact, $L^2(\mathbb{R}^n)$ is a so-called Hilbert space. As such it plays an important role in quantum physics, which has analogies with image analysis to the extent that the observer participates in (and thereby determines) the observation. In quantum physics the observables are wave functions representing either a detecting device (cf. a filter) or a source field (cf. a raw image). In particular, linear interactions between detecting devices $\phi \in L^2(\mathbb{R}^n)$ and source fields $f \in L^2(\mathbb{R}^n)$ are expressed by inner products $\langle f|\phi\rangle$, which is similar to the way an image $f \in \mathscr{S}'(\mathbb{R}^n)$ and a filter $\phi \in \mathscr{S}(\mathbb{R}^n)$ conspire to produce a number $T_f[\phi]$ in a distributional model.

The interchangeability of source fields and detecting devices in quantum physics, reflected in the isomorphism between a Hilbert space and its dual, has no counterpart in our distribution-theoretical image/filter model. The paradigmatic difference is that, in image analysis, one typically designs filters according to some axiomatic principle, thereby imposing strong regularity constraints (although one *could* imagine filter profiles as complex as any potential image). Raw images, on the other hand, are empirically determined. Thus it makes little sense to impose regularity constraints, for (if this requires interventions altering the data, think of 'smoothing') this would amount to 'tampering with the evidence'. Thus there is a natural distinction between filters (probing devices) and images (physical evidence), which is manifest in a distribution-theoretical model.

The closure property of filter and image spaces under Fourier transformation, together with the one-to-one nature of the Fourier transform, implies that to every *spatial filter* we can associate a unique *Fourier filter*, and vice versa. Likewise, every *spatial image* has a unique *Fourier image*, vice versa. Closure guarantees that both filter as well as image model apply equally well to either domain. As a counterexample, suppose we would define an image to be a function of the type $f \in L^1(\mathbb{R}^n)$, which does not seem unreasonable at first glance. Then we will have to live with the discrepancy that its Fourier transform lives in a completely different space, viz. $\hat{f} \in C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$. In other words, either we accept that a Fourier image is not (necessarily) an image, or we decide upon two distinct definitions, one for spatial images and one for Fourier images, and explicitly specify which one we are dealing with throughout. A similar argument applies to filters. It is clear that it greatly simplifies our discussion to insist on Fourier closure.

Pragmatic people may perhaps find these considerations philosophical niggling, since *in practice* there exists only one operational type of function space. After all, discrete filters and images are invariably bounded functions on a compact subset of $\mathbb{Z}^n$. But this does not take away the fact that filters and images play entirely different roles in image analysis, nor that one may want to abstract from grid details by considering continuum representations. These conceptual considerations are reflected in our axioms.

4.3.2. The Fourier Transform on $\mathscr{S}(\mathbb{R}^n)$

In this section we concentrate on the definition of the Fourier transform of admissible filters.

Definition 4.3.1. *Fourier transformation is a continuous, linear, invertible mapping* $\mathscr{F} : \mathscr{S}(\mathbb{R}^n) \longrightarrow \mathscr{S}(\mathbb{R}^n)$:

$\phi \mapsto \widehat{\phi} \stackrel{\text{def}}{=} \mathscr{F}\phi$*, defined as follows:*

$$\widehat{\phi}(\omega) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} e^{-i\omega\cdot x}\,\phi(x)\,dx\,.$$

The attributes 'continuous' and 'invertible' require a rigorous proof. Instead, we concentrate on their interpretations. Continuity indicates that small filter perturbations $\delta\phi \in \mathscr{S}(\mathbb{R}^n)$—think of deviations from exact analytical profiles in numerical approximations—will induce small perturbations $\delta\widehat{\phi}$ (and likewise for the inverse, v.i.). Without going into details of what we mean by 'small', suffice it to say that this is an obvious a priori requirement for any numerical implementation to stand a chance of being successful, for one will never be able to implement a filter without some form of approximation, recall the discussion on ill-posedness. Invertibility tells us that we can unambiguously retrieve the spatial representation of a filter given its Fourier representation. The following conjecture tells us how. Again we refrain from a rigorous proof.

Conjecture 4.3.1. Cf. Definition 4.3.1. The *inverse Fourier transformation* is given by

$$\phi(x) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{i\omega\cdot x}\,\widehat{\phi}(\omega)\,d\omega\,.$$

Various alternative definitions are encountered in the literature. To appreciate how these are interrelated, consider the following modification.

Definition 4.3.2. *Recall Definition 4.3.1 and Conjecture 4.3.1. Setting* $\left(\mathscr{F}_{(a,b)}(\phi)\right)(\omega) \stackrel{\text{def}}{=} b\,\widehat{\phi}(a\omega)$*, for fixed* $a \in \mathbb{R}\backslash\{0\}$*,* $b \in \mathbb{R}^+$*, yields an alternative definition of the Fourier transform:*

$$\left(\mathscr{F}_{(a,b)}(\phi)\right)(\omega) \stackrel{\text{def}}{=} b\int_{\mathbb{R}^n} e^{-ia\omega\cdot x}\,\phi(x)\,dx \quad \textit{with inverse} \quad \left(\mathscr{F}^{-1}_{(a,b)}(\widetilde{\phi})\right)(x) = \frac{|a|^n}{(2\pi)^n b}\int_{\mathbb{R}^n} e^{ia\omega\cdot x}\,\widetilde{\phi}(\omega)\,d\omega\,.$$

In an attempt to overcome the confusion that may arise as a consequence of the many different standards used in the literature, let us adhere to the general (a,b)-convention.

Definition 4.3.3. *See Definition 4.3.2. We shall sometimes write* $\widetilde{u}$ *instead of* $\mathscr{F}_{(a,b)}(u)$ *for the Fourier transform of a function* u *in the general Fourier convention parametrised (implicitly) by* $(a,b) \in \mathbb{R}\backslash\{0\}\times\mathbb{R}^+$*, with default* $\widehat{u} = \mathscr{F} \equiv \mathscr{F}_{(1,1)}$*.*

Observation 4.3.1. See Definitions 4.3.2–4.3.3 for definitions. The forward and inverse Fourier transform are formally related by the following reparametrization:

$$\mathscr{T} : \mathbb{R}\backslash\{0\}\times\mathbb{R}^+ \longrightarrow \mathbb{R}\backslash\{0\}\times\mathbb{R}^+ : (a,b) \mapsto \mathscr{T}(a,b) \stackrel{\text{def}}{=} (a',b') \quad \text{given by} \quad \left\{\begin{array}{rcl} a' &=& -a \\ b' &=& \dfrac{|a|^n}{(2\pi)^n b}\,. \end{array}\right.$$

In other words:

$$\mathscr{F}^{-1}_{(a,b)} = \mathscr{F}_{(a',b')}\,.$$

Note that the transformation $\mathscr{T} = \mathscr{T}^{-1}$ equals its own inverse, allowing us to 'toggle' between forward and backward Fourier transforms.

Our choice $(a,b) = (1,1)$ is fairly common in pure mathematics and systems engineering. The choice made by Horn [16] in his monograph also agrees with ours. In their book, Gonzalez and Woods [12] use a definition that corresponds to the case $(a,b) = (2\pi,1)$, which is a common choice in signal processing. In the monograph by Florack [6] the convention is $(a,b) = (-1,1)$. Mathematica's default setting is $(a,b) = (-1,1/\sqrt{2\pi}^{\,n})$. In mathematical text books one often encounters the similar case $(a,b) = (1,1/\sqrt{2\pi}^{\,n})$. Some classical physics books use $(a,b) = (-1,1/(2\pi)^{\,n})$. Cf. Table 4.1 for a summary. By using the (a,b)-convention all these cases are covered, allowing us to translate all results from the literature on Fourier transforms directly to our default convention.

a	b	**examples of references**
$+1$	$+1$	our default, Horn [16]
-1	$+1$	Florack [6]
$+2\pi$	$+1$	Gonzalez and Woods [12]
-2π	$+1$	
$+1$	$1/\sqrt{2\pi}^{\,n}$	
-1	$1/\sqrt{2\pi}^{\,n}$	Mathematica [32]
$+1$	$1/(2\pi)^{\,n}$	
-1	$1/(2\pi)^{\,n}$	

Table 4.1: Summary of Fourier conventions frequently encounterd in the literature, and some references.

Example 4.3.1. Mathematicians tend to prefer the convention $(a, b) = (\pm 1, 1/\sqrt{2\pi}^{\,n})$. The reason may be the following. Since Fourier transformation is closed on $\mathscr{S}(\mathbb{R}^n)$, $L^2(\mathbb{R}^n)$, as well as on $\mathscr{S}'(\mathbb{R}^n)$, we may concatenate Fourier transformations when applied to a function from any of these classes. Let u be such a function, then, writing $\mathscr{F}^2_{(a,b)}$ instead of $\mathscr{F}_{(a,b)} \circ \mathscr{F}_{(a,b)}$, i.e. $\mathscr{F}_{(a,b)}$ applied twice in succession, we have

$$\mathscr{F}^2_{(a,b)}(u)(x) = \frac{b^2\,(2\pi)^n}{|a|^2}\,u(-x)\,.$$

The choice $(a^*, b^*) = (\pm 1, 1/\sqrt{2\pi}^{\,n})$ appears to be such that the amplification factor becomes unity. (In $L^2(\mathbb{R}^n)$ such a Fourier transformation is referred to as a *unitary*, i.e. norm-preserving transformation.) One readily observes that if we apply the transformation $2k$ times in succession, then

$$\mathscr{F}^{2k}_{(a^*,b^*)}(u)(x) = u((-1)^k\,x) \quad (k \in \mathbb{N})\,.$$

In particular, $\mathscr{F}^2_{(a^*,b^*)}$ 'mirrors' the domain of the function to which it is applied, whereas $\mathscr{F}^4_{(a^*,b^*)}$ reduces to the identity operator.

Another advantage of Observation 4.3.1, besides its unifying potential, is that Fourier forward and backward transformations can be implemented as a *single* algorithm.

Example 4.3.2. Mathematica's built-in function `FourierTransform` (version 5.2) has an optional argument for specifying the Fourier convention [32]. The default setting is `FourierParameters->{0,1}`. Its relation to Definitions 4.3.2–4.3.3 is as follows. If we take `FourierParameters->{`α`,`β`}`, then we have in n dimensions

$$\left\{\begin{array}{rcl} a & = & -\beta \\ b & = & \left(\dfrac{|\beta|}{(2\pi)^{1-\alpha}}\right)^{n/2} \end{array}\right. .$$

Inverting this system, we may solve for Mathematica's parameters (α, β) in terms of our (a, b), yielding

$$(\alpha, \beta) = (1 + {}^{2\pi}\log(\sqrt[n]{b^2}/|a|), -a)\,.$$

In particular, our default convention $(a, b) = (1, 1)$ should be specified as `FourierParameters->{1,-1}` in Mathematica, i.e. $(\alpha, \beta) = (1, -1)$. Reversely, Mathematica's default setting corresponds to our convention $(a, b) = (-1, 1/\sqrt{2\pi}^{\,n})$, recall Table 4.1. Observation 4.3.1 may be reformulated in terms of a mapping of the Mathematica parameters (α, β).

By definition the Fourier transform of a real-valued filter is, in general, complex-valued. You may readily verify that if $\phi \in \mathscr{S}(\mathbb{R}^n)$ is even (or symmetric), i.e. if $\phi(x) = \phi(-x)$ for all $x \in \mathbb{R}^n$, then $\widehat{\phi}$ is real-valued. Likewise, if $\phi \in \mathscr{S}(\mathbb{R}^n)$ is odd (or antisymmetric), i.e. if $\phi(x) = -\phi(-x)$ for all $x \in \mathbb{R}^n$, then $\widehat{\phi}$ is purely imaginary-valued.

Filters are infinitely differentiable functions, and one may wonder how the Fourier transformation of a filter derivative is related to that of the filter itself.

Lemma 4.3.1. *Let $\phi \in \mathscr{S}(\mathbb{R}^n)$, and let $\widehat{\phi} \in \mathscr{S}(\mathbb{R}^n)$ be its Fourier transform according to Definition 4.3.1. Moreover, let $P(c, \nabla) = \sum_{\alpha, |\alpha| \leq k} c_\alpha \, \nabla_\alpha$ be a k-th order linear partial derivative operator with constant coefficients c_α, in which the sum extends over all multi-indices α of order $|\alpha| = 0, \ldots, k$. Then we have*

$$\mathscr{F}\left(P(c, \nabla)\, \phi\right)(\omega) = P(c, i\omega)\, \widehat{\phi}(\omega) \quad (\omega \in \mathbb{R}^n)\,.$$

Proof. The basic observation underlying Lemma 4.3.1 is the fact that $\nabla_\alpha e^{i\omega \cdot x} = (i\omega)^\alpha \, e^{i\omega \cdot x}$, from which it follows by linearity that $P(c, \nabla)\, e^{i\omega \cdot x} = P(c, i\omega)\, e^{i\omega \cdot x}$. In the terminology of linear algebra this states that the complex 'planar wave' function $x \mapsto e^{i\omega \cdot x}$—i.e. considered as a function of x with parameter ω—is an *eigenfunction of the linear differential operator* $P(c, \nabla)$ with *eigenvalue* $P(c, i\omega)$. Using this, it follows that

$$P(c, \nabla)\, \phi(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} P(c, \nabla)\, e^{i\omega \cdot x}\, \widehat{\phi}(\omega)\, d\omega = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\omega \cdot x}\, P(c, i\omega)\, \widehat{\phi}(\omega)\, d\omega\,.$$

The first equality follows by Conjecture 4.3.1 after interchanging the order of differentiation and integration. Inspection of the result, again with the help of Conjecture 4.3.1, shows that the function $\omega \mapsto P(c, i\omega)\, \widehat{\phi}(\omega)$ is the Fourier transform of $x \mapsto P(c, \nabla)\, \phi(x)$. □

Example 4.3.3. Let us consider a few simple cases for $n = 1$. According to Lemma 4.3.1 we have, for $\omega \in \mathbb{R}$,

$$\begin{aligned}
\mathscr{F}\left(\frac{d\phi}{dx}\right)(\omega) &= i\omega\, \widehat{\phi}(\omega)\,, \\
\mathscr{F}\left(\frac{d^2\phi}{dx^2}\right)(\omega) &= -\omega^2\, \widehat{\phi}(\omega)\,, \\
\mathscr{F}\left(\frac{d^3\phi}{dx^3}\right)(\omega) &= -i\omega^3\, \widehat{\phi}(\omega)\,,
\end{aligned}$$

and so forth. The general one-dimensional scheme should now be obvious:

$$\mathscr{F}\left(\frac{d^k\phi}{dx^k}\right)(\omega) = (i\omega)^k\, \widehat{\phi}(\omega)\,.$$

By linearity, if we substitute any linear combination of derivatives of ϕ, we obtain a corresponding linear combination of the respective Fourier transforms.

In higher dimensions results are similar, only now we have a frequency vector $\omega \in \mathbb{R}^n$ rather than a scalar frequency, let us write $\omega = (\omega_1, \ldots, \omega_n)$. Also, we may now consider any linear combination of partial derivatives with respect to any of the spatial coordinates $x^1, \ldots, x^n$.

Example 4.3.4. Recall Lemma 4.3.1, and suppose $n = 2$. Writing $\omega = (\omega_x, \omega_y)$ in this case, Lemma 4.3.1 yields:

$$\begin{aligned}
\mathscr{F}\left(\frac{\partial\phi}{\partial x}\right)(\omega) &= i\omega_x\, \widehat{\phi}(\omega) && (c_\alpha = 1 \text{ if } \alpha = (1,0),\, c_\alpha = 0 \text{ otherwise,}) \\
\mathscr{F}\left(\frac{\partial^2\phi}{\partial x \partial y}\right)(\omega) &= -\omega_x\, \omega_y\, \widehat{\phi}(\omega) && (c_\alpha = 1 \text{ if } \alpha = (1,1),\, c_\alpha = 0 \text{ otherwise,}) \\
\mathscr{F}\left(\frac{\partial^3\phi}{\partial x \partial y^2}\right)(\omega) &= -i\omega_x\, \omega_y^2\, \widehat{\phi}(\omega) && (c_\alpha = 1 \text{ if } \alpha = (1,2),\, c_\alpha = 0 \text{ otherwise.})
\end{aligned}$$

The general rule for partial derivatives is now obvious:

$$\mathscr{F}\left(\frac{\partial^{p+q}\phi}{\partial x^p \partial y^q}\right)(\omega) = (i\omega_x)^p\, (i\omega_y)^q\, \widehat{\phi}(\omega) \qquad (c_\alpha = 1 \text{ if } \alpha = (p,q),\, c_\alpha = 0 \text{ otherwise.})$$

Example 4.3.5. Linear combinations of partial derivatives map to corresponding linear combinations of their Fourier transforms. As an example, consider the Laplacian $\Delta\phi$ of ϕ. Fourier transformation yields

$$\mathscr{F}\left(\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}\right)(\omega) = -\left(\omega_x^2+\omega_y^2\right)\widehat{\phi}(\omega)\,.$$

A special case of Lemma 4.3.1, illustrated in one and two dimensions in the previous examples, is worth remembering, for the general case follows by linearity.

Result 4.3.1. *Let $\phi \in \mathscr{S}(\mathbb{R}^n)$, and let $\widehat{\phi} \in \mathscr{S}(\mathbb{R}^n)$ be its Fourier transform according to Definition 4.3.1. Then for integer orders $k_1,\ldots,k_n \geq 0$ we have, for $\omega = (\omega_1,\ldots,\omega_n) \in \mathbb{R}^n$,*

$$\mathscr{F}\left(\frac{\partial^{k_1+\ldots+k_n}\phi}{\partial x^{k_1}\ldots\partial x^{k_n}}\right)(\omega) = (i\omega_1)^{k_1}\ldots(i\omega_n)^{k_n}\,\widehat{\phi}(\omega)\,.$$

Alternatively, using mult-index notation,

$$\mathscr{F}\left(\nabla_\alpha\phi\right)(\omega) = (i\omega)^\alpha\,\widehat{\phi}(\omega)\,.$$

Lemma 4.3.1 and Result 4.3.1 may be generalized to account for the (a,b)-convention of Definition 4.3.2.

Example 4.3.6. The significance of *fractional powers of differential operators* is most clearly demonstrated in Fourier space. As an example, consider the linear operator $\sqrt{-\Delta}$, in which Δ is the Laplacean operator. A previous example has revealed that $\mathscr{F}\left(\Delta\phi\right)(\omega) = -\|\omega\|^2\mathscr{F}(\phi)$, so we may interpret the Laplacean operator in Fourier space, formally written as $\mathscr{F}(\Delta)$, as the multiplication operator $\big(\omega\mapsto f(\omega)\big)\mapsto\big(\omega\mapsto -\|\omega\|^2 f(\omega)\big)$. Then, by linearity, we have $\mathscr{F}(-\Delta) = \big(\omega\mapsto f(\omega)\big)\mapsto\big(\omega\mapsto \|\omega\|^2 f(\omega)\big)$. It appears reasonable to then define $\mathscr{F}(\sqrt{-\Delta}) = (\omega\mapsto f(\omega))\mapsto(\omega\mapsto\|\omega\|f(\omega))$. This heuristic argument at the same time explains why the Laplacean operator Δ is considered a 'negative operator'.

4.3.3. The Fourier Transform on $\mathscr{S}'(\mathbb{R}^n)$

In this section we concentrate on the definition of the Fourier transform of admissible images. We will allow ourselves some sloppiness in manipulating formulas without rigorous justification, such as changing integration order, writing and manipulating images as functions instead of proper distributions, etc. Accepting this level of rigor, the Fourier transform and its inverse take the same form as before.

Definition 4.3.4. *Fourier transformation is a continuous, linear, invertible mapping $\mathscr{F} : \mathscr{S}'(\mathbb{R}^n) \longrightarrow \mathscr{S}'(\mathbb{R}^n) : f \mapsto \widehat{f} \stackrel{\text{def}}{=} \mathscr{F}f$, defined as follows:*

$$\widehat{f}(\omega) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} e^{-i\omega\cdot x}\,f(x)\,dx\,.$$

Again we stress that the inverse Fourier transform follows unambiguously from the forward Fourier transform, but defer the proof until later.

Conjecture 4.3.2. Cf. Definition 4.3.4. The *inverse Fourier transformation* is given by

$$f(x) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{i\omega\cdot x}\,\widehat{f}(\omega)\,d\omega\,.$$

Similar alternative definitions apply as in Definition 4.3.2. Bear in mind that images are not defined point-wise, but only by virtue of being 'filterable'. In other words, functions like f and $\widehat{f}$ should only be used 'under the

integral' together with an appropriate filter. Still, abuse of notation permits us to write things like $\widehat{f}(\omega)$ and $f(x)$, even though such point evaluations make no sense. This should not cause any confusion as long as you realize that you are in fact dealing with tempered distributions.

The attribute 'continuous' in Definition 4.3.4 is particularly important in this case in view of the fact that images are measurements with intrinsically finite tolerances.

Example 4.3.7. Let $\widehat{u}_s : \mathbb{R} \longrightarrow \mathbb{R} : \omega \mapsto \widehat{u}_s(\omega) = e^{-s|\omega|}$, in which $s > 0$ is a positive constant. Then

$$u_s(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega x}\, \widehat{u}_s(\omega)\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-s|\omega| + i\omega x}\, d\omega = \frac{1}{2\pi} \left(\int_{-\infty}^{0} e^{(s+ix)\omega}\, d\omega + \int_{0}^{\infty} e^{(-s+ix)\omega}\, d\omega \right) .$$

We have

$$\int_{-\infty}^{0} e^{(s+ix)\omega}\, d\omega = \left[\frac{1}{s+ix} e^{(s+ix)\omega} \right]_{\omega \to -\infty}^{\omega=0} = \frac{1}{s+ix} .$$

Similarly,

$$\int_{0}^{\infty} e^{(-s+ix)\omega}\, d\omega = \left[\frac{1}{-s+ix} e^{(-s+ix)\omega} \right]_{\omega=0}^{\omega \to +\infty} = \frac{1}{s-ix} ,$$

so that we finally obtain

$$u_s(x) = \frac{1}{2\pi} \left(\frac{1}{s+ix} + \frac{1}{s-ix} \right) = \frac{1}{\pi} \frac{s}{x^2+s^2} .$$

That we obtain a real-valued solution is no coincidence. You should be able to figure out yourself how we might have anticipated this. (Hint: $\widehat{u}_s(\omega) = \widehat{u}_s(-\omega)$.)

As we have seen, the variety of images is significantly larger than that of filters ($\mathscr{S}'(\mathbb{R}^n)$ is significantly larger than $\mathscr{S}(\mathbb{R}^n)$). In particular one may wonder what the Fourier transformation of the Dirac point distribution looks like. This follows directly from Definition 4.3.4 by taking $f(x) = \delta(x)$.

Lemma 4.3.2. *The Fourier transform of the Dirac point distribution at the origin, $x \mapsto \delta(x)$, is a constant function. In default Fourier convention we have*

$$\mathscr{F}\left(\delta\right)(\omega) = 1 .$$

Proof. Apply the definition of the Dirac point distribution[3], Lemma 3.3.1, Page 50.

An equivalent lemma expresses the Fourier inverse formulation.

Lemma 4.3.3. *The inverse Fourier transform of a constant function is proportional to the Dirac point distribution at the origin, $\delta(x)$. In default Fourier convention we have*

$$\mathscr{F}^{-1}\left(1\right)(x) = \delta(x) .$$

Note that the symbol 1 here actually stands for $1 : \mathbb{R}^n \longrightarrow \mathbb{R} : \omega \mapsto 1$.

Proof. This is basically the hitherto unproven Conjecture 4.3.2, with $\widehat{f}(\omega) = 1$. □

Note that the lemma is equivalent to the following result, which is a useful one to remember, because it does not depend on the particular Fourier convention:

$$\int_{\mathbb{R}^n} e^{\pm i\omega\cdot x}\, d\omega = (2\pi)^n\, \delta(x) . \tag{4.10}$$

[3]Note that $e^{i\omega x}$ is not a Schwartz function. Nevertheless, it is clear that the Dirac point distribution can be generalised to any function space in which point evaluation makes sense.

Both Lemma 4.3.3 and Eq. (4.10) express the fact that the Dirac point distribution can be thought of as a linear superposition of 'planar waves' of all possible frequencies, all with the same amplitude.

You will encounter Eq. (4.10) in various disguises. Here is an example:

$$\int_{\mathbb{R}^n} e^{\pm i\omega\cdot x}\,dx = (2\pi)^n\,\delta(\omega)\,. \tag{4.11}$$

Of course this is *exactly* the same equation as the previous one, with merely a change of symbols. In view of our generalized Fourier convention, substitution of variables, $y = |a|x$ in Eq. (4.11) yields the following useful generalization:

Result 4.3.2. *For fixed $a \in \mathbb{R}\backslash\{0\}$ we have* $\displaystyle\int_{\mathbb{R}^n} e^{\pm ia\omega\cdot x}\,dx = \frac{(2\pi)^n}{|a|^n}\,\delta(\omega)$.

We refrain from a proof here. Note that Conjectures 4.3.1 and 4.3.2 can be proven using this result.

In the next section we concentrate on operational issues that will eventually lead us to the *discrete Fourier transform*, and its inverse, in Section 4.5.

4.3.4. Algebraic Fourier Theorems

Recall Definitions 4.3.2–4.3.3. In the following theorems and results $u_{1,2} : \mathbb{R}^n \longrightarrow \mathbb{K}$ and $\widetilde{u}_{1,2} : \mathbb{R}^n \longrightarrow \mathbb{K}$ are suitably defined functions, meaning that all stipulated products and convolutions are well-defined and have well-defined (inverse) Fourier transforms. (The $\tilde{}$ is inessential and merely indicates that the corresponding function is to be interpreted as a function in Fourier space.)

Theorem 4.3.1 (Fourier Transform of a Function Product)**.**

$$\mathscr{F}_{(a,b)}(u_1u_2) = \frac{|a|^n}{(2\pi)^n\,b}\,\mathscr{F}_{(a,b)}(u_1) * \mathscr{F}_{(a,b)}(u_2)\,.$$

In particular, following our convention $(a, b) = (1, 1)$, we have

$$\mathscr{F}(u_1u_2) = \frac{1}{(2\pi)^n}\,\mathscr{F}(u_1) * \mathscr{F}(u_2)\,.$$

This theorem explains the popularity of the Fourier convention in which $(a, b) = (\pm 2\pi, 1)$, or $(a, b) = (\pm 1, 1/(2\pi)^n)$.

Proof. For notational convenience we set $u \equiv u_1u_2$ and $\mathscr{F}_{(a,b)}(u) \equiv \widetilde{u}$ in this proof. We have

$$\widetilde{u}(\omega) \stackrel{\text{def}}{=} b\int_{\mathbb{R}^n} u_1(x)\,u_2(x)\,e^{-ia\omega\cdot x}\,dx\,.$$

Substituting the respective inverse Fourier transforms for $u_{1,2}$, i.e.

$$u_{1,2}(x) = \frac{|a|^n}{(2\pi)^n}\frac{1}{b}\int_{\mathbb{R}^n} e^{ia\omega\cdot x}\,\widetilde{u}_{1,2}(\omega)\,d\omega\,,$$

yields

$$\widetilde{u}(\omega) = b\int_{\mathbb{R}^n}\left[\frac{|a|^n}{(2\pi)^n}\frac{1}{b}\int_{\mathbb{R}^n} e^{ia\omega_1\cdot x}\,\widetilde{u}_1(\omega_1)\,d\omega_1\right]\left[\frac{|a|^n}{(2\pi)^n}\frac{1}{b}\int_{\mathbb{R}^n} e^{ia\omega_2\cdot x}\,\widetilde{u}_2(\omega_2)\,d\omega_2\right]\,e^{-ia\cdot\omega x}\,dx\,.$$

We may rewrite this expression as

$$\widetilde{u}(\omega) = \frac{|a|^n}{(2\pi)^n}\frac{1}{b}\int_{\mathbb{R}^n}\frac{|a|^n}{(2\pi)^n}\frac{1}{b}\int_{\mathbb{R}^n}\widetilde{u}_1(\omega_1)\,\widetilde{u}_2(\omega_2)\left[b\int_{\mathbb{R}^n} e^{ia(\omega_1+\omega_2-\omega)\cdot x}\,dx\right]d\omega_1\,d\omega_2\,,$$

provided we are allowed to interchange the order of integrations. (We take this for granted.) We recognize the innermost integral as a variant of Result 4.3.2. Thus we conclude that

$$\widetilde{u}(\omega) = \frac{|a|^n}{(2\pi)^n\,b}\int_{\mathbb{R}^n}\widetilde{u}_1(\omega_1)\,\widetilde{u}_2(\omega-\omega_1)\,d\omega_1\,,$$

in other words,

$$\widetilde{u}(\omega) = \frac{|a|^n}{(2\pi)^n\,b}\,\left(\widetilde{u}_1 * \widetilde{u}_2\right)(\omega)\,.$$

The rest follows by inspection of this result. □

An immediate result of Theorem 4.3.1 is the following.

Theorem 4.3.2 (Fourier Transform of a Convolution Product)**.**

$$\mathscr{F}_{(a,b)}\,(u_1 * u_2) = \frac{1}{b}\mathscr{F}_{(a,b)}(u_1)\mathscr{F}_{(a,b)}(u_2)\,.$$

In particular, following our convention $(a,b) = (1,1)$*, we have*

$$\mathscr{F}\,(u_1 * u_2) = \mathscr{F}(u_1)\mathscr{F}(u_2)\,.$$

Proof. Set $v_{1,2} = \mathscr{F}_{(a,b)}(u_{1,2})$. Theorem 4.3.1 applied to the pair $v_{1,2}$ states that

$$\mathscr{F}_{(a',b')}(v_1 v_2) = \frac{|a'|^n}{(2\pi)^n\,b'}\,\left(\mathscr{F}_{(a',b')}(v_1) * \mathscr{F}_{(a',b')}(v_2)\right)\,.$$

for all $(a',b') \in \mathbb{R}\backslash\{0\}\times\mathbb{R}^+$. By Observation 4.3.1, Page 66, we may rewrite this as

$$\mathscr{F}^{-1}_{(a,b)}\left(\mathscr{F}_{(a,b)}(u_1)\mathscr{F}_{(a,b)}(u_2)\right) = \frac{|a'|^n}{(2\pi)^n\,b'}\,\left(\mathscr{F}^{-1}_{(a,b)}(\mathscr{F}_{(a,b)}(u_1)) * \mathscr{F}^{-1}_{(a,b)}(\mathscr{F}_{(a,b)}(u_2))\right) = b\,(u_1 * u_2)\,.$$

Application of $\mathscr{F}_{(a,b)}$ to left and right hand sides completes the proof. □

Theorems 4.3.1–4.3.2 are powerful results that are often used in the analysis of linear shift invariant, i.e. convolution systems. A diagrammatic representation of Theorem 4.3.1 is

$$\begin{array}{ccc} (u_1,u_2) & \longrightarrow & u_1\,u_2 \\ {\scriptstyle\mathscr{F}_{(a,b)}}\downarrow & & \uparrow{\scriptstyle\mathscr{F}^{-1}_{(a,b)}} \\ (\mathscr{F}_{(a,b)}(u_1),\mathscr{F}_{(a,b)}(u_2)) & \longrightarrow & \dfrac{|a|^n}{(2\pi)^n b}\,\mathscr{F}_{(a,b)}(u_1) * \mathscr{F}_{(a,b)}(u_2)\,, \end{array} \tag{4.12}$$

whereas that of Theorem 4.3.2 is

$$\begin{array}{ccc} (u_1,u_2) & \longrightarrow & u_1 * u_2 \\ {\scriptstyle\mathscr{F}_{(a,b)}}\downarrow & & \uparrow{\scriptstyle\mathscr{F}^{-1}_{(a,b)}} \\ (\mathscr{F}_{(a,b)}(u_1),\mathscr{F}_{(a,b)}(u_2)) & \longrightarrow & \dfrac{1}{b}\,\mathscr{F}_{(a,b)}(u_1)\,\mathscr{F}_{(a,b)}(u_2)\,. \end{array} \tag{4.13}$$

In each diagram the upper horizontal route expresses an algebraic operation in the spatial domain (ordinary product, respectively convolution). The lower horizontal route expresses the equivalent operation in Fourier

space. For each operation (moving from one corner in the diagram to another) there are always two distinct but equivalent alternatives. Note that you may reverse the vertical arrows if you replace $\mathscr{F}_{(a,b)}$ by $\mathscr{F}^{-1}_{(a,b)}$, vice versa.

The diagrams suggests how one could efficiently implement numerical schemes for convolution (either in the spatial domain or in Fourier space), viz. by avoiding the convolution corner in the applicable diagram. After all, convolution is a multi-local operation, whereas multiplication acts point-wise, which potentially saves us a significant amount of computation time. Of course, the concomitant cost of the forward and inverse Fourier transformations encountered along the detour one has to make in the diagram has to be taken into consideration as well. Still, in certain cases this may be the most efficient way to compute a convolution product.

Other equivalent formulations of Theorem 4.3.1 and 4.3.2 are stated in the following results.

Result 4.3.3 (Fourier Transform of a Function Product in Fourier Space)**.**

$$\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_1)\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_2) = \frac{|a|^n}{(2\pi)^n\, b}\,\mathscr{F}^{-1}_{(a,b)}\left(\widetilde{u}_1 * \widetilde{u}_2\right) .$$

In particular, following our convention $(a, b) = (1, 1)$*, we have*

$$\mathscr{F}^{-1}(\widetilde{u}_1)\mathscr{F}^{-1}(\widetilde{u}_2) = \frac{1}{(2\pi)^n}\,\mathscr{F}^{-1}\left(\widetilde{u}_1 * \widetilde{u}_2\right) .$$

Proof. This is basically Theorem 4.3.1, with $u_{1,2} = \mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_{1,2})$, and with $\mathscr{F}^{-1}_{(a,b)}$ applied to both sides. □

Result 4.3.4 (Fourier Transform of a Convolution Product in Fourier Space)**.**

$$\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_1) * \mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_2) = \frac{1}{b}\,\mathscr{F}^{-1}_{(a,b)}\left(\widetilde{u}_1\widetilde{u}_2\right) .$$

In particular, following our convention $(a, b) = (1, 1)$*, we have*

$$\mathscr{F}^{-1}(\widetilde{u}_1) * \mathscr{F}^{-1}(\widetilde{u}_2) = \mathscr{F}^{-1}\left(\widetilde{u}_1\widetilde{u}_2\right) .$$

Proof. This is Theorem 4.3.2, subject to the same adaptation as in the proof of the previous result. □

Diagrams similar to Diagrams (4.12–4.13) can be constructed for the latter two results.

4.4. Intermezzo: Towards an Operational Definition

There are two unrealistic aspects of the previously discussed Fourier transform, viz. the assumption of infinite function domain, and the assumption that this domain is a continuum rather than a discrete grid. In other words, these Fourier models tacitly assume that there are no limiting constraints on the possible scales involved. Scales range from infinitesimal (implicit in the continuum assumption) to infinity (due to infinite support).

Recall furthermore that the filter and image spaces have been carefully selected so as to meet certain consistency demands, such as the requirement of closure under convolution and Fourier transformation. Thus we cannot simply truncate and discretize the Fourier model *just like that* and expect it still to satisfy the same basic consistency demands. Instead, we have to investigate carefully what happens if we truncate and discretize our model.

Instead of maintaining the conceptual difference between filters and images throughout, it will simplify our discussion to handle these collectively as 'suitably defined' functions. The reason for this is that, unlike previously,

realistic representations of both filters as well as images on a finite grid will constitute finite vector spaces isomorphic to some $\mathbb{K}^N$ (typically for very large number of pixels, $N \in \mathbb{N}$). Recall that images live in the space dual to that of the filters. The mathematical justification for handling discrete filters and images on the same foot is the fact that the topological dual of a finite-dimensional vector space is isomorphic to the vector space itself.

We proceed as follows. We adhere to the established Fourier paradigm throughout. In Section 4.4.1, however, we consider its implications when applied to filters and images of finite size[4]. The spatial domain of definition will still be considered a continuum though. We will see that this does not lead to a consistent theory (Fourier transformation will take us out of our defining function space), but on the fly we establish some useful results.

Subsequently we impose the constraint that the images are defined on a discrete grid in Section 4.4.2, but now we disregard the effect of truncation. Again it will become apparent that we have to sacrifice consistency, although results as such will be instructive.

In Sections 4.4.3 and 4.4.4 we exploit the analogies between forward and inverse Fourier transforms in combination with previous results in order to understand the effects of truncation, respectively discretization in the frequency domain.

We combine both constraints in Section 4.5. It will turn out that if both discretization as well as truncation are taken into account, then not only do we obtain an operationally well-defined Fourier transform, but also restore consistency! The resulting Fourier transform is known as the *Discrete Fourier Transform*, or *DFT*.

For simplicity we restrict ourselves to the one-dimensional case when discussing intermediate results in Sections 4.4.1–4.4.4.

4.4.1. Fourier Transformation on a Finite Spatial Domain

Recall Definitions 4.3.1, Page 65, and 4.3.4, Page 69, and Theorems 4.3.1, Page 66, and 4.3.2, Page 69. These also apply to finite support functions, as long as they are within the respective function classes. Let $u : [-L/2, L/2] \longrightarrow \mathbb{K}$ be such a function. Then in the general Fourier convention of Definition 4.3.2 we have

$$\mathscr{F}_{(a,b)}(u)(\omega) = b \int_{-L/2}^{L/2} u(x)\, e^{-ia\omega x}\, dx\,. \tag{4.14}$$

To understand the impact of a finite integration interval on the Fourier representation of a function, we could make the following assumption.

Assumption 4.4.1. *A finite support function $u_0 : [-L/2, L/2] \longrightarrow \mathbb{K}$ is a windowed version of an infinite support function, $u : \mathbb{R} \longrightarrow \mathbb{K}$, say:*

$$u_0(x) \stackrel{\text{def}}{=} u(x)\, \chi_{[-L/2,L/2]}(x)\,,$$

in which $\chi_{[-L/2,L/2]}$ is the indicator function on the interval $[-L/2, L/2]$.

The subscript reminds us of the 'zero-padding' outside the original, finite domain of definition. Thus

$$u_0(x) \stackrel{\text{def}}{=} \begin{cases} u(x) & \text{if } x \in [-L/2, L/2], \\ 0 & \text{otherwise.} \end{cases} \tag{4.15}$$

[4] In $\mathscr{S}'(\mathbb{R}^n)$ this can be accomplished simply by multiplication of the image with a suitable characteristic function of compact support, in $\mathscr{S}(\mathbb{R}^n)$ this requires a more subtle procedure using smooth 'bump functions' instead. The details are beyond our scope.

Another choice that is of interest for reasons yet to be discussed is the following. The *entier* of a real number $x \in \mathbb{R}$, notation $\lfloor x \rfloor$, denotes the largest integer i smaller than or equal to x. In other words, $\lfloor x \rfloor = i \in \mathbb{Z}$ if $i \leq x$ but $i + 1 > x$. With this definition we introduce

$$u_p(x) \stackrel{\text{def}}{=} \begin{cases} u(x) & \text{if } x \in [-L/2, L/2], \\ u(x - kL) & \text{if } x \notin [-L/2, L/2], \text{ and } k = \lfloor (x + L/2)/L \rfloor. \end{cases} \tag{4.16}$$

Thus u_p is a periodic replication of u_0 outside the interval $[-L/2, L/2]$ with period L, whence the mnemonic subscript. Note that by virtue of the definition of $k \in \mathbb{Z}$, $x - kL \in [-L/2, L/2]$, thus $u_0(x - kL)$ is well-defined.

In summary, we can always write the extended function as $u = u_0 + u_1$, i.e. the sum of a nontrivial, zero-padded function u_0, and a function u_1 that vanishes on the interval of interest, but is otherwise arbitrary. In the case of zero-padding we have $u_1 = 0$, and in the case of periodic replication, $u_1 = u_p - u_0$.

Of course u_0 and u_p are entirely different functions and therefore they have different Fourier transforms, as does any alternative extension. Somehow, the *relevant* information they convey is the same, and so we must take the differences of their Fourier transforms with a grain of salt, and figure out what the effects of the respective domain extensions are on the Fourier representations. This will tell us how the relevant spatial information (i.e. the spatial information actually contained in the finite support image) manifests itself in Fourier space, and how it can be segregated from the irrelevant information that merely expresses the effect of the arbitrary domain extension (in particular zero-padding or periodic replication).

4.4.1.1 The Effect of Zero-Padding in the Spatial Domain

The effect of zero-padding follows from Theorems 4.3.1 and 4.3.2 or Results 4.3.3 and 4.3.4. Since a zero-padded one-dimensional function can be written as a product of an arbitrarily extended function and the characteristic function on the support interval, $\chi_{[-L/2,L/2]}$, we can use Theorem 4.3.1 to compute its Fourier transform. According to the theorem the Fourier transform of $u_0 = u\,\chi_{[-L/2,L/2]}$ is a convolution product in Fourier space.

Definition 4.4.1. *The function* sinc *is defined as*

$$\operatorname{sinc} x \stackrel{\text{def}}{=} \begin{cases} \dfrac{\sin x}{x} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0. \end{cases}$$

(Note that $\lim_{x \longrightarrow 0} \operatorname{sinc} x = 1$*, so the* sinc *function is actually continuous.)*

Theorem 4.4.1. *The Fourier transform of a compact support function* $u_0 = u\,\chi_{[-L/2,L/2]}$ *is given by*

$$\mathscr{F}_{(a,b)}(u_0)(\omega) = \frac{|a|}{2\pi b} \left(\mathscr{F}_{(a,b)}(u) * \mathscr{F}_{(a,b)}(\chi_{[-L/2,L/2]}) \right)(\omega)\,,$$

with

$$\mathscr{F}_{(a,b)}(\chi_{[-L/2,L/2]})(\omega) = b\,L \operatorname{sinc}\left(\frac{a\omega L}{2} \right).$$

Proof. The proof is left as an exercise. □

Apparently the Fourier transform of the indicator function on a finite interval is, unlike the indicator function itself, *not* a compact support function. Figure 4.1 shows the graph of the sinc -function in Theorem 4.4.1 for several interval widths L. Note that its zero-crossings are found at discrete frequencies

$$\omega_k \stackrel{\text{def}}{=} k\omega_1 \quad \text{for } k \in \mathbb{Z}\backslash\{0\}, \text{ and ground frequency} \quad \omega_1 \stackrel{\text{def}}{=} \frac{2\pi}{|a|L}\,. \tag{4.17}$$

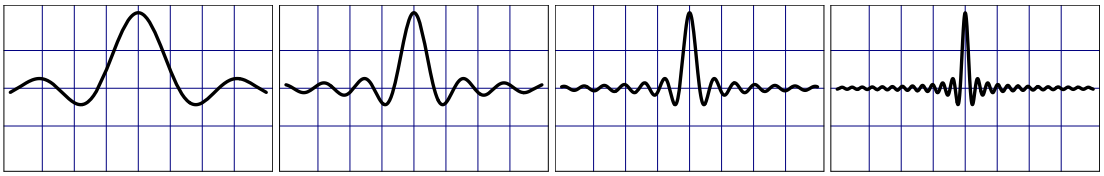


Figure 4.1: The function $\widehat{\chi}_{[-L/2,L/2]}(\omega) = L\,\mathrm{sinc}\,(\omega L/2)$ for various interval widths. From left to right: $L = 1, 2, 4, 8$. Graphs have been scaled: Their peaks have (increasing) magnitudes $L = 1, 2, 4, 8$.

Clearly we have

$$\lim_{L\longrightarrow\infty} u_0(x) = u(x)\,,$$

or, put differently,

$$\lim_{L\longrightarrow\infty} \chi_{[-L/2,L/2]}(x) = 1\,,$$

for all $x \in \mathbb{R}$. In which form does this limiting behavior manifest itself in the Fourier domain? According to Definition 4.3.4 and Result 4.3.2 the Fourier transform of the above spatial limiting function is proportional to the Dirac point distribution, viz.

$$\mathscr{F}_{(a,b)}\left(\lim_{L\longrightarrow\infty} \chi_{[-L/2,L/2]}\right)(\omega) = \mathscr{F}_{(a,b)}\,(1)\,(\omega) = \frac{2\pi b}{|a|}\,\delta(\omega)\,.$$

By virtue of continuity of Fourier transformation this is also equal to

$$\lim_{L\longrightarrow\infty} \mathscr{F}_{(a,b)}\left(\chi_{[-L/2,L/2]}\right)(\omega) = \lim_{L\longrightarrow\infty} b\,\mathrm{sinc}\left(\frac{a\omega L}{2}\right)\,. \tag{4.18}$$

In other words,

$$\lim_{L\longrightarrow\infty} b\,\mathrm{sinc}\left(\frac{a\omega L}{2}\right) = \frac{2\pi b}{|a|}\,\delta(\omega)\,. \tag{4.19}$$

The latter identity expresses the theoretical limit of what can basically be observed in Figure 4.1. Whereas the indicator function evolves towards a constant function as $L \longrightarrow \infty$, its Fourier transform develops into a spike at the origin. Indeed, the Fourier transform in Theorem 4.4.1 has the following limit:

$$\lim_{L\longrightarrow\infty} \mathscr{F}_{(a,b)}(u_0)(\omega) = \lim_{L\longrightarrow\infty} \frac{|a|}{2\pi b}\,\left(\mathscr{F}_{(a,b)}(u) * \mathscr{F}_{(a,b)}(\chi_{[-L/2,L/2]})\right)(\omega) = (\mathscr{F}_{(a,b)}(u)*\delta)(\omega) = \mathscr{F}_{(a,b)}(u)(\omega)\,.$$

In the second step we have tacitly assumed that it is allowed to take the $\lim$-operator inside the convolution integral, and used Eqs. (4.18–4.19).

In summary, Theorem 4.4.1 leads us to conclude that *the Fourier transform of a zero-padded function equals the Fourier transform of an arbitrarily extended, not necessarily compactly supported function, 'blurred' with a* sinc*-function, the width of which is inversely proportional to the size of the support interval*. Thus the physical constraint of a finite spatial scope manifests itself as blurring in the frequency domain. More precisely, neighbouring frequencies separated by a distance

$$|\Delta\omega| \leq \frac{2\pi}{|a|L}\,, \tag{4.20}$$

in Fourier space, i.e. half the distance between the innermost zero-crossings of the sinc-function, cf. Figure 4.1 and Theorem 4.4.1, are 'smeared out' in Fourier space and cannot be resolved. The right hand side of Eq. (4.20) is a measure of the *frequency blur* induced by the spatial cut-off stipulated in Assumption 4.4.1 or Eq. (4.15). Note that this frequency blur equals the ground frequency defined in Eq. (4.17).

For n-dimensional images on a rectangular domain results are similar. As a rule of thumb one may replace ω and L by the respective components ω_i and L_i in each perpendicular direction $i = 1, \ldots, n$.

Similar arguments can be given for zero-padding in the Fourier domain. In the spatial domain this will be reflected in a spatial convolution of the original signal with a convolution filter with a sinc-profile. In this case spatial blur will occur, inversely proportional to the frequency interval width. Details are given in Section 4.4.3.1.

4.4.1.2 The Effect of Periodic Replication in the Spatial Domain

Periodic replication is another way of extending the domain of definition of a finite support image. Again we restrict our attention to one-dimensional signals for the sake of argument. We are interested in the effect of periodic replication, particularly in comparison with that of zero-padding as discussed in the previous paragraph.

Our point of departure is Eq. (4.16). Fourier transformation yields

$$\mathscr{F}_{(a,b)}(u_p)(\omega) = b \int_{-\infty}^{\infty} u_p(x)\, e^{-ia\omega x}\, dx\,.$$

Since $u_p(x) = u_p(x + kL)$ for all $k \in \mathbb{Z}$ and all $x \in \mathbb{R}$, we may rewrite this as

$$\begin{aligned}
\mathscr{F}_{(a,b)}(u_p)(\omega) &= \sum_{k\in\mathbb{Z}} b \int_{-L/2}^{L/2} u_p(x)\, e^{-ia\omega(x-kL)}\, dx \\
&= b \int_{-L/2}^{L/2} u_p(x)\, e^{-ia\omega x}\, dx \sum_{k\in\mathbb{Z}} e^{ia\omega kL} \\
&= b \int_{-\infty}^{\infty} u_0(x)\, e^{-ia\omega x}\, dx \sum_{k\in\mathbb{Z}} e^{ia\omega kL} \\
&= \mathscr{F}_{(a,b)}(u_0)(\omega) \sum_{k\in\mathbb{Z}} e^{ia\omega kL}\,. \qquad (4.21)
\end{aligned}$$

In the last steps we encounter the zero-padded function, Eq. (4.15), again. The final step is to evaluate the infinite sum on the right. We present the essential fact as a lemma, and only give a sketch of the proof.

Lemma 4.4.1 (Poisson Summation Formula)**.**

$$\sum_{k\in\mathbb{Z}} e^{ikx} = 2\pi \sum_{k\in\mathbb{Z}} \delta(x - 2\pi k)\,.$$

Proof. Recall Theorem 2.2.3 on Page 26. Here we actually have $N = \infty$. The inner product space of interest is $L^2([-\pi, \pi])$, i.e. the class of square-integrable functions on a finite interval, endowed with the usual inner product, Eq. (2.10), with the integration interval $[0, 1]$ replaced by $[-\pi, \pi]$. The complex exponential functions constitute a countable orthonormal basis in the following sense. Set

$$e_k(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}}\, e^{ikx} \quad k \in \mathbb{Z}\,,$$

then for $k, l \in \mathbb{Z}$ we have

$$\langle e_k | e_l \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(l-k)x}\, dx = \operatorname{sinc}\left((l-k)\pi\right) = 0 \quad \text{if } k \neq l, \text{ and} \quad \langle e_k | e_k \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} dx = 1\,.$$

Thus the set $\{e_k\}_{k\in\mathbb{Z}}$ is indeed orthonormal. We refer to the mathematical literature on Hilbert space theory for a proof that it actually constitutes a basis for $L^2([-\pi, \pi])$. Let us take this for granted here, and apply Theorem 2.2.3, Page 26, to an arbitrary element $u \in L^2([-\pi, \pi])$:

$$u = \sum_{k\in\mathbb{Z}} \langle e_k | u \rangle\, e_k\,.$$

In other words,

$$u(x) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} \int_{-\pi}^{\pi} u(t)\, e^{ik(x-t)}\, dt\,.$$

(If u is continuous this equation holds point-wise, otherwise it relates $L^2([-\pi,\pi])$-equivalent functions.) Let us now extend the domain of u to all of $\mathbb{R}$ in a suitable way. Then obviously we have for any $m \in \mathbb{Z}$

$$u(x+2m\pi) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} \int_{-\pi}^{\pi} u(t+2m\pi)\, e^{ik(x-t)}\, dt\,.$$

Summing this over all $m \in \mathbb{Z}$, assuming this sum exists, we obtain

$$\sum_{m\in\mathbb{Z}} u(x+2m\pi) = \frac{1}{2\pi} \sum_{m\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \int_{-\pi}^{\pi} u(t+2m\pi)\, e^{ik(x-t)}\, dt\,.$$

The right hand side can be rewritten, after a change of variables $s = t + 2m\pi$, and an interchange of summation order, as

$$\sum_{m\in\mathbb{Z}} u(x+2m\pi) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} \sum_{m\in\mathbb{Z}} \int_{-\pi+2m\pi}^{\pi+2m\pi} u(s)\, e^{ik(x-s+2m\pi)}\, ds = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} \int_{-\infty}^{\infty} u(s)\, e^{ik(x-s)}\, ds\,.$$

In the last identity we have used the fact that $e^{\ell 2\pi i} = 1$ for any integer $\ell \in \mathbb{Z}$, allowing us to remove the m-dependent term in the complex exponential. This is in fact Lemma 4.4.1 in disguise, for we can rewrite the left hand side alternatively as

$$\sum_{m\in\mathbb{Z}} u(x+2m\pi) = \sum_{m\in\mathbb{Z}} \int_{-\infty}^{\infty} u(s)\, \delta(x-s+2m\pi)\, ds\,.$$

Since this holds for all u, we may rewrite the identity of the right hand sides of the previous two equations in a more abstract form, by setting $\xi = x - s$, as

$$\sum_{k\in\mathbb{Z}} e^{ik\xi} = 2\pi \sum_{m\in\mathbb{Z}} \delta(\xi + 2m\pi)\,.$$

Relabelling $m \longrightarrow -k$ on the right hand side completes the proof of Lemma 4.4.1. □

A thorough proof requires some mathematical scrutiny beyond our scope, but with the above outline we get the gist of it. Lemma 4.4.1 can be viewed as a discrete counterpart of Eqs. (4.10–4.11), involving discrete frequencies only. Using this lemma together with Eq. (4.21) we arrive at the following conclusion.

Theorem 4.4.2. *Let $u_p : \mathbb{R} \longrightarrow \mathbb{K}$ be a periodic function with periodicity L, and $u_0 = u_p\, \chi_{[-L/2,L/2]}$, then*

$$\mathscr{F}_{(a,b)}(u_p)(\omega) = 2\pi\, \mathscr{F}_{(a,b)}(u_0)(\omega) \sum_{k\in\mathbb{Z}} \delta(a\omega L - 2\pi k)\,.$$

If desired you may rewrite the right hand side using

$$\delta(a\omega L - 2\pi k) = \frac{1}{|a|L}\, \delta\left(\omega - \frac{2\pi k}{|a|L}\right)\,, \tag{4.22}$$

which can be verified by substitution of variables $Ax = y$ in an integral of the type $\int_{\mathbb{R}} \delta(Ax - B)\, \phi(x)\, dx$.

Theorem 4.4.2 establishes the relation between the Fourier transforms of a zero-padded function and its periodic replication. It shows that *a periodic function has a discrete Fourier spectrum*. The relevant frequencies are multiples of a ground frequency ω_1, i.e. $\omega_k = k\omega_1$, $k \in \mathbb{Z}$, with

$$\omega_1 \stackrel{\text{def}}{=} \frac{2\pi}{|a|L} \tag{4.23}$$

Compare this ground frequency to the most significant frequency interval of the sinc-function, and its zero-crossings, recall Eqs. (4.17) and (4.20).

In retrospection, we observe that the Fourier transforms of a zero-padded function u_0 and a periodically replicated one, u_p, even though they basically convey the same information, differ qualitatively. Whereas u_0 has a Fourier transform with a continuous spectrum of frequencies, the Fourier transform of u_p admits only a discrete spectrum. Still, despite these apparent differences, the frequency representations of u_0 and u_p do have the same essential characteristics. In some sense the frequency continuum of u_0 is 'essentially discrete', since it is of intrinsically finite resolution, characterized by a frequency blur equal to the sampling frequency interval of the discrete spectrum of u_p. Moreover, if we evaluate the Fourier amplitudes of u_0 on the same discrete frequency grid as for u_p we observe that they agree up to an 'infinite constant' ('$\frac{1}{|a|L}\,\delta(0)$', cf. Theorem 4.4.2 and Eq. (4.22)).

4.4.2. Fourier Transformation on a Discrete Spatial Grid

In order to study the effect of discrete sampling without confounding it with that of a finite scope we consider a digital signal in the form of an infinite discrete sequence.

Let the samples be given by $u_d(k) \in \mathbb{K}$, with $k \in \mathbb{Z}$. In order to subject it to our previously established Fourier transform we have to reformat it into the form $u_c(x) \in \mathbb{K}$, with $x \in \mathbb{R}$. We could do this through interpolation, recall Section 1.3. However, interpolation requires an ad hoc model, and would greatly complicate the analysis. Instead, recall that the class $\mathscr{S}'(\mathbb{R})$ of tempered distributions accommodates a huge variety of 'generalized functions', including discrete ones in the form of (linear combinations of derivatives of) the Dirac distribution. This allows us to circumvent the arbitrariness of interpolation.

Definition 4.4.2. *Let $u_d : \mathbb{Z} \longrightarrow \mathbb{K}$ be a suitably defined discrete signal. Then its continuum representation is defined as the function $u_c : \mathbb{R} \longrightarrow \mathbb{K}$ as follows:*

$$u_c(x) \stackrel{\text{def}}{=} \Delta x \sum_{k\in\mathbb{Z}} u_d(k)\,\delta(x - k\Delta x)\,,$$

in which Δx is the sampling width.

(Recall the similar connection between discrete and continuum representations encountered in Theorem 4.4.2.)

The attribute 'suitably defined' pertains to a (very mild) constraint on the sample values, viz. such that the above continuum representation is indeed a tempered distribution. Recall that it is customary to set $\Delta x = 1$ throughout, but here we prefer to make the sampling width explicit for the sake of argument.

With Definition 4.4.2 we model a discrete signal as a signal defined on a continuum, with 'spikes' of various relative amplitudes placed at the actual grid points. The relative amplitudes are just the actual discrete sample values. Note that the absolute amplitudes are all infinite. This is not a problem, since an element of $\mathscr{S}'(\mathbb{R})$ is not intended to be read out point-wise, but is only defined by virtue of the fact that it produces a well-defined output when exposed to a linear filter taken from the class of smooth, rapid decay functions, $\mathscr{S}(\mathbb{R})$.

Analyzing Definition 4.4.2 we conclude that, by virtue of the conventional factor Δx, the physical dimensions of the signals u_c and u_d are the same, so discrete and continuum samples represent the same physical 'stuff'. (In the n-dimensional generalization we will need a factor $\Delta x_1 \ldots \Delta x_n$.) A second reason for the factor Δx is that it allows us to identify

$$u_d(k) \stackrel{\text{def}}{=} u_c(x_k) \quad \text{with } x_k = k\Delta x, \tag{4.24}$$

and assign a function value $u_c(x)$ to any $x \in \mathbb{R}$ by the following hypothetical grid refinement procedure:

$$u_c(x) \stackrel{\text{def}}{=} \underbrace{\lim_{\Delta x \longrightarrow 0} \lim_{k \longrightarrow \infty}}_{x = k\Delta x \text{ fixed}} u_c(k\Delta x)\,. \tag{4.25}$$

As indicated the joint limiting procedure is constrained by keeping $x \equiv k\Delta x$ fixed. To see that this is a consistent definition, take any test function $\phi \in \mathscr{S}(\mathbb{R})$, and consider

$$\int_{-\infty}^{\infty} u_c(x)\,\phi(x)\,dx = \Delta x \sum_{k\in\mathbb{Z}} u_d(k)\,\phi(k\Delta x) = \sum_{k\in\mathbb{Z}} u_c(k\Delta x)\,\phi(k\Delta x)\,\Delta x\,. \tag{4.26}$$

Note that, with the identification of Eq. (4.24), this identity is *exact* only for the *particular* signal of Definition 4.4.2. Eq. (4.24) guarantees that the limiting case, Eq. (4.25), is consistent with our definition of a continuous integral via Riemann sums, Definition 3.1.1, for *any* signal $u_c \in \mathscr{S}'(\mathbb{R})$.

We now arrive at the main result of this section.

Theorem 4.4.3. *Let $u_d : \mathbb{Z} \longrightarrow \mathbb{K}$ and $u_c : \mathbb{R} \longrightarrow \mathbb{K}$ be given according to Definition 4.4.2. Then*

$$\mathscr{F}_{(a,b)}(u_c)(\omega) = b\,\Delta x \sum_{k\in\mathbb{Z}} u_d(k)\,e^{-ia\omega k\Delta x}\,.$$

Proof. The Fourier transform of $\delta(x - k\Delta x)$ is $b\int_{-\infty}^{\infty} e^{-ia\omega x}\,\delta(x - k\Delta x)\,dx = b\,e^{-ia\omega k\Delta x}$. The rest follows, assuming it is allowed to interchange the order of summation and Fourier transformation. □

Theorem 4.4.3 shows that *the Fourier transform of a discrete signal is a periodic function*. The period equals

$$\Omega_1 \stackrel{\text{def}}{=} \frac{2\pi}{|a|\Delta x}\,. \tag{4.27}$$

Notice the similarity in form between this period Ω_1 and the ground frequency ω_1 in Eq. (4.23). The grid constant Δx now formally takes over the role of the signal's scope L. This is not a coincidence, but a result of the analogy that exists between the Fourier transform and its inverse. Recall from Theorem 4.4.2 that a periodic function has a discrete Fourier spectrum. Due to the analogy the reverse also holds: A discrete function has a periodic Fourier transform. This implies that all frequency information is contained in the frequency interval $[-\Omega_1/2, \Omega_1/2]$. In analogy to our treatment of spatial replication versus truncation by spatial zero-padding we may want to compare the result of Theorem 4.4.3, which pertains to a periodically replicated Fourier transform, to that in which we take out only the relevant portion of the frequency domain by zero-padding. By analogy we should expect to find a sinc-convolved spatial signal defined for all $x \in \mathbb{R}$ upon Fourier inversion, rather than a discrete signal. Let us work out the details.

4.4.3. Inverse Fourier Transformation on a Finite Frequency Domain

In view of the analogies with the theory presented in Section 4.4.1 we will only briefly discuss the effect of limited scope in Fourier space. Again we do this by considering the two possible options for linking this case to the theory established for an unbounded domain, viz. zero-padding and periodic replication. To this end we define the following functions in the Fourier domain:

$$\widetilde{u}_0(\omega) \stackrel{\text{def}}{=} \begin{cases} \widetilde{u}(\omega) & \text{if } \omega \in [-\Omega/2, \Omega/2], \\ 0 & \text{otherwise,} \end{cases} \tag{4.28}$$

and

$$\widetilde{u}_p(\omega) \stackrel{\text{def}}{=} \begin{cases} \widetilde{u}(\omega) & \text{if } \omega \in [-\Omega/2, \Omega/2], \\ \widetilde{u}(\omega - k\Omega) & \text{if } \omega \notin [-\Omega/2, \Omega/2], \text{ and } k = \lfloor(\omega + \Omega/2)/\Omega\rfloor. \end{cases} \tag{4.29}$$

The tilde on top of the functions serves to remind you of the fact that we are working in the Fourier domain. The subscripts are self-explanatory.

4.4.3.1 The Effect of Zero-Padding in the Frequency Domain

Recall Theorems 4.3.1 and 4.3.2 and Results 4.3.3 and 4.3.4

Result 4.4.1. *The inverse Fourier transform of a compact support function $\widetilde{u}_0 = \widetilde{u}\,\chi_{[-\Omega/2,\Omega/2]}$, in which $\widetilde{u}$ is an arbitrarily extended function in the frequency domain, is given by*

$$\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_0)(x) = b\,\left(\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}) * \mathscr{F}^{-1}_{(a,b)}(\chi_{[-\Omega/2,\Omega/2]})\right)(x)\,,$$

with

$$\mathscr{F}^{-1}_{(a,b)}(\chi_{[-\Omega/2,\Omega/2]})(x) = \frac{|a|}{2\pi b}\,\Omega\,\mathrm{sinc}\left(\frac{a\Omega x}{2}\right)\,.$$

Proof. One can prove Result 4.4.1 by brute force, as was done for Theorem 4.4.1. This goes as follows.

$$\mathscr{F}^{-1}_{(a,b)}(\chi_{[-\Omega/2,\Omega/2]})(x) = \frac{|a|}{2\pi b}\int_{-\infty}^{\infty} e^{ia\omega x}\,\chi_{[-\Omega/2,\Omega/2]}(\omega)\,d\omega = \frac{|a|}{2\pi b}\int_{-\Omega/2}^{\Omega/2} e^{ia\omega x}\,d\omega = \frac{|a|}{2\pi b}\,\Omega\,\mathrm{sinc}\left(\frac{a\Omega x}{2}\right)\,.$$

However, slick use of Observation 4.3.1 allows us to base the proof on the already established result of Theorem 4.4.1:

$$\mathscr{F}^{-1}_{(a,b)}(\chi_{[-\Omega/2,\Omega/2]})(x) \overset{\star}{=} \mathscr{F}_{(a',b')}(\chi_{[-\Omega/2,\Omega/2]})(x) \overset{\circ}{=} b'\,\Omega\,\mathrm{sinc}\left(\frac{a'\Omega x}{2}\right) \overset{\star}{=} \frac{|a|}{2\pi b}\,\Omega\,\mathrm{sinc}\left(\frac{a\Omega x}{2}\right)\,.$$

The identities marked with $\star$ make use of Observation 4.3.1, Page 66, the one marked with $\circ$ exploits Theorem 4.4.1. Note that this formal way of manipulating established results frees us from having to compute the explicit (inverse) Fourier integral, which is a potentially hard problem. □

Looking at the spatial zero-crossings of the sinc -function above, the analogy of Eq. (4.17) on Page 75 becomes

$$x_k \overset{\mathrm{def}}{=} k\,x_1 \quad \text{for } k \in \mathbb{Z}\backslash\{0\}, \text{ with} \quad x_1 \overset{\mathrm{def}}{=} \frac{2\pi}{|a|\Omega}\,. \tag{4.30}$$

Note the exclusion of $k = 0$.

Result 4.4.1 answers the question we set out to solve in this section. Note that all results involve inverse Fourier transforms, and that all proofs have been based on manipulating already established results involving forward Fourier transforms, exploiting the symmetry captured by Observation 4.3.1 on Page 66.

A spatial function u that has a Fourier transform $\widetilde{u} = \mathscr{F}_{(a,b)}(u)$ with nontrivial values confined to a finite interval is called *band-limited*. Theorem 4.4.1 tells us that *a band-limited function equals the inverse Fourier transform of an arbitrarily extended frequency function 'blurred' with a* sinc*-function, the width of which is inversely proportional to the bandwidth, i.e. the size of the frequency support interval*. Thus the physical constraint of band-limitedness manifests itself as blurring in the spatial domain. More precisely, neighbouring spatial locations separated by a distance

$$|\Delta x| \leq \frac{2\pi}{|a|\Omega}\,, \tag{4.31}$$

in Fourier space, i.e. half the distance between the innermost zero-crossings of the sinc-function, cf. Theorem 4.4.1, are 'smeared out' in space and cannot be resolved. The right hand side of Eq. (4.31) is a measure of the *spatial blur* induced by the frequency cut-off stipulated in Eq. (4.28). Note the analogy with Eq. (4.20), Page 76.

For n-dimensional images on a rectangular domain results are similar. As a rule of thumb one may replace x and Ω by the respective components x_i and Ω_i in each perpendicular direction $i = 1, \ldots, n$.

Example 4.4.1. Define

$$u_b(x) = \mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_0)(x) \quad \text{and} \quad s_{a,b,\Omega}(x) = \frac{|a|}{2\pi b}\,\Omega\,\mathrm{sinc}\left(\frac{a\Omega x}{2}\right).$$

Using Result 4.4.1 we observe that $u_b = u_b * s_{a,b,\Omega}$, i.e., u_b is invariant under blurring with the given sinc-filter.

4.4.3.2 The Effect of Periodic Replication in the Frequency Domain

To find the spatial function corresponding to the periodic Fourier function Eq. (4.29), recall Theorem 4.4.2, Page 76.

Result 4.4.2. *Let $\widetilde{u}_p : \mathbb{R} \longrightarrow \mathbb{K}$ be a periodic function in the frequency domain with periodicity Ω, and $\widetilde{u}_0 = \widetilde{u}_p\,\chi_{[-\Omega/2,\Omega/2]}$, then*

$$\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_p)(x) = 2\pi\,\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_0)(x)\sum_{k\in\mathbb{Z}}\delta(ax\,\Omega - 2\pi k)\,.$$

Proof. Combine Theorem 4.4.2 with Fourier parameters (a', b') and Observation 4.3.1, Page 66. □

Result 4.4.2 establishes the relation between the spatial representations of a periodically replicated and a zero-padded frequency function. It shows that *a periodic frequency function has a discrete spatial representation*. The spatial grid points are multiples of a sampling width x_1, i.e.

$$x_k = k\,x_1 \quad \text{with} \quad x_1 \stackrel{\text{def}}{=} \frac{2\pi}{|a|\Omega} \quad (k \in \mathbb{Z})\,. \tag{4.32}$$

Compare this sampling width to the zero-crossings of the sinc-function, Eq. (4.30), and to the induced spatial blur given by the width of its central mode, Eq. (4.31).

Again it seems as if the spatial representations corresponding to a compact Fourier function and a periodic one are quite different qualitatively, suggesting that it is of crucial importance how one imagines the Fourier function to be continued beyond its finite borders. But again this is a misconception, for although the spatial representation of the former is defined on a continuous domain, it is characterized by an amount of intrinsic spatial blur roughly equal to the sampling width of the discrete spatial representation of the latter. Both spatial representations have equal spatial resolution, and their values on the aforementioned discrete grid are equal up to an infinite constant. Therefore they convey the same number of essential spatial degrees of freedom with equal information content. The spatial continuum of $\mathscr{F}^{-1}(\widetilde{u}_0)(x)$ is 'essentially discrete', so to speak, and thus not so much different from the discrete grid of $\mathscr{F}^{-1}(\widetilde{u}_p)(x)$ after all.

4.4.3.3 Sampling Theorem and Nyquist Frequency

We return to a remark made in Section 4.4.3.1 concerning the spatial representation of a band-limited function. If in Result 4.4.1 on Page 81 we insert the particular choice $\widetilde{u} = \widetilde{u}_p$, in which $\widetilde{u}_p$ is a periodic function in the frequency domain with period Ω, we obtain the following result.

Result 4.4.3 (Sampling Theorem)**.** *Let $\widetilde{u}_0 = \widetilde{u}\,\chi_{[-\Omega/2,\Omega/2]}$ be the Fourier transform of a band-limited function $u_B : \mathbb{R} \longrightarrow \mathbb{K}$, i.e. $u_B(x) = \mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_0)(x)$, then*

$$u_B(x) = \sum_{k\in\mathbb{Z}} u_B\left(\frac{2\pi k}{|a|\Omega}\right)\,\mathrm{sinc}\left(\frac{|a|\Omega x}{2} - k\pi\right).$$

This result states that, although it is defined for all $x \in \mathbb{R}$, *a band-limited function is completely determined by its values on a discrete grid*. The grid points are precisely those given by Eq. (4.30). The cut-off frequency (i.e. the size of the frequency window divided by two) is also known as the *Nyquist frequency*.

Proof. Take $\widetilde{u} = \widetilde{u}_p$ in Result 4.4.1, in which $\widetilde{u}_p$ is a periodic function in the frequency domain with periodicity Ω:

$$\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_0)(x) = b\left(\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_p) * \mathscr{F}^{-1}_{(a,b)}(\chi_{[-\Omega/2,\Omega/2]})\right)(x)\,.$$

Recall that combining Theorem 4.4.2 and Observation 4.3.1, Page 66, led to the conclusion (Result 4.4.2) that

$$\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_p)(x) = 2\pi\,\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_0)(x)\sum_{k\in\mathbb{Z}}\delta(ax\Omega - 2\pi k)\,.$$

Furthermore we have

$$\mathscr{F}^{-1}_{(a,b)}(\chi_{[-\Omega/2,\Omega/2]})(x) = \frac{|a|}{2\pi b}\,\Omega\,\mathrm{sinc}\left(\frac{a\Omega x}{2}\right)\,.$$

Using these expressions and working out the convolution product (interchanging the order of summation and integration), and setting $u_B = \mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_0)$, establishes Result 4.4.3. □

4.4.4. Inverse Fourier Transformation on a Discrete Frequency Grid

Finally we arrive at the analogue of Section 4.4.2 for a discrete signal in Fourier space. To this end we can duplicate Definition 4.4.2, Page 79, and simply substitute $(x, \Delta x) \longleftrightarrow (\omega, \Delta\omega)$, and $(u_d, u_c) \longleftrightarrow (\widetilde{u}_d, \widetilde{u}_c)$.

Definition 4.4.3. *Let $\widetilde{u}_d : \mathbb{Z} \longrightarrow \mathbb{K}$ be a suitably defined discrete signal in frequency space. Then its continuum representation is defined as the function $\widetilde{u}_c : \mathbb{R} \longrightarrow \mathbb{K}$ as follows:*

$$\widetilde{u}_c(\omega) \stackrel{\text{def}}{=} \Delta\omega \sum_{k\in\mathbb{Z}} \widetilde{u}_d(k)\,\delta(\omega - k\Delta\omega)\,,$$

in which $\Delta\omega$ is the frequency sampling width.

Likewise, the counterpart of Theorem 4.4.3 on Page 80 is found by the procedure of using Observation 4.3.1 on Page 66.

Result 4.4.4. *Let $\widetilde{u}_d : \mathbb{Z} \longrightarrow \mathbb{K}$ and $\widetilde{u}_c : \mathbb{R} \longrightarrow \mathbb{K}$ be given according to Definition 4.4.3. Then*

$$\mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_c)(x) = \frac{|a|}{2\pi b}\,\Delta\omega \sum_{k\in\mathbb{Z}} \widetilde{u}_d(k)\,e^{iaxk\Delta\omega}\,.$$

You should be able to provide the details of the proof as suggested.

Result 4.4.4 confirms what we should already have expected, viz. that the *inverse Fourier transform of a discrete signal in frequency space is a periodic function*. The spatial period is given by

$$X_1 \stackrel{\text{def}}{=} \frac{2\pi}{|a|\Delta\omega}\,.$$

Cf. Theorem 4.4.3, Page 80, and the comments following its proof, in particular Eq. (4.27). Again we observe a spatiofrequency trade-off: The finer the frequency sampling, the longer the spatial period. The product of frequency sampling width and spatial period is a fixed, dimensionless constant. Note also the analogy with Eq. (4.32) on Page 82.

4.5. The Discrete Fourier Transform

Let us summarize what we have achieved sofar. In Section 4.3 we have introduced the Fourier transform and its inverse for functions defined on an unbounded, continuous domain. In the various subsections of Section 4.4 we have worked out the implications of truncating or discretizing the function domain, but we have not yet combined these two operations. This is what this section intends to cover. We consider the operationally well-defined Fourier transform and its inverse on discrete functions with finite support, i.e. the one actually used in computer algorithms. It is therefore important to understand the exact relationship between Section 4.3 and the following, so-called *Discrete Fourier Transform* (*DFT*), and its inverse. The intermezzo in Sections 4.4.1–4.4.4 was intended to provide some basic insight.

In the finite-dimensional case there is no need to distinguish between images and filters. Mathematically speaking, a finite-dimensional vector space (of filters, say) is isomorphic to its own dual (the class of images). For this reason we will collectively denote the digital image and filter functions by $u : \{1, \dots, N_1\} \times \dots \times \{1, \dots, N_n\} \longrightarrow \mathbb{K}$. As before it is instructive to introduce the DFT using a general Fourier convention. The DFT of a discrete function can be seen as a special instance of the Fourier transform on $\mathscr{S}'(\mathbb{R}^n)$—which after all contains both compactly supported as well as discrete functions—as defined in Section 4.3. We start with the one-dimensional case.

Notation. For the sake of clarity we use capital symbols for discrete, and small symbols for analogue functions.

Definition 4.5.1. *The one-dimensional Discrete Fourier Transform is defined as follows:*

$$\widetilde{U}(k) \stackrel{\text{def}}{=} \beta \sum_{\ell=1}^{N} U(\ell)\, e^{-2\pi i \alpha (k-1)(\ell-1)/N}\,,$$

in which $(\alpha, \beta) \in \{-1, 1\} \times \mathbb{R}^+$ *are constant parameters.*

Again, DFT is a continuous, linear, invertible mapping.

Proof. The thing to prove is that Definition 4.5.1 is a good definition, in the sense that it is indeed compatible with our definition of Fourier transformation on the subclass of discrete functions in[5] $\mathscr{S}'(\mathbb{R}^n)$. To this end, let us embed a discrete signal $U(\ell)$, $\ell = 1, \dots, N$, into the class of tempered distributions by defining $u(x)$ for (almost all) $x \in \mathbb{R}$ such that $u \in \mathscr{S}'(\mathbb{R})$ as follows:

$$u(x) \stackrel{\text{def}}{=} \Delta x \sum_{\ell=1}^{N} U(\ell)\, \delta(x - x_\ell) \quad \text{with} \quad x_\ell \stackrel{\text{def}}{=} \frac{\ell - 1}{N} L \quad \text{and thus} \quad \Delta x = \frac{L}{N}\,.$$

Recall Definition 4.4.2, Page 79, and the subsequent discussion of a similar procedure. In the present situation we have accounted for both discretization and truncation: The parameter L corresponds to the width of the physical interval[6] $[x_1 = 0, x_{N+1} = L]$ on which the signal is defined, and N is the number of discrete sample points. Recall Definition 4.3.2 on Page 66:

$$\widetilde{u}(\omega) \stackrel{\text{def}}{=} b \int_{\mathbb{R}^n} e^{-ia\omega\cdot x}\, u(x)\, dx \quad \text{with inverse} \quad u(x) = \frac{|a|^n}{(2\pi)^n b} \int_{\mathbb{R}^n} e^{ia\omega\cdot x}\, \widetilde{u}(\omega)\, d\omega\,,$$

in which $(a, b) = (1, 1)$ expresses our default. Thus for the one-dimensional discrete signal at hand we get

$$\mathscr{F}_{(a,b)}(u)(\omega) = b\,\Delta x \int_{-\infty}^{\infty} \sum_{\ell=1}^{N} U(\ell)\, \delta(x - x_\ell)\, e^{-ia\omega x}\, dx = b\,\Delta x \sum_{\ell=1}^{N} U(\ell)\, e^{-ia\omega x_\ell}\,.$$

[5] Since $\mathscr{S}'(\mathbb{R}^n) \supset \mathscr{S}(\mathbb{R}^n)$ it is most convenient to take this as a point of departure.

[6] Note that x_{N+1} is not a sample point for the discrete signal; there is exactly one sample point per sampling interval.

This establishes the Fourier transform if the signal is interpreted as a tempered distribution. In other words, it is defined for all frequencies $\omega \in \mathbb{R}$, as opposed to the DFT. By sampling the frequency domain according to

$$\omega_k \stackrel{\text{def}}{=} \frac{k-1}{L}\,,$$

for $k = 1, \ldots, N$, we obtain the exact DFT:

$$\mathscr{F}_{(a,b)}(u)(\omega_k) = \frac{bL}{N} \sum_{\ell=1}^{N} U(\ell) e^{-ia(k-1)(\ell-1)/N}\,.$$

Adopting the Fourier convention $(a, b) = (2\pi\alpha, \beta/\Delta x) = (2\pi\alpha, \beta N/L)$ reproduces the Fourier convention of Definition 4.5.1. □

Sampling the Fourier domain as we did in the proof might seem ad hoc. But if you recall the discussion on (infinitely long) discrete signals, this entails *no approximation*. Indeed, the following theorem gives us the *exact* inversion formula corresponding to Definition 4.5.1 based solely on the discretized frequency spectrum.

Theorem 4.5.1. *With the DFT convention of Definition 4.5.1 the inverse DFT is given by*

$$U(\ell) = \frac{1}{\beta N} \sum_{k=1}^{N} \widetilde{U}(k)\, e^{2\pi i\alpha(k-1)(\ell-1)/N}\,.$$

Before we turn to the proof, let us try to understand the rationale behind the particular frequency sampling $\omega \longrightarrow \omega_k = (k-1)/L$ needed to relate DFT to its continuous counterpart. In fact we have discussed all issues involved so as to be able to understand the situation at least from a heuristic point of view. Recall that truncation leads to *frequency blur* by an amount $\Delta\omega \approx 2\pi/(|a|L)$, Eq. (4.20), Page 76. Thus it makes little sense to consider frequencies separated by a distance smaller than $\Delta\omega$. On top of that, discretization in steps of Δx leads to a periodic Fourier transform with period $\Omega_1 = 2\pi/(|a|\Delta x)$, so that all relevant frequencies are contained in the interval $[-\Omega_1/2, \Omega_1/2]$, Eq. (4.27), Page 80. Thus it seems plausible that it suffices to consider a sampling of this finite frequency interval in steps of $\Delta\omega$. The number of reasonable sample points consequently equals $\Omega_1/\Delta\omega = L/\Delta x$. With $\Delta x = L/N$ this indeed yields a total number of N, as it should. The proof of Theorem 4.5.1, as we will see, confirms the correctness of this heuristic argument and shows that there is no loss of information if we use this frequency sampling strategy.

Let us turn to the rigorous proof, after the following lemma, which you should be able to prove yourself.

Lemma 4.5.1. *For $\xi \in \mathbb{C}$, $\xi \neq 1$, we have*

$$\sum_{k=1}^{N} \xi^{k-1} = \frac{1-\xi^N}{1-\xi}\,.$$

Proof. Set $\xi = e^{2\pi i\alpha(\ell-m)/N}$, and notice that $-N < \alpha(\ell-m) < N$ for all $\ell, m = 1, \ldots, N$ and $\alpha = \pm 1$. It follows that $\xi \neq 1$ if $\ell \neq m$, $\xi = 1$ if and only if $\ell = m$, and $\xi^N = 1$ for all $\ell, m = 1, \ldots, N$, whence, by Lemma 4.5.1,

$$\sum_{k=1}^{N} \xi^{k-1} = \begin{cases} 0 & \text{if } \ell \neq m, \\ N & \text{if } \ell = m. \end{cases}$$

Therefore,

$$\frac{1}{\beta N} \sum_{k=1}^{N} \widetilde{U}(k)\, e^{2\pi i\alpha(k-1)(\ell-1)/N} = \frac{1}{N} \sum_{k=1}^{N} \sum_{m=1}^{N} U(m)\, e^{2\pi i\alpha(k-1)(\ell-m)/N} = \frac{1}{N} \sum_{m=1}^{N} U(m) \sum_{k=1}^{N} \xi^{k-1} = U(\ell)\,.$$

The last step uses the previous observation. This proves Theorem 4.5.1. □

α	β	**examples of references**
1	1	our default, Horn[a] [16]
-1	1	
1	$1/N$	Gonzalez and Woods [12]
-1	$1/N$	
1	$1/\sqrt{N}$	Petrou [23]
-1	$1/\sqrt{N}$	Mathematica [32]

Table 4.2: Summary of Fourier conventions frequently encounterd in the literature for the one-dimensional DFT, and some references. Cf. Table 4.1.

[a]Horn uses an implicit periodicity doubling, which makes it look as if $\alpha = 1/2$ in his formulas at first glance.

The general DFT convention in Definition 4.5.1 and Theorem 4.5.1 differs somewhat from the one adopted previously for the continuum counterpart in Definitions 4.3.2–4.3.3, Page 66. We will adhere to the default $(\alpha, \beta) = (1, 1)$. Gonzalez and Woods [12] use a convention with $(\alpha, \beta) = (1, 1/N)$. Horn [16], on the other hand, agrees with ours. Maria Petrou has opted for $(\alpha, \beta) = (1, 1/\sqrt{N})$, which leads to a more symmetrical relation between forward and inverse DFT. Mathematica, of course, has a built-in option for specifying the Fourier convention, with default $(\alpha, \beta) = (-1, 1/\sqrt{N})$. Cf. Table 4.2 for a summary and some references. Note that in all cases the product of factors in front of the forward and inverse Fourier sums equals $1/N$. This gives us a consistency check that can be applied to all (one-dimensional) cases.

A final remark concerns the range of indices. Recall from Chapter 1 that we start our discrete labelling with index value 1, as opposed to many text books, but in agreement with Mathematica [32]. This explains the 'minus one' in the exponents of the Fourier transform and its inverse.

Example 4.5.1. Besides Definition 4.5.1 there exists an equally popular DFT definition

$$\widetilde{U}(k) \stackrel{\text{def}}{=} \beta \sum_{\ell=0}^{N-1} U(\ell)\, e^{-2\pi i \alpha k \ell / N}\,.$$

With this definition, the inverse DFT is given by

$$U(\ell) = \frac{1}{\beta N} \sum_{k=0}^{N-1} \widetilde{U}(k)\, e^{2\pi i \alpha k \ell / N}\,.$$

There exists an optimized implementation of DFT, known as the *Fast Fourier Transform*, or *FFT*. We refer to the literature for details [2, 12, 22].

The n-dimensional DFT and its inverse follow from the 1-dimensional case.

Definition 4.5.2. *The n-dimensional Discrete Fourier Transform is defined as follows:*

$$\widetilde{U}(k_1, \ldots, k_n) \stackrel{\text{def}}{=} \beta \sum_{\ell_1=1}^{N_1} \cdots \sum_{\ell_n=1}^{N_n} U(\ell_1, \ldots, \ell_n)\, e^{-2\pi i \alpha [(k_1-1)(\ell_1-1)/N_1 + \ldots + (k_n-1)(\ell_n-1)/N_n]}\,,$$

in which $(\alpha, \beta) \in \{-1, 1\} \times \mathbb{R}^+$ are constant parameters.

Theorem 4.5.2. *With the DFT convention of Definition 4.5.2 the inverse DFT is given by*

$$U(\ell_1, \ldots, \ell_n) = \frac{1}{\beta N_1 \ldots N_n} \sum_{k_1=1}^{N_1} \cdots \sum_{k_n=1}^{N_n} \widetilde{U}(k_1, \ldots, k_n)\, e^{2\pi i \alpha [(k_1-1)(\ell_1-1)/N_1 + \ldots + (k_n-1)(\ell_n-1)/N_n]}\,.$$

Proof. The proof proceeds in the same fashion as for Theorem 4.5.1. Note that the product of factors in front of the DFT and its inverse formula is always $1/(N_1 \dots N_n)$. The same Fourier conventions apply as for the one-dimensional case, with the same default in our case. □

Example 4.5.2. The relationship between our Fourier parameters (α, β) and Mathematica's (version 5.2) optional specification `FourierParameters->{a,b}` for the one-dimensional case, with N discrete samples, is given by

$$\begin{cases} \alpha &= -b \\ \beta &= \sqrt{N^{a-1}}\,. \end{cases}$$

In particular, our default corresponds to `FourierParameters->{1,-1}`.

4.5.1. DFT Subjected to the Continuum Limit

In the previous section we have derived the DFT from the continuum formulation of Fourier transformation in Sections 4.3.2 and 4.3.3 by means of discretization and truncation. Obviously we could have opted for the reverse approach, starting out to define the DFT, and subsequently investigating its continuum limit. Here we sketch the procedure for the one-dimensional case for the sake of simplicity, recall Definition 4.5.1.

Let us make the following identifications:

$$x_\ell \overset{\text{def}}{=} \overline{x}_\ell + x_c \quad \text{with} \quad x_\ell \overset{\text{def}}{=} \frac{\ell - 1}{N} L \quad \text{and} \quad x_c \overset{\text{def}}{=} \frac{L}{2} \quad \text{and} \quad \omega_k \overset{\text{def}}{=} \frac{k-1}{L}\,,$$

for $k, \ell = 1, \dots, N+1$, and define $u(\overline{x}_\ell) = U(\ell)$. Thus the spatial and frequency grid constants are given by, respectively, $\Delta x = L/N$ and $\Delta\omega = 1/L$. With a modest amount of foresight we have shifted the mid-point of the physical interval, such that the physical boundaries are now symmetric with respect to the origin: $\overline{x}_1 = -L/2$, respectively $\overline{x}_{N+1} = L/2$.

Furthermore, let us switch Fourier conventions as before by setting

$$(\alpha, \beta) = \left(\frac{a}{2\pi}, b\Delta x\right)\,.$$

Now consider the limit in which $N \longrightarrow \infty$ and $\Delta x \longrightarrow 0$ simultaneously, such that the physical interval width $L = N\Delta x$ remains fixed. Using the definition of integrals via Riemann sums explained in Section 3.1, and defining $\widetilde{u}(\omega_k) = \widetilde{U}(k)\, e^{ia\omega_k x_c}$, we readily find that in this limit Definition 4.5.1 yields

$$\widetilde{U}(k) = \underbrace{\lim_{\Delta x \longrightarrow 0} \lim_{N \longrightarrow \infty}}_{L = N\Delta x \text{ fixed}} \beta \sum_{\ell=1}^{N} U(\ell)\, e^{-2\pi i a \omega_k x_\ell}\,,$$

or

$$\widetilde{U}(k) = b\, e^{-ia\omega_k x_c} \underbrace{\lim_{\Delta x \longrightarrow 0} \lim_{N \longrightarrow \infty}}_{L = N\Delta x \text{ fixed}} \sum_{\ell=1}^{N} u(\overline{x}_\ell)\, e^{-ia\omega_k \overline{x}_\ell}\, \Delta x = b\, e^{-ia\omega_k x_c} \int_{-L/2}^{L/2} u(x)\, e^{-ia\omega_k x}\, dx\,.$$

In other words,

$$\widetilde{u}(\omega_k) = b \int_{-L/2}^{L/2} u(x)\, e^{-ia\omega_k x}\, dx\,.$$

Note that the parameter β is proportional to Δx and therefore belongs inside the limit-operator. We recognize the familiar continuous Fourier transform of Definition 4.3.2 on Page 66, evaluated at the sample points ω_k. Also note that in this limit the index k now runs over all natural numbers.

Although we have established the connection at the discrete frequencies ω_k with fixed grid spacing $\Delta\omega = 1/L$ only, the result can be generalized to arbitrary frequencies via sinc-interpolation according to a modification of the sampling theorem, recall Result 4.4.3 on Page 82. To this end, modify Result 4.4.3 on Page 82 by making the following formal changes, cf. Observation 4.3.1, Page 66:

Result 4.4.3		modification
$\widetilde{u}_0$	$\longleftrightarrow$	u_0
$\widetilde{u}$	$\longleftrightarrow$	u
Ω	$\longleftrightarrow$	L
$u_B \equiv \mathscr{F}^{-1}_{(a,b)}(\widetilde{u}_0)$	$\longleftrightarrow$	$\widetilde{u}_C \equiv \mathscr{F}^{-1}_{(a',b')}(u_0)$
x	$\longleftrightarrow$	ω
(a,b)	$\longleftrightarrow$	(a',b')

For a compactly supported spatial signal u_0 it suffices to know its discrete frequency spectrum $\widetilde{u}_C(\omega_k) = b\int_{-L/2}^{L/2} u_0(x)\, e^{-ia\omega_k x}\, dx$. Via the sinc-interpolation formula we may obtain the full continuum spectrum.

4.6. Interpretation of the Fourier Transform

In order to understand the significance of Fourier transformation let us consider, as an example, the two-dimensional case. According to Theorem 4.5.2 we can interpret the inverse Fourier transformation formula as an expansion of a spatial image in terms of a linear combination of complex exponential basis functions. For $n = 2$ we may write

$$u(\ell_1,\ell_2) = \frac{1}{\beta N_1 N_2} \sum_{k_1=1}^{N_1} \sum_{k_2=1}^{N_2} \widetilde{U}(k_1,k_2)\, e_{k_1,k_2}(\ell_1,\ell_2)\,,$$

in which the (k_1,k_2)-labelled spatial basis function is given by

$$e_{k_1,k_2}(\ell_1,\ell_2) \stackrel{\text{def}}{=} e^{2\pi i\alpha[(k_1-1)(\ell_1-1)/N_1+(k_2-1)(\ell_2-1)/N_2]}\,.$$

Apparently the set $\{e_{k_1,k_2} \mid k_1 = 1,\ldots,N_1\,,\ k_2 = 1,\ldots,N_2\}$ of discrete images forms a complete basis that spans the space of all possible discrete $N_1 \times N_2$ images. Recall that this is just what Theorem 2.2.3 on Page 26 expressed. Figures 4.2–4.3 show all basis functions for the case $N_1 = N_2 = 8$. Notice that the cases $(k_1,k_2) \in \{(1,1),(1,5),(5,1),(5,5)\}$ yield a sine part with an argument that is a multiple of π. Only for the first one the cosine part yields the constant 1.

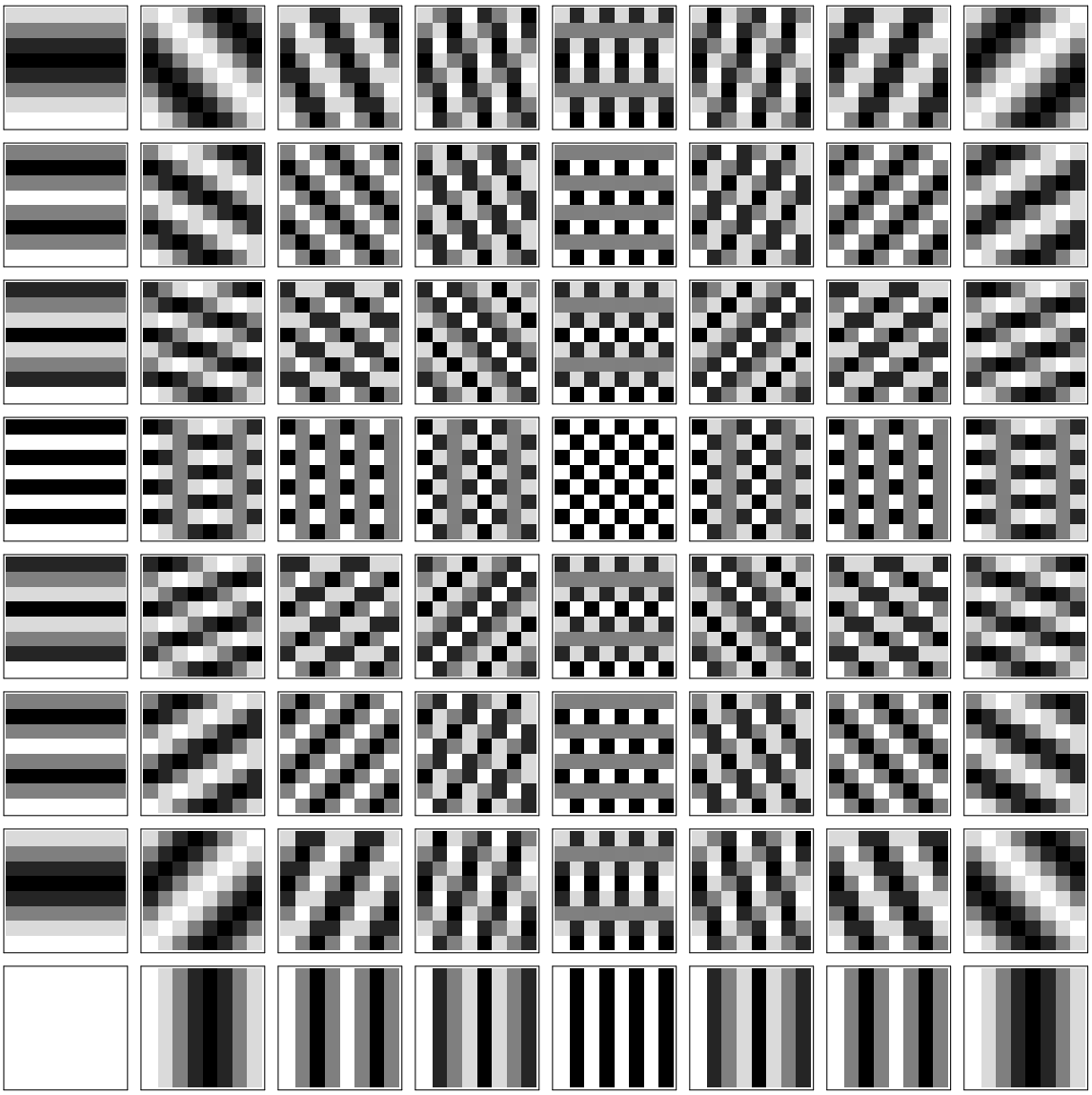


Figure 4.2: All 8×8 images can be decomposed into a unique linear combination of the 64 8×8 basis images e_{k_1,k_2}, $k_1, k_2 = 1, \ldots, 8$. This figure displays the real part, $\mathrm{Re}(e_{k_1,k_2})$. The lower left corner corresponds to $(k_1, k_2) = (1, 1)$, k_1 increases from left to right, k_2 from bottom to top. White means positive, black means negative, mid-grey means zero.

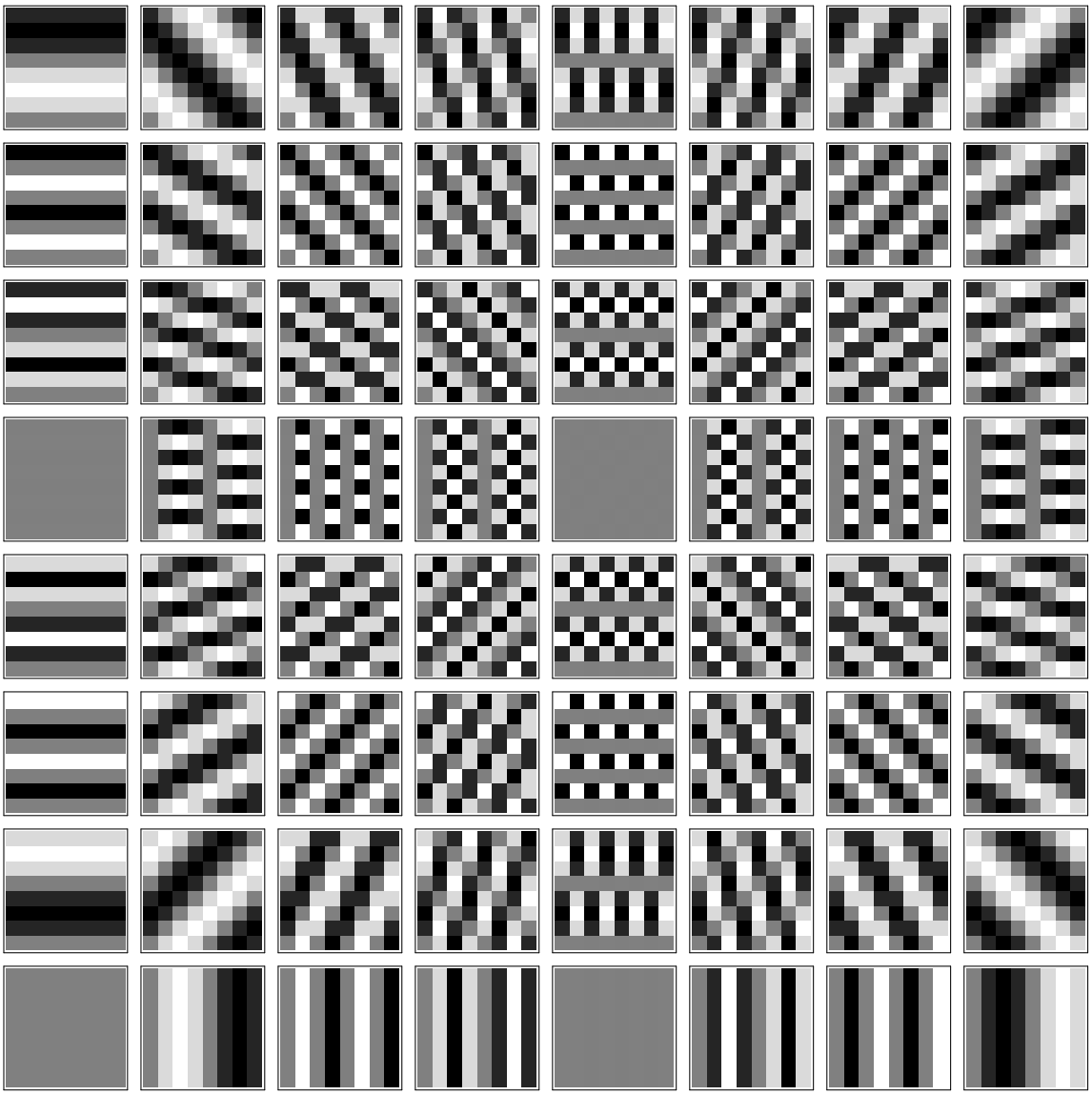


Figure 4.3: All 8×8 images can be decomposed into a unique linear combination of the 64 8×8 basis images e_{k_1,k_2}, $k_1, k_2 = 1,\dots,8$. This figure displays the imaginary part, $\mathrm{Im}(e_{k_1,k_2})$. The lower left corner corresponds to $(k_1,k_2) = (1,1)$, k_1 increases from left to right, k_2 from bottom to top. White means positive, black means negative, mid-grey means zero.

4.7. Applications of the Continuous Fourier Transform

The continuous Fourier transform is a helpful conceptual tool, especially in the context of linear differential operators and partial differential equations, and in the context of convolutions. Fourier transformation should be the first thing that comes to your mind in these situations.

As an example of the use of Fourier techniques in the context of linear partial differential equations, recall

Theorem 3.4.1 on Page 55, where it was conjectured that the family of normalized Gaussian filters,

$$\psi_t(x) = \frac{1}{\sqrt{4\pi\, t}^{\,n}} \exp\left[-\frac{\|x\|^2}{4t}\right] \quad \text{with } t > 0, \tag{4.33}$$

satisfies the *isotropic heat equation*

$$\frac{\partial u}{\partial t} - \Delta u = 0\,. \tag{4.34}$$

Given this conjecture the proof boils down to a verification by substitution. But would we have found this solution of the heat equation if we had no clue to begin with? In other words, how would one proceed if one were to solve the heat equation from scratch?

Example 4.7.1. Problem: Solve Eq. (4.34). Solution: Since the equation is linear, we should immediately think of Fourier transformation. Here is how it works.

Substitute

$$u(x,t) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\omega\cdot x}\, \widehat{u}(\omega,t)\, d\omega\,, \tag{4.35}$$

recall Conjecture 4.3.1 on Page 66, assuming that the solution exists in a 'suitably defined' function space. Note that we treat t as a constant. After interchanging the order of differentiation and integration, the result becomes

$$\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\omega\cdot x} \left[\frac{\partial \widehat{u}(\omega,t)}{\partial t} + \|\omega\|^2\, \widehat{u}(\omega,t)\right] d\omega = 0\,.$$

In the integrand, between the square brackets, we recognize the Fourier transform of the left hand side of Eq. (4.34). Since the Fourier transform of the null function is again the null function, we conclude that

$$\frac{\partial \widehat{u}(\omega,t)}{\partial t} + \|\omega\|^2\, \widehat{u}(\omega,t) = 0\,. \tag{4.36}$$

Note that we could have found this immediately by applying Result 4.3.1, Page 69, to the left hand side of Eq. (4.34), using the fact that Fourier transformation is linear, and by noticing that the Fourier transform of the null function on the right hand side is again the null function—a direct consequence of linearity.

Equation (4.36)—which is in fact an ordinary differential equation—is easily solved:

$$\widehat{u}(\omega,t) = C\, \exp\left(-t\, \|\omega\|^2\right)\,,$$

for some constant $C \in \mathbb{K}$, which depends on the initial condition. Suppose

$$u(x,0) = \delta(x)\,. \tag{4.37}$$

It is well-known from linear systems theory that the response of a linear system upon exposure to a 'δ-spike' produces the system's transfer function at the output side. Thus this is indeed the way to probe the heat equation to reveal its 'fundamental solution'. Again we recast this initial condition into Fourier language (now we need Conjecture 4.3.2 on Page 69, since we are dealing with a tempered distribution) by substituting Eq. (4.35), with $t = 0$, using the Fourier representation of the Dirac point distribution, Lemma 4.3.3, Page 70 (or Eq. (4.10), Page 70). After bringing everything to the left hand side we obtain

$$\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\omega\cdot x}\, \left[\widehat{u}(\omega,0) - 1\right]\, d\omega = 0\,.$$

By the same token as before this implies that $\widehat{u}(\omega,0) - 1$ is the null function, from which it follows that $C = 1$. Thus the solution to the full initial value problem, Eqs. (4.34–4.37), is given by

$$\widehat{u}(\omega,t) = \exp\left(-t\, \|\omega\|^2\right)\,, \tag{4.38}$$

in the Fourier domain. The final step is to invert the Fourier representation of this solution to the spatial domain (no free lunch here: Fourier inversion tends to be the hardest part).

Knowing $\widehat{u}(\omega, t)$ we apply Eq. (4.35) to obtain the spatial representation we are looking for:

$$u(x,t) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\omega\cdot x - t\|\omega\|^2}\, d\omega\,. \tag{4.39}$$

Indeed, here we run into some technical difficulties, since even though we started out from a real-valued linear system, we have to deal with a complex integral. It is at moments like this, when the real trouble begins, that details are usually left as an exercise. We shall adhere to this widespread character building malevolence, but not without some helpful hints on how to proceed.

You will not find it hard to solve the integral if you are familiar with the basics of complex analysis (in particular if you know what *poles* are, and the role they play in complex integrals). If, on the other hand, you are inclined to confuse poles with certain geographic regions with a harsh climate, then the following lemma should help you out in this specific case.

Lemma 4.7.1. *For all $x \in \mathbb{R}$ we have*

$$\int_{-\infty}^{\infty} e^{-(\omega+ix)^2}\, d\omega = \sqrt{\pi}\,.$$

The lemma states, *deus-ex-machina*, that the imaginary part in the argument of the exponential function has no effect whatsoever. It can be generalized to multiple dimensions, with $x = (x_1, \ldots, x_n)$ and $\omega = (\omega_1, \ldots, \omega_n)$:

$$\int_{\mathbb{R}^n} e^{-\|\omega+ix\|^2}\, d\omega = \sqrt{\pi}^{\,n}\,.$$

As a final hint, note that we may rewrite

$$i\omega\cdot x - t\,\|\omega\|^2 = -\left\|\sqrt{t}\,\omega - \frac{1}{2}i\frac{x}{\sqrt{t}}\right\|^2 - \frac{\|x\|^2}{4t}\,.$$

The details are left as an exercise to the reader.

Example 4.7.2. As a second demonstration of the power of Fourier analysis, consider the autoconvolution property of Theorem 3.4.1, Page 55. Recall also the equivalent formulation in an exercise on Page 55 in terms of the t-parametrization, which we shall exploit here:

$$\psi_t * \psi_u = \psi_{t+u}\,,$$

with ψ_t defined as in Eq. (4.33). Since $\widehat{\psi}_t(x) = \widehat{u}(x,t)$, given by Eq. (4.38), it is immediately obvious that

$$\widehat{\psi}_t\,\widehat{\psi}_u = \widehat{\psi}_{t+u}\,.$$

This looks similar to the previous formula, except that ordinary multiplication takes over the role of convolution. But according to Theorem 4.3.2, Page 72, this implies the autoconvolution property in the spatial domain.

We conclude this section with a few exercises.

Exercise. Solve, via the Fourier detour of Eq. (4.13), the following initial value problem:

$$\left\{\begin{array}{rcl} \dfrac{\partial u}{\partial t} - \Delta\, u & = & 0\,, \\ u(x,0) & = & f(x)\,, \end{array}\right.$$

in which $f \in \mathscr{S}'(\mathbb{R}^n)$ is a raw image. Do you recognize the result? Recall Definition 3.4.2, Page 56, and a similar exercise on Page 57.

Exercise. A scale space representation $u(x,t) = u_t(x)$ of a raw image $f(x)$ may be written symbolically as

$$u_t = e^{t\Delta}\, f\,.$$

Try to explain this notation. Hint: (i) Eq. (3.10), Page 57, or (ii) Lemma 4.3.1, Page 68 or Result 4.3.1, Page 69.

Exercise. Recall Theorem 3.4.1, Page 55. Let u_t be the scale space representation of a raw image f at some fixed scale $t > 0$, in other words $u_t = f * \psi_t$. Show that $u_t * \psi_s = u_{s+t}$ for any $s > 0$. Explain in words what this means.

Exercise. The previous two exercises suggest a method for *deblurring*, or enhancement of spatial resolution, viz. by considering $f(x) = e^{-t\Delta}\, u(x,t)$, or $f = u_t * \psi_{-t}$. Do you expect this procedure of deblurring to work in practice? Hint: Verify whether the above schemes are well-posed by assuming that $u(x,t) = u_t(x)$ is known up to some (arbitrarily) small, but finite additive noise function $\delta u(x) \approx 0$.

Exercise. Sketch in one figure the graphs of the one-dimensional Gaussian scale space filters

$$\psi_t(x) = \frac{1}{\sqrt{4\pi\, t}}\, \exp\left[-\frac{x^2}{4t}\right]$$

for $t = 1/2, 1, 2$, using different colours to identify each parameter setting. Do the same with their Fourier equivalents,

$$\widehat{\psi}_t(\omega) = \exp\left[-t\omega^2\right]\,,$$

using the same colour labelling. Several observations can be made. Describe these.

Exercise. Cf. the previous exercise. What happens to the respective representations in the asymptotic limit $t \longrightarrow 0$? Are the limiting spatial and Fourier functions still filters in the sense of $\mathscr{S}(\mathbb{R})$-functions? Are they still related by Fourier transformation?

Exercise. Solve, along similar lines via the Fourier detour, the following initial value problem. To this end, use the intuitive explanation of the linear differential operator $\sqrt{-\Delta}$ discussed in the example on Page 69:

$$\left\{\begin{array}{rcl} \dfrac{\partial u}{\partial t} + \sqrt{-\Delta}\, u & = & 0\,, \\ u(x,0) & = & \delta(x)\,. \end{array}\right.$$

Hint: The actual Fourier inversion has been carried out in an example on Page 70.

4.8. Applications of the Discrete Fourier Transform

The DFT comes into play whenever we want to actually implement a Fourier based method, recall the filtering diagram of Eq. (4.13), Page 72, or Theorem 4.3.2, Page 72. In this section we present a few examples.

The first example demonstrates the principle of *image restoration* via frequency filtering.

Example 4.8.1. The left image in Figure 4.4 shows an MR slice, f say, corrupted by a spurious artifact with a more or less periodic structure. Periodicity suggests that we take a closer look at its frequency representation. The second image shows its DFT magnitude, i.e. $|\widehat{f}| = [\mathrm{Re}^2(\widehat{f}) + \mathrm{Im}^2(\widehat{f})]^{1/2}$. Greyvalues have been modified by a monotonic mapping (a so-called 'histogram equalization') for the sake of display. In addition, the origin has been shifted to the middle of the figure. Figure 4.5 shows the lower left part of the DFT magnitude image.

An important observation—typical for *natural images*—is that most frequency information is confined to a central portion of the DFT domain. Apparently low frequencies are most dominant. This expresses the fact that natural

images capture *spatially coherent*, not random objects. Since such objects have a certain extent within which they are more or less 'smooth', they exhibit only weak variability at small scales. Since spatial scale somehow corresponds to inverse frequency, this implies that high frequency components must have low amplitudes.

Based on this heuristic some exceptionally strong peaks located on the diagonals (barely visible as a result of their narrow widths), as well as some vertical line artifacts at $\omega_1 \approx 0$, should be marked as 'suspicious'. In order to remove these, the third image shows a binary frequency mask, $\widehat{\beta}$ say, with which the image's DFT is multiplied. Black and white correspond to pixel values 0 and 1, respectively. The fourth image shows the magnitude of the result, $|\widehat{f}\,\widehat{\beta}|$, again after histogram equalization. The suspicious structures are now suppressed. If we apply the inverse DFT to $\widehat{f}\,\widehat{\beta}$, we obtain the image on the right, in which the spurious pattern has indeed disappeared.

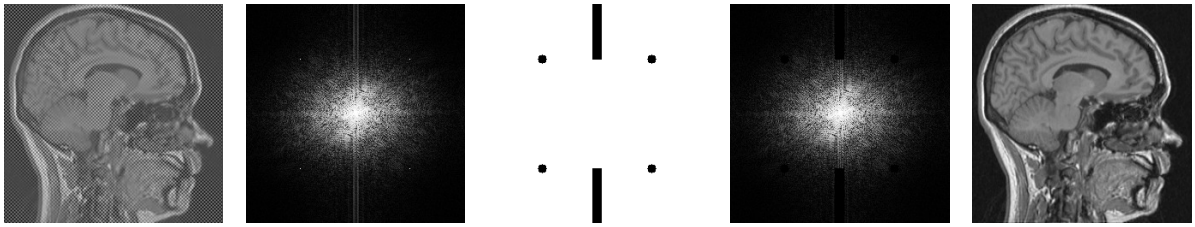


Figure 4.4: Removal of frequency artifacts in Fourier space. The spurious frequency spikes are barely visible. Figure 4.5 shows the lower left part of the DFT magnitude image to reveal the locus of one of the artifacts.

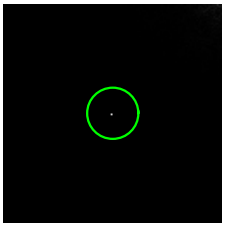


Figure 4.5: Magnification of the lower left quadrant of Figure 4.4 *without* histogram equalization. The centre of the circle contains a spurious frequency spike that, among other artifacts, needs to be removed. The spike subtends only a single pixel (which makes it particularly suspicious), but is evidently quite strong.

The previous example shows the strength of Fourier decomposition, for it would be nearly impossible to devise a filter capable of restoring the image directly in the spatial domain! To get an impression of the complexity of the corresponding spatial restoration filter, see Fig 4.6.

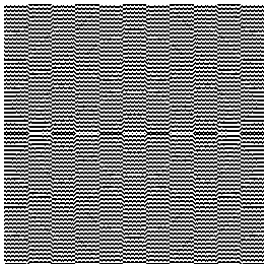


Figure 4.6: Spatial convolution mask corresponding to the frequency restoration filter shown in the middle of Figure 4.4. For clarity only the signs of the filter values are given: Black means negative, white means positive.

The following examples illustrate the use of Eq. (4.13) or Theorem 4.3.2.

Example 4.8.2. The commutative diagram for convolution filtering via the Fourier domain is not only of computational relevance (speed considerations), but is also of interest in those cases where we design our filters in the frequency domain. A simple example is that of an *ideal lowpass filter*. This multiplicative filter depends on a threshold frequency ω_c below which it leaves all information unaltered. All higher frequency components are suppressed entirely, whence the attribute 'ideal'. The general form of an ideal lowpass filter is

$$\widehat{\phi}_{\text{l.p.}}(\omega;\omega_c) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \|\omega\| \leq \|\omega_c\|, \\ 0 & \text{otherwise,} \end{cases} \tag{4.40}$$

in which $\|\omega\|$ denotes a suitable norm in frequency space, recall Section 2.2.3. A natural choice is a rotationally invariant filter (recall that for $\omega \in \mathbb{R}^n$, $\|\omega\|_2 = \sqrt{\omega_1^2 + \ldots + \omega_n^2}$):

$$\widehat{\phi}_{\text{l.p.}}(\omega;\omega_c) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \|\omega\|_2 \leq \|\omega_c\|_2, \\ 0 & \text{otherwise.} \end{cases}$$

Another popular instance is that of a separable filter (recall that for $\omega \in \mathbb{R}^n$, $\|\omega\|_\infty = \max\{|\omega_1|, \ldots, |\omega_n|\}$):

$$\widehat{\phi}_{\text{l.p.}}(\omega;\omega_c) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \|\omega\|_\infty \leq \|\omega_c\|_\infty, \\ 0 & \text{otherwise.} \end{cases}$$

Figure 4.7 illustrates the case of a separable lowpass filter in two dimensions. Since high frequencies mainly encode fine scale details one expects lowpass filtering to yield a blurred image. Recall that this sort of DFT filtering corresponds, at least in the continuum case, to spatial convolution with a sinc-filter.

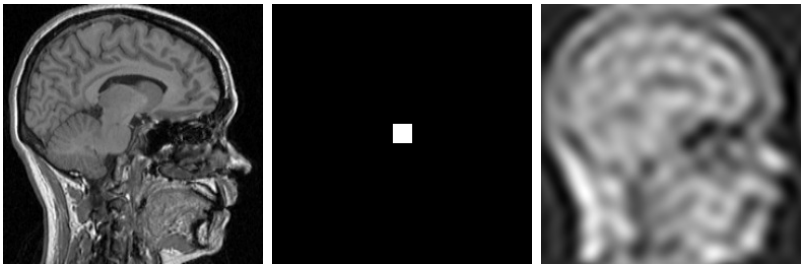


Figure 4.7: Left: Test image. Middle: Ideal lowpass filter defined in the DFT domain. The origin has been shifted to the middle for the sake of display. Right: Lowpass filtered image, obtained by multiplying the DFT of the test image with the ideal lowpass filter (with origins in the DFT domain coinciding), and subsequent DFT inversion.

Ideal lowpass filtering is a somewhat awkward thing to do if the intention is merely to obtain a blurred image. The problem is that there is no gradual attenuation of higher frequencies, but an abrupt cutoff. We know from Section 4.4.3 that such an abrupt cutoff manifests itself in the spatial domain as a blurring with a convolution filter in the form of a sinc-function. Since this is not a nonnegative definite filter, a nonnegative input image may yield an output image that is not everywhere nonnegative[7]. Thus we cannot interpret ideal lowpass filtering as a local, spatially weighted averaging in the spatial domain, because an averaging operator must respect the lower and upper bounds of its operand, in particular it must preserve positivity.

It would be more natural to suppress higher frequencies in a more graceful fashion. An example of this is *Gaussian lowpass filtering*, which brings us back to scale space theory, recall Section 3.4.2.

[7]From the physics of image acquisition many image modalities are known to produce nonnegative pixel values, simply because they are based on some form of *counting*, possibly accompanied by some positivity preserving nonlinearity. Examples are proton density MRI (number of protons per volume element), X-ray attenuation (relative number of transmitted photons), et cetera. Thus one should not blindly rely on the (band-limitedness assumption underlying the) sampling theorem, i.e. Result 4.4.3, Page 82.

Example 4.8.3. Recall Section 3.4.2, in particular the example of Figure 3.13 on Page 57. According to Section 4.7, Eqs. (4.33–4.38), the Fourier transform of a Gaussian is again a Gaussian. Exploiting Diagram (4.13), Page 72, or Theorem 4.3.2, Page 72, we may approximate a Gaussian blurred image by taking the DFT inverse of the product of the DFT of the image and a suitably scaled, unit-amplitude Gaussian. Figure 4.8 shows the procedure, which is completely analogous to that of Figure 4.7. Note that because we know the analytical expression of the Gaussian scale space filter in the Fourier domain, we do not need to apply any DFT to obtain the Fourier filter itself.

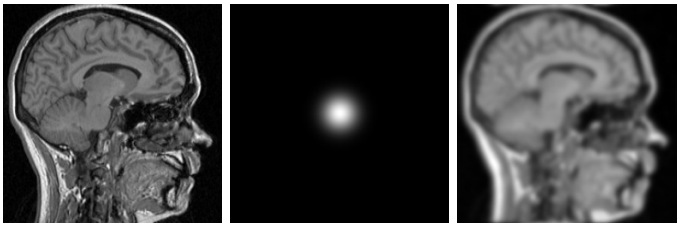


Figure 4.8: Left: Test image. Middle: Gaussian lowpass filter defined in the DFT domain. The origin has been shifted to the middle for the sake of display. Right: Gaussian filtered image, obtained by multiplying the DFT of the test image with the Gaussian lowpass filter (with origins in the DFT domain coinciding), and subsequent DFT inversion. This is one way of obtaining the result shown in Figure 3.13, Page 57.

As opposed to ideal lowpass filtering, Gaussian scale space filtering can be interpreted as a local, spatially weighted averaging in the spatial domain. Thus for lowering resolution the latter is the preferred method.

Instead of lowpass filtering one may want to extract the high frequency information from an image. To this end we define a *highpass filter*, $\widehat{\phi}_{\text{h.p.}}$, as the complementary operator of a lowpass filter:

$$\widehat{\phi}_{\text{l.p.}}(\omega) + \widehat{\phi}_{\text{h.p.}}(\omega) = 1 \quad \text{for all } \omega \in \mathbb{R}^n. \tag{4.41}$$

Example 4.8.4. Figure 4.9 illustrates what happens if we subject an image to highpass filtering. The corresponding lowpass filters are the ones discussed in the previous examples.

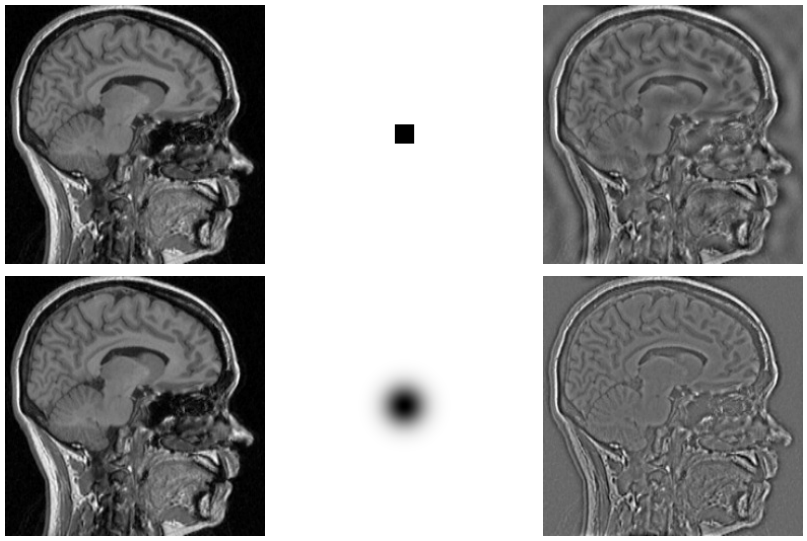


Figure 4.9: Top left: Test image. Top middle: Ideal highpass filter in the DFT domain. The origin has been shifted to the middle for the sake of display. Top right: Highpass filtered image, obtained by multiplying the DFT

of the test image with the ideal highpass filter (with origins in the DFT domain coinciding), and subsequent DFT inversion. Bottom row: Same procedure as shown on top, but for a Gaussian instead of an ideal highpass filter.

By the same token as before one may argue that ideal highpass filtering is not the ideal way of extracting small scale information, despite the fact that high frequency does pertain to small scale spatial variations. The *vox populi* of identifying 'high (low) frequency' with 'fine (coarse) scale' may be misleading. A single frequency component does not encode a spatially confined structure at a scale inversely proportional to the frequency of interest, but captures the *global* spatial variations at that scale. Thus if one is interested in *local* structure one should use the scale space paradigm explained in Section 3.4.2 instead of ideal lowpass/highpass filtering.

Example 4.8.5. Figure 4.10 shows the Fourier method for taking a Gaussian scale space derivative, recall Definition 3.4.3, Page 57, in casu the first order derivative in horizontal direction. Scale is the same as in Figure 4.8. The DFT profile looks similar to the corresponding spatial profile in Figure 3.14, Page 58. To see how it is obtained, recall Theorem 3.4.1, Page 55, and Definitions 3.4.2 and 3.4.3, Pages 56 and 57. For the zeroth order filter we have

$$\phi_\sigma(x,y) = \frac{1}{2\pi\,\sigma^2}\,\exp\left[-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}\right],$$

so that

$$\frac{\partial\phi_\sigma}{\partial x}(x,y) = -\frac{x}{\sigma^2}\,\frac{1}{2\pi\,\sigma^2}\,\exp\left[-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}\right].$$

We already established that

$$\mathscr{F}\left(\phi_\sigma\right)(\omega_x,\omega_y) = \exp\left[-\frac{1}{2}\sigma^2\left(\omega_x^2+\omega_y^2\right)\right],$$

whence, using Result 4.3.1 on Page 69, or the example on Page 68,

$$\mathscr{F}\left(\frac{\partial\phi_\sigma}{\partial x}\right)(\omega_x,\omega_y) = i\omega_x\,\mathscr{F}\left(\phi_\sigma\right)(\omega_x,\omega_y).$$

Thus

$$\mathscr{F}\left(\frac{\partial\phi_\sigma}{\partial x}\right)(\omega_x,\omega_y) = i\omega_x\,\exp\left[-\frac{1}{2}\sigma^2\left(\omega_x^2+\omega_y^2\right)\right]$$

is the filter profile shown in the middle of Figure 4.10 prior to discrete sampling. One can use Result 4.3.1 to obtain the Fourier representation of the entire Gaussian family in a similar fashion, cf. Figure 3.14. Thus if f denotes the original image, and u its scale space representation, then the rightmost image in Figure 4.10 corresponds to

$$\frac{\partial u}{\partial x}(x,y,\sigma) = \mathscr{F}^{-1}\left((\omega_x,\omega_y)\mapsto i\omega_x\,\exp\left[-\frac{1}{2}\sigma^2\left(\omega_x^2+\omega_y^2\right)\right]\mathscr{F}(f)(\omega_x,\omega_y)\right)(x,y).$$

By the same token one can extract all partial image derivatives at any sensible scale of interest. Again, we need not run a DFT to obtain our Fourier filter, as we may simply sample its analytically known Fourier expression.

Figure 4.10: Left: Test image. Middle: Sampled analytical Fourier transform of the horizontal first order partial derivative of the Gaussian scale space filter. The origin has been shifted to the middle for the sake of display.

Right: Filtered image, i.e. the first order horizontal partial derivative at the scale of the underlying Gaussian. This is one way of implementing scale space differentiation.

Figure 4.10 has in common with Figure 4.9 that it reveals grey-level transitions rather than absolute grey-values. It should be clear from the previous discussion on scale versus frequency that, unlike Figure 4.9, Figure 4.10 captures *local* transitions, or 'edges'.

BIBLIOGRAPHY

[1] Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick. *Analysis, Manifolds, and Physics. Part I: Basics*. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1991.

[2] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computing*, 19:297–301, 1965.

[3] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology: Functional and Variational Methods*, volume 2. Springer-Verlag, Berlin, 1988.

[4] B. S. DeWitt. Quantum theory of gravity. II. The manifestly covariant theory. *Physical Review*, 162:1195–1239, 1967.

[5] J. J. Duistermaat. Distributies. Course notes published by the Mathematics Institute, University of Utrecht, The Netherlands, 1992.

[6] L. M. J. Florack. *Image Structure*, volume 10 of *Computational Imaging and Vision Series*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.

[7] O. Fogh Olsen, L.M.J. Florack, and A. Kuijper, editors. *Deep Structure, Singularities and Computer Vision*, volume 3753 of *Lecture Notes in Computer Science*. Springer-Verlag, 2005.

[8] J. Fourier. *The Analytical Theory of Heat*. Dover Publications, Inc., New York, 1955. Replication of the English translation that first appeared in 1878 with previous corrigenda incorporated into the text, by Alexander Freeman, M.A. Original work: "Théorie Analytique de la Chaleur", Paris, 1822.

[9] F. G. Friedlander. *Introduction to the Theory of Distributions*. Cambridge University Press, Cambridge, 1982.

[10] I. M. Gelfand and G. E. Shilov. *Generalized Functions*, volume 1. Academic Press, New York and London, 1968.

[11] I. M. Gelfand and G. E. Shilov. *Spaces of Fundamental and Generalized Functions*, volume 2. Academic Press, New York and London, 1968.

[12] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice-Hall, New Jersey, second edition, 2002.

[13] B. M. ter Haar Romeny and L. M. J. Florack. Front-end vision: A multiscale geometry engine. In S.-W. Lee, H. H. Bülthoff, and T. Poggio, editors, *Biologically Motivated Computer Vision: Proceedings of the First IEEE International Workshop, BMCV 2000 (Seoul, Korea, May 2000)*, volume 1811 of *Lecture Notes in Computer Science*, pages 297–307, Berlin, May 2000. Springer-Verlag.

[14] B. M. ter Haar Romeny, L. M. J. Florack, J. J. Koenderink, and M. A. Viergever, editors. *Scale-Space Theory in Computer Vision: Proceedings of the First International Conference, Scale-Space'97, Utrecht, The Netherlands*, volume 1252 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, July 1997.

[15] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Bul. Univ. Princeton*, 13:49–62, 1902.

[16] B. K. P. Horn. *Robot Vision*. MIT Press, Cambridge, 1986.

[17] M. Kerckhove, editor. *Scale-Space and Morphology in Computer Vision: Proceedings of the Third International Conference, Scale-Space 2001, Vancouver, Canada*, volume 2106 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, July 2001.

[18] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[19] J. J. Koenderink and A. J. van Doorn. Receptive field families. *Biological Cybernetics*, 63:291–298, 1990.

[20] T. Lindeberg. *Scale-Space Theory in Computer Vision*. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[21] M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors. *Scale-Space Theories in Computer Vision: Proceedings of the Second International Conference, Scale-Space'99, Corfu, Greece*, volume 1682 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, September 1999.

[22] H. J. Nussbaumer. *Fast Fourier Transform and Convolution Algorithms*. Springer-Verlag, New York, 1982.

[23] M. Petrou and P. Bosdogianni. *Image Processing—The Fundamentals*. John Wiley & Sons, Ltd., Chichester etc., 1999.

[24] L. Schwartz. *Théorie des Distributions*, volume I, II of *Actualités scientifiques et industrielles; 1091,1122*. Publications de l'Institut de Mathématique de l'Université de Strasbourg, Paris, 1950–1951.

[25] L. Schwartz. *Théorie des Distributions*. Publications de l'Institut Mathématique de l'Université de Strasbourg. Hermann, Paris, second edition, 1966.

[26] F. Sgallari, A. Murli, and N. Paragios, editors. *Scale Space and Variational Methods in Computer Vision: Proceedings of the First International Conference, SSVM 2007, Ischia, Italy*, volume 4485 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, May–June 2007.

[27] J. Sporring, M. Nielsen, L. M. J. Florack, and P. Johansen, editors. *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision Series*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.

[28] G. Strang. *Linear Algebra and its Applications*. Harcourt Brace Jovanovich Inc., third edition, 1988.

[29] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, second edition, 1998.

[30] F. Trêves. *Topological Vector Spaces, Distributions and Kernels*. Academic Press, 1969.

[31] R. L. Wheeden and A. Zygmund. *Measure and Integral: an Introduction to Real Analysis*. Marcel Dekker, New York, 1977.

[32] S. Wolfram. *The Mathematica Book*. Wolfram Media/Cambridge University Press, fourth edition, 1999.