Applied Mathematics in Music Processing

MD

February 1, 2012
**Contents**

1 **Representing sound and the Fourier Transforms**  
   1.1 Fourier Series ........................................... 5  
      1.1.1 Motivation 1: Simple harmonic motion ................. 5  
      1.1.2 Motivation 2: The vibrating string and the wave equation .......... 6  
      1.1.3 Definition of Fourier series and examples ............... 7  
   1.2 Fourier transforms on $\mathbb{Z}$ and $\mathbb{R}$ ............. 14  
      1.2.1 The transition to other domains ...................... 14  
      1.2.2 The discrete Fourier transform ..................... 14  
      1.2.3 The Fourier transform on $\mathbb{R}$ ................. 15  
      1.2.4 Filters and convolution ............................. 16  
      1.2.5 The finite discrete Fourier transform ................. 19  
      1.2.6 Excursus: Periodization ......................... 21  
2 **How does the Music end up on a CD? Sampling and Filtering**  
   2.0.7 Aliasing .............................................. 23  
   2.0.8 The mathematics of aliasing: sampling is periodization in the Fourier  
        domain ................................................. 29  
   2.0.9 Reconstruction from samples ........................... 31  
   2.0.10 Prefiltering of analog signals ........................ 32  
3 **Digging the audio structure: Frames and Redundancy** .......... 33  
   3.1 Introduction - Uncertainty principle and time-frequency molecules .... 33  
   3.2 Approximation ............................................ 34  
      3.2.1 Least squares method ............................... 34  
      3.2.2 Eigenvalues and singular values .................... 38  
      3.2.3 Pseudoinverse: a generalization of matrix inversion .......... 41  
      3.2.4 Pseudoinverse and least squares .................... 45  
   3.3 The method of frames ..................................... 45  
      3.3.1 Analysis of a time-variant signal: Short-time Fourier transform .... 46  
      3.3.2 Gabor frames ....................................... 55
1 Representing sound and the Fourier Transforms

The most basic concept in most signal processing applications is the Fourier transform in its various appearances: continuous Fourier transform, Fourier series, discrete Fourier transform, FFT. The basic idea is always the same: using some kind of "basis", a function, sequence or signal is decomposed into a sum of weighted frequencies. In Chapter 2, the insights gained from the study of Fourier transforms will be applied to the problem of sampling ("how does a CD-player work?") and to a basic understanding of compression (JPEG).

1.1 Fourier Series

"Sinusoids describe many natural, periodic processes"

1.1.1 Motivation 1: Simple harmonic motion

Recall Newton’s second law: The acceleration $a$ of a body is parallel and directly proportional to the net force $F$ and inversely proportional to the mass $m$, i.e., $F = ma = m \cdot x''$. Now consider a tuning fork, that is struck and thus produces a sound. What happens as the tuning fork is struck? It will be deformed and a restoring force $F$ strives to restore the equilibrium, the fork overshoots, etc. This motion produces air pressure waves that are picked up by our ears. How can the motion be modeled?

1. $F$ is proportional to the displacement $x(t)$ from equilibrium:
   \[ F = -kx \] (1.1)
   where $k$ is an elasticity constant (whose unit would be $N/m$, )

2. Since $F$ acts on the tine, it produces acceleration proportional to $F$:
   \[ F = ma, \] (1.2)
   where $m$ is the mass of the tine.

3. Now recall that $a = d^2x/dt^2$, i.e. acceleration is the second derivative of displacement $x$ with respect to time $t$. We thus obtain the ordinary differential equation for the harmonic oscillator:
   \[ \frac{d^2x}{dt^2} = -\frac{k}{m}x(t) \] (1.3)
1. Representing sound and the Fourier Transforms

4. In order to understand the motion of the struck tine, we therefore have to find functions $x(t)$ that are proportional to their second derivative by a negative number $c = -\frac{k}{m}$:

$$\frac{d^2}{dt^2} \sin(\omega t) = -\omega^2 \sin(\omega t) \quad (1.4)$$

and the $\cos(\omega t)$ fulfills (1.3) analogously. In both cases $\omega = \sqrt{\frac{k}{m}}$ and the period of the oscillation is then given by $T = 2\pi \sqrt{\frac{m}{k}} = \frac{2\pi}{\omega} = \frac{1}{f}$, where $f$ is the usual frequency in Hertz (whereas $\omega$ is measured in radians per seconds, thus $\omega = 2\pi f$).

Let us next look at some properties of the sinusoids, i.e. sines and cosines, which we often don’t discriminate since we have $\sin(\omega t + \pi/2) = \cos(\omega t)$.

**Proposition 1.**

1. Sinusoids are closed under time-shift.
2. Sinusoids are closed under addition.
3. Adding sinusoids of close frequencies produces beats.

1.1.2 Motivation 2: The vibrating string and the wave equation

"Trigonometric series can represent arbitrary functions"

A vibration in a string is a wave. Usually a vibrating string produces a sound whose frequency in most cases is constant. Therefore, since frequency characterizes the pitch, the sound produced is a constant note. Vibrating strings are the basis of any string instrument like guitar, cello, or piano. Their behavior is described by the wave equation:

Let the vertical position (displacement) of a point $x$ on a given string at time $t$ be described by the function $y(x,t)$. Then the vibrating string is guided by the following partial differential equation (PDE):

$$\frac{\partial^2 y}{\partial t^2} = \frac{k}{m} \frac{\partial^2 y}{\partial x^2} \quad (1.5)$$

This is the wave equation for $y(x,t)$ and, as above, $k$ is a constant corresponding to tension (stiffness) and $m$ is the mass of the string. We set $c = \sqrt{\frac{k}{m}}$. It is then easy to see (verify!) that a general solution of this PDE must satisfy $y(x,t) = f(t-x/c) + g(t+x/c)$, where $f, g$ are smooth but arbitrary function of one variable. We now make use of the obvious initial conditions $y(0,t) = y(L,t) = 0$, for all (!) $t$, where $L$ is the length of the string under investigation - these conditions simply state that the string is attached at both ends. We then deduce:

$$y(0,t) = 0 \Rightarrow f(t) = -g(t)$$
$$y(L,t) = 0 \Rightarrow f(t - L/c) + g(t + L/c) = f(t - L/c) - f(t + L/c) \text{ for all } t$$

set $t' = t + L/c \Rightarrow 0 = f(t') - f(t' + 2L/c)$, hence $f(t') = f(t' + 2L/c)$ for all $t'$
Therefore, the solutions of (1.5) must be periodic in $t$ with period $\frac{2L}{c}$. We already know the general form of functions with this property, and therefore, we assume that $y$ must be of the form $y(x, t) = e^{2\pi i \frac{c}{2L} t} Y(x)$. We now plug this "Ansatz" into our PDE (1.5) and obtain:

$$
\frac{\partial^2 y}{\partial t^2} = -4\pi^2 \left( \frac{c}{2L} \right)^2 e^{2\pi i \frac{c}{2L} t} Y(x)
$$

$$
\frac{\partial^2 y}{\partial x^2} = e^{2\pi i \frac{c}{2L} t} \frac{\partial^2 Y}{\partial x^2}(x)
$$

$$
\Rightarrow -4\pi^2 \left( \frac{c}{2L} \right)^2 Y(x) = c^2 \frac{\partial^2 Y}{\partial x^2}(x)
$$

$$
\Rightarrow \frac{\partial^2 Y}{\partial x^2}(x) = \frac{\pi^2}{L^2} Y(x)
$$

which is the differential equation we have solved in the previous section! We therefore know that the shape of our string is given by a general sinusoid: $Y(x) = \sin \left( \frac{2\pi x}{L} + \Phi \right)$.

However, in the assumption on the form of $y$, namely, that $y(x, t) = e^{2\pi i \frac{c}{2L} t} Y(x)$, we have so far ignored, that not only the sinusoid $e^{2\pi i \frac{c}{2L} t}$ is periodic with period $\frac{2L}{c}$, but also $e^{2\pi in \frac{c}{2L} t}$, for any integer $n \in \mathbb{Z}$, as we will investigate more precisely in the following.

**Definition 1.** A function $f$ on $\mathbb{R}$ is periodic with period $p$ ($p$-periodic), for $p > 0$, if $f(x + p) = f(x)$ for all $x \in \mathbb{R}$.

**Example 1.** The functions $e^{2\pi in \frac{c}{2L} t}$ are $\frac{2L}{c}$-periodic for any $n \in \mathbb{Z}$, since

$$
e^{2\pi i \frac{c}{2L} (t+\frac{2L}{c})} = e^{2\pi i \frac{c}{2L} t} \cdot e^{2\pi in} = e^{2\pi in \frac{c}{2L} t},$$

since $e^{2\pi in} = 1$ for any $n \in \mathbb{Z}$.

Since we have shown that linear combinations of sinusoids are sinusoids, we come to the conclusion that the general solution for the shape of a string is in fact given by arbitrary combinations of sinusoids of the form $\sin \left( \frac{2\pi nx}{L} + \Phi \right)$.

**1.1.3 Definition of Fourier series and examples**

**Definition 2.** For a $p$-periodic function $f(x)$ that is integrable on $[-\frac{p}{2}, \frac{p}{2}]$, the numbers

$$
a_n = \frac{2}{p} \int_{-p/2}^{p/2} f(x) \cos \left( \frac{2\pi nx}{p} \right) dx, \quad n \geq 0
$$

and

$$
b_n = \frac{2}{p} \int_{-p/2}^{p/2} f(x) \sin \left( \frac{2\pi nx}{p} \right) dx, \quad n \geq 1
$$

are called the Fourier coefficients of $f$. The expression

$$
(S_N f)(x) = \frac{a_0}{2} + \sum_{n=1}^{N} \left[ a_n \cos \left( \frac{2\pi nx}{p} \right) + b_n \sin \left( \frac{2\pi nx}{p} \right) \right], \quad N \geq 0.
$$


is called trigonometric polynomial of degree $N$. The infinite sum

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} \left[ a_n \cos\left(\frac{2\pi nx}{p}\right) + b_n \sin\left(\frac{2\pi nx}{p}\right) \right]$$

is called the Fourier series of $f$.

**Example 2.** The Fourier series of a square wave Consider the $[0,1]$-periodic function

$$f(x) := \begin{cases} 
1 & \text{for } 0 \leq x < \frac{1}{2} \\
-1 & \text{for } \frac{1}{2} \leq x < 1
\end{cases}$$

Then its Fourier series is given by

$$f(x) = \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{1}{(2k-1)} \sin(2\pi(2k-1)x)$$

**Proposition 2.** For a bounded, piecewise continuous function $f$, the Fourier coefficients (1.6) and (1.8) yield the best approximation with a trigonometric polynomial of degree $N$. Furthermore, if $f$ is piecewise smooth with finitely many discontinuities, its Fourier series converges pointwise.
1.1 Fourier Series

Remark 1. Note that best approximation means, that the error which occurs, when approximating a given, \( p \)-periodic functions by a trigonometric polynomial of degree \( N \), as in (1.8), is minimal, if the coefficients \( a_n, b_n \) are the Fourier coefficients. This is an immediate consequence of the fact, that the sinusoids form an orthonormal basis for "all" periodic functions (which are sufficiently nice). We will consider this property in Proposition 3.

The complex version of Fourier series: We can use Euler's formula, 
\[ e^{2\pi i \frac{2}{p} x} = \cos(2\pi \frac{2}{p} x) + i\sin(2\pi \frac{2}{p} x) \]
where \( i \) is the imaginary unit, to give a more concise formula of the Fourier series of a function \( f \):

\[ f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i \frac{2}{p} x}, \tag{1.10} \]
with the Fourier coefficients \(^2\) given by:

\[ \hat{f}[n] = c_n = \frac{1}{p} \int_{-1/2}^{1/2} f(x)e^{-2\pi i \frac{2}{p} x} \, dx. \tag{1.11} \]

if we assume here \( p = 1 \), the above formulas simplify to We can use Euler's formula, 
\[ e^{2\pi inx} = \cos(2\pi nx) + i\sin(2\pi nx) \]
where \( i \) is the imaginary unit, to give a more concise formula of the Fourier series of a function \( f \):

\[ f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi inx}, \tag{1.12} \]
and

\[ \hat{f}[n] = c_n = \int_{-1/2}^{1/2} f(x)e^{-2\pi inx} \, dx. \tag{1.13} \]

We will often denote the Fourier coefficients of a function \( f \) by \( F[n] \) or \( \hat{f}(n) \). More precise explanations on scalar product, (norm and minimal error), orthogonality, and the correspondence to ONBs:

**Proposition 3.** The family of functions \( \{e^{2\pi i k \frac{x}{p}}\}_{k \in \mathbb{Z}} \) is an orthonormal basis of \( L^2([-\frac{p}{2}, \frac{p}{2}]) \).

Alternatively, we state for the real sinusoids: The sines and cosines form an orthogonal set: (note that the constant function is \( \cos(2\pi \frac{m}{p} x) \) for \( m = 0 \)).

\[ \int_{-\frac{p}{2}}^{\frac{p}{2}} \cos(2\pi \frac{m}{p} x) \cos(2\pi \frac{n}{p} x) \, dx = \delta_{mn}, \quad m \geq 0, n \geq 1 \tag{1.14} \]

---

\(^1\)In der Vorlesung haben wir mehrere Beweise für diese Tatsache der Minimalität des Fehlers besprochen. Diese Beweise werden ins Skriptum noch nachgeliefert, da sie strukturell auch durch ihre Verbindung zur Differentialrechnung (Minima finden...) interessant sind.

\(^2\)The Fourier coefficients \( c_n \) are often denoted by \( f[n] \), since \( f \) is the most common notation for the Fourier transform of \( f \).
We now show that this leads to a contradiction. We compute:

\[ \|f - M_N\|_2^2 = \langle f - M_N, f - M_N \rangle = \langle f, f \rangle + \langle M_N, M_N \rangle - 2\Re \langle f, M_N \rangle \]

\[ = \|f\|_2^2 + \int_{-\frac{1}{2}}^{\frac{1}{2}} M_N(t)\overline{M_N(t)} dt - 2\Re \left[ \sum_{k=-N}^{N} c_k e^{2\pi ikt} \right] \]

\[ = \|f\|_2^2 + \sum_k \sum_{k'} c_k \overline{c_{k'}} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi ikt} e^{-2\pi i k't} dt - 2\Re \left[ \sum_{k=-N}^{N} c_k \langle f, e^{2\pi ikt} \rangle \right] \]

\[ = \|f\|_2^2 + \sum_k |c_k|^2 - 2\Re \left[ \sum_{k=-N}^{N} \overline{c_k} \langle f, e^{2\pi ikt} \rangle \right], \]

where the last step follows from the orthogonality of the basis functions \( \{e^{2\pi ikt}\} \). We carry out the same steps for \( S_n \) and obtain:

\[ \|f - S_n\|_2^2 = \|f\|_2^2 + \sum_k |\hat{f}[k]|^2 - 2\Re \left[ \sum_{k=-N}^{N} \hat{f}[k] \langle f, e^{2\pi ikt} \rangle \right] \]

\[ = \|f\|_2^2 + \sum_k |\hat{f}[k]|^2 - 2 \sum_{k=-N}^{N} |\hat{f}[k]|^2 = \|f\|_2^2 - \sum_k |\hat{f}[k]|^2 \]

\(^3\text{This is the proof for Übungsbeispiel 9b} \)
1.1 Fourier Series

Hence, our assumption (1.17) is equivalent to assuming

\[ \sum_{k} |c_k|^2 - 2 \Re \{ \sum_{k=-N}^{N} c_k \langle f, e^{2\pi i k t} \rangle \} < - \sum_{k} |\hat{f}[k]|^2 \]

for some \( c_k, k = -N, \ldots, N \). We rewrite this as

\[ \sum_{k} |c_k|^2 - 2 \Re \{ \sum_{k=-N}^{N} c_k \hat{f}[k] \} + \sum_{k} |\hat{f}[k]|^2 < 0 \]

hence

\[ \sum_{k} |(c_k|^2 - 2 \Re \{ c_k \hat{f}[k] \} + |\hat{f}[k]|^2 = \sum_{k} |c_k - \hat{f}[k]|^2 < 0 \]

and obviously the sum of positive values can never be negative. This contradiction concludes the proof.

**Example 3. The Fourier series of a sum of sinusoids**

We consider the functions \( f_1(t) = \sin 2\pi \omega_0 t \) and \( f_2(t) = \cos 2\pi 3\omega_0 t \), for arbitrary \( \omega_0 \in \mathbb{N} \) and \( h(t) = f_1(t) + \frac{1}{2} \cdot f_2(t) \). We want to compute and interpret the Fourier series of these three functions. Obviously, \( f_1 \) and \( f_2 \) are pure sinusoids with frequencies \( \omega_0 \) and \( 3\omega_0 \), respectively, hence, with periods \( p_1 = \frac{1}{\omega_0} \) and \( p_2 = \frac{1}{3\omega_0} \). It is clear, that \( f_2 \) is also periodic with the longer period \( \frac{1}{\omega_0} \). (Check this by invoking the definition of periodic functions!)

We first consider the Fourier coefficients of \( f_1 \). Since its period is \( \frac{1}{\omega_0} \), we are looking for the coefficients \( a_n, b_n \) in the expansion

\[ f_1(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(2\pi n \omega_0 t) + b_n \sin(2\pi n \omega_0 t)], \quad n \geq 0. \quad (1.18) \]

We can now compute the coefficients (do it!!), according to the formulas given in Definition 2, to find, using the orthogonality relations, that \( a_n = 0 \forall n \) and \( b_1 = 1, b_n = 0 \) for \( n \neq 1 \). Alternatively we simply look at the given form of \( f_1 \) and argue that, since the expansion in an orthogonal system is unique, the coefficients have to be of the very same form. As a third version, compute the Fourier coefficients according to (1.13). We immediately derive (do it!) that \( c_1 = \hat{f}[1] = \frac{1}{2i} \) and \( c_{-1} = \hat{f}[-1] = -\frac{1}{2i} \), which leads us directly to the expression of the sine-function via Euler's formula:

\[ f_1(t) = \sin 2\pi \omega_0 t = \frac{e^{2\pi i \omega_0 t} - e^{-2\pi i \omega_0 t}}{2i} ! \]

Let us first interpret these findings: obviously, the coefficients \( a_n, b_n \) in the Fourier series express the "contribution" or energy of the cosine (or sine) function to the periodic signal we wish to express. If we use the complex form, we split the energy contained in one sinusoid into a positive and a negative part of equal absolute value (in the case of real functions). If we use the real part, the contributions to "one frequency component" may
be split in cosine and sine parts. Since this is usually more complicated, the complex form is usually preferred.

We now turn to $f_2$, periodic with period $p_2$, hence, if we consider the orthonormal basis \{cos($2\pi n3\omega_0$), $n \geq 0$\} \cup \{sin($2\pi n3\omega_0$), $n \geq 1$\} in complete analogy to before, we compute, or derive from the properties of our orthonormal basis, that $a_n = 0 \forall n \neq 1$ and $a_1 = 1$, $b_n = 0 \forall n$. On the other hand, if we consider the basis \{cos($2\pi n\omega_0$), $n \geq 0$\} \cup \{sin($2\pi n\omega_0$), $n \geq 1$\}, which we will also have to use for $h(t)$, we find that

\[
f_2(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(2\pi n\omega_0 t) + b_n \sin(2\pi n\omega_0 t)], \quad n \geq 0.
\]

with $a_n = 0 \forall n \neq 3$ and $a_3 = 1$, $b_n = 0 \forall n$! We also derive the coefficients of the complex form:

Combining all the above considerations, we now derive the Fourier coefficients of $h(t)$ according to (1.13): $c_1 = \hat{f}[1] = \frac{1}{2\pi}$, $c_{-1} = \hat{f}[-1] = -\frac{1}{2\pi}$, $c_3 = \hat{f}[3] = \frac{1}{2}$, $c_{-3} = \hat{f}[-3] = \frac{1}{2}$. The absolute values of these Fourier coefficients as well as the functions $h(t)$ are shown in Figure 3 for $\omega_0 = 10$. Please also write out the real form of the Fourier series and verify that the two forms are identical. In the figure, note that the x-axis is labeled with the frequencies in Hertz. Of course, this is an interpretation of our observation that the coefficients in the Fourier series correspond to the pure frequencies in the function (signal) of interest: $c_0$ corresponds to $0 \cdot \omega_0 Hz$, $c_1$ corresponds to $1 \cdot \omega_0 Hz$, etc.

From general properties of ONBs we can now easily deduce the following properties of Fourier series:

**Proposition 4** (Parseval Identity).

\[
(f, g)_{L^2([-\frac{p}{2}, \frac{p}{2}])} = p \sum_{k \in \mathbb{Z}} \hat{f}(k) \bar{\hat{g}(k)} =: p(\hat{f}, \hat{g})_\ell^2
\]

In particular, setting $f = g$, it follows, that

\[
\|f\|_{L^2([-\frac{p}{2}, \frac{p}{2}])}^2 = (f, f)_{L^2([-\frac{p}{2}, \frac{p}{2}])} = \int_{-\frac{p}{2}}^{\frac{p}{2}} |f(x)|^2 dx = p \sum_{k \in \mathbb{Z}} |\hat{f}(k)|^2.
\]

**Proof.** A direct proof, assuming that the interchange of sum and integral is justified:

\[
(f, f)_{L^2([-\frac{p}{2}, \frac{p}{2}])} = \int_{-\frac{p}{2}}^{\frac{p}{2}} f(x) \overline{g(x)} dx
\]

\[
= \int_{-\frac{p}{2}}^{\frac{p}{2}} f(x) \sum_{n \in \mathbb{Z}} G[n] e^{2\pi i \frac{n}{p} x} dx
\]

\[
= \sum_{n \in \mathbb{Z}} \bar{G}[n] \int_{-\frac{p}{2}}^{\frac{p}{2}} f(x) e^{-2\pi i \frac{n}{p} x} dx = p \sum_{k \in \mathbb{Z}} F[k] \bar{G}[k]
\]

\[
= p \sum_{k \in \mathbb{Z}} \hat{f}(k) \bar{\hat{g}(k)} =: p(\hat{f}, \hat{g})_\ell^2.
\]

\[\square\]
1.1 Fourier Series

Figure 1.2: Fourier coefficients of the sum of two sinusoids
1 Represen ting sound and the Fourier Transforms

1.2 Fourier transforms on $\mathbb{Z}$ and $\mathbb{R}$

1.2.1 The transition to other domains

We first introduced the Fourier series, since they are, in a certain sense, the most natural instance of Fourier transforms. The basic idea should be clear by now: a (periodic, so far) function can be represented by a sum of weighted sinusoids, and the sinusoids can be interpreted as the frequencies present in the function (signal). We will now push this concept a bit further. First, we will simply turn around the interpretation of the two variables involved in the definition of Fourier series, namely time $t$ and frequency, which has so far been chosen to live on a discrete subset of $\mathbb{R}$ and labeled by the integers. The next step is of vital importance for understanding the world of digital signal processing, which is, in fact behind almost any modern technical tools we use. Indeed, if we assume, that the frequency information contained in a signal is contained in an interval of finite length, in other words, if the signal (function) is band-limited, we may - mutatis mutandis - expand its frequency-information in a "Fourier series", and the corresponding coefficients will then contain the time-information. However, as we have seen so far, this is discrete information, indexed by $k \in \mathbb{Z}$ - in Section 1.2.2 we thus arrive naturally at a concept of Fourier transform for discrete signals - as a dual concept of the Fourier series.

On the other hand, and complementary to the approach just described, we may think of a periodic time-signal for gradually growing period, which means that, for $p \to \infty$, the size of $k/p$ in the definition of the sinusoids $\{e^{2\pi i k \frac{s}{p}}\}_{k \in \mathbb{Z}}$ becomes infinitely small. This idea leads to Fourier transforms on $\mathbb{R}$, introduced in Section 1.2.3, with an integral replacing the sum in the representation.

1.2.2 The discrete Fourier transform

Let $f : \mathbb{Z} \mapsto \mathbb{C}$ be a function defined on the integers. We consider the complex exponentials $e^{2\pi is}$, $n \in \mathbb{Z}$, $s \in \mathbb{R}$ and observe immediately, that

$$e^{2\pi i (s+m)n} = e^{2\pi isn} \text{ for all } n, m \in \mathbb{Z}.$$  

In other words, the exponentials $e^{2\pi in}, e^{2\pi i(s\pm 1)n}, e^{2\pi i(s\pm 2)n}, \ldots$ cannot be distinguished if $n$ are integer values. Hence, in order to avoid ambiguity, we will synthesize $f$ from $e^{2\pi is/p}$, for $0 \leq s < p$ and $n \in \mathbb{Z}$.

**Definition 3** (DFT). The discrete Fourier transform of a (suitably regular) function $f$ on $\mathbb{Z}$ is defined as

$$\hat{f}(s) = F(s) = \frac{1}{p} \sum_{n=-\infty}^{\infty} f[n] e^{-2\pi isn/p} \text{ for } 0 \leq s < p \quad (1.21)$$

$f$ can then be written as

$$f[n] = \int_{s=0}^{p} F(s)e^{2\pi isn/p} ds. \quad (1.22)$$

14
Note that the Fourier transform \( \hat{f} = F \) of a discrete-valued function is a function on the circle with diameter of length \( p \).

**Remark 3.** Note that normally \( p \) is set to 1, so that the frequencies that occur in a signal are normalized, on the unit circle. We will see later, when we discuss sampling (in fact, any signal on \( \mathbb{Z} \) is a digital, hence sampled signal, unless it stems from a, inherently discrete process, e.g. a time-series of stock exchange values), that \( \frac{1}{2} \) corresponds to the highest frequency that occurs in a real signal. The frequencies in the interval \( \left[ \frac{1}{2}, 1 \right] \) are then the negative frequencies.

We introduced \( p \) in the above definition to guarantee generality and to emphasize the parallelism to Fourier series.

**Example 4 (Dirac Impulse).** Consider the function \( \delta \) defined on \( \mathbb{Z} \), that is equal to 0 everywhere, except for \( \delta[0] = 1 \). It is then easy to see, that the DFT of \( \delta \) is given by \( \hat{\delta}(s) = \frac{1}{p} \) for all \( s \in [0, p] \).

**Example 5 (Sinusoid).**

### 1.2.3 The Fourier transform on \( \mathbb{R} \)

We now consider integrable functions on \( \mathbb{R} \).

**Remark** on the spaces \( L^1 \) and \( L^2 \) of integrable and square-integrable functions.

**Definition 4.** The Fourier transform of a function \( f \in L^1(\mathbb{R}) \) is defined by

\[
\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i \omega t} dt.
\]

(1.23)

Note that, if \( f \) is integrable, the integral in (1.23) converges and

\[
|\hat{f}(\omega)| \leq \int_{-\infty}^{\infty} |f(t)| dt < \infty.
\]

**Proposition 5 (Inverse Fourier transform).** If \( \hat{f} \in L^1(\mathbb{R}) \), then \( f \) is given by the inverse Fourier transform of \( \hat{f} \):

\[
f(t) = \int_{-\infty}^{\infty} \hat{f}(\omega)e^{2\pi i \omega t} d\omega
\]

(1.24)

**Remark 4.** Note that the Fourier transform is usually extended to all functions in \( L^2(\mathbb{R}) \) by using a density argument, similar to our approach in the proof of Proposition 3. Then, as before, an inner product can be defined on \( L^2(\mathbb{R}) \), and most arguments work similar to the case of periodic functions.

**Example 6 (Fourier transform of the box function).** Consider the function

\[
\Pi(x) := \begin{cases} 
1 & \text{for } -\frac{1}{2} < x < \frac{1}{2} \\
0 & \text{else}
\end{cases}
\]
To compute the Fourier transform, first note that $\Pi$ is even, so that we can omit the sine-part (generally we can observe, that the Fourier transform of even (symmetric) functions is always real! On the other hand, the Fourier transform of real functions is symmetric and we only have to consider the positive frequencies. This property is heavily exploited in the processing of speech and music signals, which are always real.) We therefore have

\[
\hat{\Pi}(\omega) = \int_{\mathbb{R}} \Pi(x)e^{-2\pi i x \omega}dx = \int_{\mathbb{R}} \Pi(x) \cos(2\pi i x \omega)dx
\]

\[
= \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos(2\pi i x \omega)dx = \frac{\sin(2\pi i x \omega)}{2\pi \omega} \big|_{x=-\frac{1}{2}}^{\frac{1}{2}}
\]

\[
= \frac{\sin(\pi \omega)}{2\pi \omega} - \frac{\sin(-\pi \omega)}{2\pi \omega} = \frac{\sin(\pi \omega)}{\pi \omega} =: \text{sinc}(\omega)
\]

Note that $\text{sinc}(x)$ can be defined as $\text{sinc} = \frac{\sin(x)}{\pi x}$ only for $x \neq 0$. However according to L'Hôpital's rule, we have that $\lim_{x \to 0} \text{sinc}(x) = 1$, since, for any open interval $I$ around 0, we have $h'(x) = 1 \neq 0$ for $h(x) = x$ and, since $\frac{d}{dx} \sin x = \cos x$ and $\lim_{x \to 0} \cos(x) = 1$, we have $\lim_{x \to 0} \frac{\sin x}{x} = 1$, and so $\lim_{x \to 0} \frac{\sin x}{x} = 1$. More directly, we may consider the Taylor series $\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}$, such that $\text{sinc}(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n}$ for $x \neq 0$, and convergence to 1 is obvious.

We will next address two very basic operators that can act on a function or signal, namely translation, or time-shift and modulation, or frequency-shift.

**Example 7** (Translation and Modulation). For any real number $x_0$, if $g(x) = T_{x_0}f(x) := f(x-x_0)$, then $\hat{g}(\omega) = e^{-2\pi i x_0 \omega} \hat{f}(\omega)$.

For any real number $\omega_0$, if $g(x) = M_{\omega_0}f(x) := e^{2\pi i \omega_0 x} f(x)$, then $\hat{g}(\omega) = \hat{f}(\omega - \omega_0)$.

**Example 8** (Dilation). Let $a \neq 0 \in \mathbb{R}$. Set $g(x) = D_a f(x) := f(ax)$. Let $\hat{f}$ be the Fourier transform of $f$. Then, the Fourier transform of $g$ is given by

\[
\hat{g}(\omega) = \frac{1}{a} \hat{f}\left(\frac{\omega}{a}\right).
\]

### 1.2.4 Filters and convolution

We all know filters, since they are all around us. Every room is a filter, our own mouth is a filter, and of course filters are part of any modern audio equipment. Light is filtered by the air etc.

If we think about the characteristics of filters, then one of the most striking one is the fact that it shouldn’t matter whether a signal is filter at an earlier time or later on. In other words, a filter is a time-invariant system. Let us denote our filter by $L$, and we assume that any input signal $f$ is then mapped to an output $Lf$. We will hope to work with linear filters, so that we arrive at the class of linear, time-invariant systems.

**Definition 5** (Linear, time-invariant (LTI) systems). A linear operator $L$ that maps functions $f \in V$ to $Lf \in V$, where $V$ is a vector space, is called time-invariant, if

\[
L(f(t-u)) = L(f)(t-u), \text{ equivalently: } L(T_u f) = T_u(Lf).
\]

16
1.2 Fourier transforms on $\mathbb{Z}$ and $\mathbb{R}$

Let us now look at a very fundamental concept, the impulse response. Any LTI-system is completely characterized by its impulse response. That is, for any input function, the output function can be calculated in terms of the input and the impulse response. The impulse response of a linear transformation is the image of Dirac’s delta function under the transformation.  

We now consider the mathematical derivation of impulse response of an LTI-system $L$. Note that

$$f(t) = \int f(u)\delta_u(t)du = \int f(u)\delta(t-u)du$$

(1.25)

hence, because $L$ is linear

$$Lf(t) = \int f(u)L\delta_u(t)du$$

(1.26)

Finally, we use the last property of $L$, namely time-invariance, to see that $L\delta_u(t) = L(\delta(t-u)) = (L\delta)(t-u)$, hence

$$Lf(t) = \int f(u)L\delta_u(t)du = \int f(u)(L\delta)(t-u)du$$

(1.27)

Setting $h(t) := (L\delta)(t)$, we achieve

$$Lf(t) = \int f(u)h(t-u)du =: h \ast f.$$  

(1.28)

As we see from (1.28), an LTI-system is completely characterized by its impulse response.

**Definition 6 (Impulse Response).** Let $L$ be an LTI-system. Its impulse response is defined as $h(t) = L\delta(t)$.

**Example 9 (Discrete Impulse response).**

**Definition 7 (Convolution).** The convolution of two (suitably regular and decaying\(^5\)) functions $f,g$ on $\mathbb{R}$ is formally defined by

$$(f \ast g)(t) = \int f(u)g(t-u)du = \int g(u)f(t-u)du = (g \ast f)(t)$$

(1.29)

For functions $f,g$ on $\mathbb{Z}$, we define

$$(f \ast g)[n] = \sum_{m=-\infty}^{\infty} f[m]g[n-m]du = \sum_{m=-\infty}^{\infty} g[m]f[n-m]du = (g \ast f)[n]$$

(1.30)

---

\(^4\)In practical situations, it is not possible to produce a true impulse used for testing. Therefore, some other brief, explosive sound is sometimes used as an approximation of the impulse. In acoustic and audio applications, impulse responses enable the acoustic characteristics of a location, such as a concert hall, to be captured. These impulse responses can be used in applications to mimic the acoustic characteristics of a particular location.

\(^5\)"Decaying" heißt auf Deutsch "abklingend", zur Erinnerung, damit das Integral überhaupt definiert ist, müssen die beteiligten Funktionen schnell genug gegen $0$ gehen.
1 Representing sound and the Fourier Transforms

For functions \( f, g \) on \( \mathbb{C}^N \), we define

\[
(f * g)[n] = \sum_{m=0}^{N-1} f[m]g[n-m]du = \sum_{m=-\infty}^{\infty} g[m]f[n-m]du = (g * f)[n].
\] (1.31)

**Example 10.** Linear averaging over \([-T,T]\):

\[
Lf(t) = \frac{1}{2T} \int_{t-T}^{t+T} f(u)du = \frac{1}{2T} \int_{t}^{1}_{-T,T} l(t-u)f(u)du = (1_{[-T,T]}*f)(t)
\]

**Excursus: eigenfunctions**

In mathematics, an eigenfunction of a linear operator \( L \) defined on some function space is any non-zero function \( h \) such that

\[
Lh = \lambda h
\]

for some scalar \( \lambda \) the corresponding eigenvalue. In the theory of of signals and systems, the eigenfunction of a system is the signal \( h \) which produces a scalar multiple (with a possibly complex scalar \( \lambda \in \mathbb{C} \) of itself as response to the system. Now assume that there is an ONB \( \{\varphi_k\}_{k \in \mathbb{Z}} \) of \( V \) of eigenfunctions of a mapping (system, operator) \( L \), i.e.

\[
f = \sum_k c_k \varphi_k, \text{ for all } f \in V.
\]

Then:

\[
Lf = \sum_k c_k L\varphi_k = \sum_k \lambda_k c_k \varphi_k
\]

In other words, \( L \) can be evaluated as a multiplication on the coefficients of the function’s expansion. Formel ..

**End of excursus.**

Since we have already seen, that complex exponentials provide nice signal expansion, with a natural interpretation, let us next consider, what happens to them, if we let an LTI system act on them. Intuitively, if we recall, that LTI systems "are filters", we should expect that complex exponentials corresponding to a particular frequency should merely be amplified, damped and maybe there phase can be shifted. Now, for an LTI system \( L \) with impulse response \( h \) we have in fact:

\[
Le^{2\pi i \omega t} = \int_u e^{2\pi i \omega u} h(t-u)du = \int_u e^{2\pi i \omega (t-u)} h(u)du = e^{2\pi i \omega t} \hat{h}(\omega)
\] (1.32)

More generally, we have the following
1.2 Fourier transforms on $\mathbb{Z}$ and $\mathbb{R}$

**Proposition 6** (Fourier transformation of convolution). Let $f, g$ absolutely integrable (summable), then $f \ast g$ is also absolutely integrable (summable) and

$$
\hat{f \ast g} = \hat{f} \cdot \hat{g}
$$

(1.35)

**Example 11.** Convolution of box functions.

**Example 12.** Convolution with sinc

1.2.5 The finite discrete Fourier transform

So far, we have been dealing with functions/signals that are

1. continuous in time and have infinite duration; they have a Fourier transform that is continuous in time and has infinite bandwidth ("duration in frequency").

2. discrete in time and have infinite duration; they have a Fourier transform that is continuous in frequency and has finite bandwidth.

3. continuous in time and have finite duration (in other words; they are periodic); they have a Fourier transform that is discrete in frequency and has infinite duration.

We notice that finite duration (or, equivalently, periodicity) in one domain (i.e. in time or frequency) leads to discreteness in the other domain. We now address the fourth case, which is the case of signals, that are both finite in duration and discrete, or, discrete and periodic.

**Remark 5.** We want to point out that finite duration is equivalent to periodicity only in the sense, that the entire information that is contained in the signal is actually contained in an interval of finite length. In this sense, a periodic function can be identified with a function supported on an interval of finite length or on the torus $T = \{ z \in \mathbb{C} : |z| = 1 \}$. As we will see later, in the context of sampling, we may have to distinguish meticulously between periodic signals and signals that are supported in an interval of finite duration and are zero elsewhere.

**Definition 8** (Finite discrete Fourier transform). The finite discrete Fourier transform of $f \in \mathbb{C}^N$, i.e. of a vector of $N$ complex numbers is given by

$$
\mathcal{F}' f[k] = \hat{f}[k] = F[k] = \frac{1}{N} \sum_{n=0}^{N-1} f[n] \cdot e^{-i2\pi \frac{k}{N}n}.
$$

(1.36)

The inverse transform yields the expansion of $f$ as

$$
\mathcal{F}' \mathcal{F}' f[n] = \mathcal{F} \hat{f}[n] = F[n] = \sum_{k=0}^{N-1} F[k] \cdot e^{i2\pi \frac{n}{N}k}.
$$

(1.37)
1 Representing sound and the Fourier Transforms

Remark 6. The Fourier transforms we discussed so far, gave us information about the amount of any pure frequency, i.e. complex exponentials, present in a given signal. Obviously, the sinusoid must have the same basic properties as the underlying signal under consideration: for periodic signals with a certain period $p$, we only considered sinusoids with the same period, for discrete-time signals we only considered discrete-time sinusoids, whereas, for continuous time signals of infinite duration, any complex exponential is a candidate in the expansion (1.23).

For the finite, discrete signals, which are in fact vectors in $\mathbb{C}^N$, we may ask, how many complex exponentials are eligible for the definition of a corresponding Fourier transform. We have two criteria:

(a) they should be periodic with length $N$, that is, we require that

$$e^{2\pi is(n+N)} = e^{2\pi isn}, \text{ for all } n,$$

which means that $s = \frac{k}{N}$.

(b) we observe that for all $m \in \mathbb{Z}$:

$$e^{2\pi ik \frac{n}{N}} = e^{2\pi i(k+mN) \frac{n}{N}}, \text{ for all } n,$$

which is a similar phenomenon as observed for discrete-time sinusoids before. This means, that, since $s = \frac{k}{N}, s = \frac{k+N}{N}, s = \frac{k+2N}{N}, \ldots$ all give the same signals, we have only $N$ distinct sinusoids adequate for analyzing our $N$-finite, discrete signals, namely $e^{2\pi ik \frac{n}{N}}$, for $k = 0, \ldots, N-1$.

Of course, this is exactly what you should have expected: since the complex exponentials have provided ONBs so far, they should provide an orthonormal basis for $\mathbb{C}^N$ as well. Obviously, this means, that there should be $N$ of them.

Example 13. Note that for $k = 0$, $e^{2\pi ik \frac{n}{N}}$ is constantly equal to 1, then, the rotation of the vector $e^{2\pi ik \frac{n}{N}}$, that rotates, as $n$ goes from 0 to $N - 1$, accelerates with growing $k$: $k = 1$ corresponds to a single rotation, $k = 2$ to 2 rotations, etc., up to $N/2$, from where the frequencies decrease, since they become negative.

We next show, that the vectors $e^{2\pi ik \frac{n}{N}}$ in fact form an orthogonal basis.

Proposition 7. The vectors $s_k, k = 0, \ldots, N - 1$, with entries $s_k[n] = e^{2\pi ik \frac{n}{N}}$ are orthogonal in $\mathbb{C}^N$. The set $\left\{ \frac{1}{\sqrt{N}} s_k, k = 0, \ldots, N - 1 \right\}$ is an ONB.

Proof.

$$\langle s_k, s_l \rangle = \sum_{n=0}^{N-1} s_k[n] \overline{s_l[n]} = \sum_{n=0}^{N-1} e^{2\pi ik \frac{n}{N}} e^{-2\pi il \frac{n}{N}} = \sum_{n=0}^{N-1} e^{2\pi i(k-l) \frac{n}{N}} = 1 - e^{2\pi i(k-l) / N} \frac{1}{1 - e^{2\pi i(k-l) / N}} (1.38)$$

20
where the last step follows from the well-known formula for geometric series: \( \sum_{n=0}^{N-1} z^n = \frac{1-z^N}{1-z} \). Now, (1.38) is zero, if \( k \neq l \), and for \( k = l \), we evaluate the sum as \( \sum_{n=0}^{N-1} e^{2\pi i (k-l) \frac{n}{N}} = 1 \). Therefore, the normalization \( \frac{1}{\sqrt{N}} s_k \) leads to
\[
\langle \frac{1}{\sqrt{N}} s_k, \frac{1}{\sqrt{N}} s_k \rangle = \frac{1}{N} \| s_k \|^2 = 1.
\]

**Example 14.**

- The Delta Function
- The Constant Function
- The Delta train or Dirac comb on \( \mathbb{C}^N \)

When \( m = 1, 2, \ldots \) divides \( N \), we define the Dirac comb as
\[
\mathbb{D}_m[n] = \begin{cases} 1 & \text{if } n = 0, \pm m, \pm 2m, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{1.39}
\]

Here, \( m \) specifies the spacing between the "teeth" hence \( m' := N/m \) is the number of teeth. We can easily verify that \( \mathbb{D}_m \) has the Fourier transform
\[
\hat{\mathbb{D}}_m[k] = \frac{1}{m} \mathbb{D}_{N/m}[k]. \tag{1.40}
\]

- Periodicity on \( \mathbb{C}^N \):
  Let \( N = m \cdot m' \), \( m, m' \in \mathbb{N}^+ \) and assume that \( f \) is \( m \)-periodic on \( \mathbb{C}^N \). We show that \( \hat{f}[k] = 0 \) if \( k \) is not a multiple of \( m' \):

Since \( f \) is \( m \)-periodic, we have \( f[n + m] - f[n] = 0 \). Now, since, for \( n_0 \in \mathbb{Z} \),
\[
g[n] = f[n - n_0] \text{ has the Fourier transform } \hat{g}[k] = e^{-2\pi ikn_0/N} \hat{f}[k],
\]
we may write
\[
(e^{2\pi ikm/N} - 1) \hat{f}[k] = (e^{2\pi ik/m'} - 1) \hat{f}[k] = 0
\]
hence \( \hat{f}[k] = 0 \) if \( m' \neq k \).

1.2.6 Excursus: Periodization

Any function \( f \) that decays rapidly (in more technical terms: that is at least absolutely integrable/summable) may be periodized by summation of translates \( \ldots, f(x+2p), f(x+p), f(x), f(x-p), f(x-2p), \ldots \) to produce the \( p \)-periodic function:
\[
g(x) := \sum_{m \in \mathbb{Z}} f(x - mp), \ -\infty < x < \infty \tag{1.41}
\]

Similarly, for \( f \) defined on \( \mathbb{Z} \), we obtain a function \( \gamma \) on \( \mathbb{C}^N \) by
\[
\gamma[n] := \sum_{m \in \mathbb{Z}} f[n - mN], \ n \in \mathbb{Z} \tag{1.42}
\]
1 Representing sound and the Fourier Transforms

**Definition 9 (Dirac Impulse).** The delta distribution $\delta$ on $\mathbb{R}$ is a functional that yields the value $f(0)$ of a given function by integration:

$$\int_{-\infty}^{\infty} f(x)\delta(x) \, dx = f(0)$$

Then, the shifted Dirac gives:

$$\int_{-\infty}^{\infty} f(x)\delta(x_0 - x) \, dx = (\delta \ast f)(x_0) = f(x_0)$$

The Dirac comb (impulse train) is a periodic distribution constructed from Dirac delta functions as

$$X_T(t) \overset{\text{def}}{=} \sum_{k=-\infty}^{\infty} \delta(t - kT)$$

The integral notation is a convenient "abuse of notation", and not a standard (Riemann or Lebesgue) integral.

Now, since $f(x - mp) = (T_{mp} \delta \ast f)(x)$, the periodization $g$ of $f$ can be written as

$$g(x) := \sum_{m \in \mathbb{Z}} (T_{mp} \delta \ast f)(x) = (\Pi_p \ast f)(x) \quad (1.43)$$
2 How does the Music end up on a CD? Sampling and Filtering

In the previous chapter we introduced continuous and discrete-time signals in a somewhat unrelated manner. This chapter deals with the core of modern digital signal processing: the idea and the basic theory of sampled signals. The principal idea is the following: which conditions of a continuous signal guarantee perfect reconstructions from discrete signal samples?

In order to understand the principal idea, let us first look at what happens to the Fourier transform, if we sample a signal as to obtain $f_d$ from $f$. In Figure 2.1, you see the plot of the excerpt of a (pseudo-)continuous signal (a piano sound), with its Fourier transform. In the lower plots, a rather coarsely sampled signal (Sampling rate 11025 samples per second) and its Fourier transform are shown. It should be immediately obvious, what happens to the Fourier transform, if we sample $f$: the Fourier transform $\hat{f}$ of $f$ is periodic!

So, the answer to the next question, namely, how to obtain the original signal from the sampled version, should be really easy: since the sampling process leads to repeated copies of the (hopefully bandlimited) spectrum, all we need to do is multiply with a lowpass filter in order to get rid of the unwanted copies: $\hat{f} = \hat{f}_d \cdot \Pi$, hence $f = f_d \ast \Pi$. Here we are intentionally sloppy and don’t specify any of the involved parameters, since we only want to get across the basic idea - and this seems almost perfect!

However, if we look a bit closer at the spectrum of $f$, namely, if we apply a logarithmic scale (which actually corresponds to our perception of audio), we can see, that the spectrum of $f$ has not actually dropped to anything close to 0, see Figure 2.2 so, what will happen to the frequencies above the cut-off? In fact, if we don’t suppress them by highpass-filtering before the sampling process, those samples will show up as - usually unwanted - aliases in the lower frequency bands.

2.0.7 Aliasing

One of the limitations of discrete-time sampling is an effect called aliasing. An example of aliasing can be seen in old movies, e.g. when watching wagon but also car wheels: the wheels appear to go in reverse. This phenomenon can be observed if the rate of the wagon wheel’s spokes spinning approaches the rate of the sampler (the camera operating at about 30 frames per second) \(^1\).

The same thing happens in data acquisition between the sampler and the signal we are sampling. For an example, have a look at Figure 2.3. Here, the effect of undersampling

\(^1\)This effect is even called "wagon-wheel effect".
Figure 2.1: Subsampling and resulting Fourier transform
Figure 2.2: Subsampling and resulting Fourier transform
is immediately obvious: the sinusoid of 330Hz appears as a sinudoid with much lower frequency, namely 30Hz. In the lower plot, the wrong sampling rate of 320Hz maps the frequency 330Hz to 10Hz. We will now study a simple case of this phenomenon mathematically.

**Example 15.** Consider a complex exponential (a phasor) with frequency \( \omega_0 \), i.e. \( \varphi(t) = e^{2\pi i \omega_0 t} \). Now, assume that we sub-sample this phasor to obtain

\[ \varphi_d(n) = e^{2\pi i \omega_0 (nT)} \]

i.e., \( T \) is the sampling interval. We have seen many times by now, that adding \( 2\pi i nk \) to exponent doesn’t change this (discrete) function:

\[ \varphi_d(n) = e^{2\pi i \omega_0 (nT) + 2\pi i nk} = e^{2\pi i T n (\omega_0 + k/T)} \text{, for all } k \in \mathbb{Z}. \]

This equation tells us that, after sampling, a sinusoid with frequency \( \omega_0 \) cannot be distinguished from a sinusoid with frequency \( \omega_0 + k/T \), \( k \in \mathbb{Z} \). Note that \( F_s = 1/T \) is the sampling rate.

If we sample real-valued signals, however, we always have to consider positive and negative frequencies, so, to a real sinusoid (sine or cosine) with frequency \( \omega_0 \), we have in fact aliases at \( \pm \omega_0 k/T \).

**Example 16.** Recall now the Fourier series of a square wave as defined in Example 2:

\[ f(x) = \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{1}{(2k-1)} \sin(2\pi (2k-1)x) \]

Obviously, this periodic function does NOT have a finite number of frequencies in it: its spectrum, i.e. the frequencies contained in the square wave decay like \( 1/n \) - and this really slow! We note that, due to this infinite bandwidth, the square-wave cannot be sampled properly: sampling, no matter how densely must always lead to aliasing, as we shall see next.

We assume a sampling rate of \( F_s = 44100\text{Hz} \) and consider a square wave with fundamental frequency \( F = 700\text{Hz} \). Then, the Fourier series of this function is simply

\[ f(x) = \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{1}{(2k-1)} \sin(2\pi * 700 * (2k-1)x) \]

since now we have 700 oscillations per second. In this case, the highest frequency that is still below the Nyquist frequency of 22050Hz is the 31st harmonic which belongs to the frequency 31 \cdot 700 = 21700. The next frequency contained in the signal, with index 33 is 23100Hz and is above Nyquist. It will therefore show up as an alias at \((23100 - 44100)\text{Hz} = -21000\text{Hz}. \) This effect continues for all higher frequencies, and of course, all the negative frequencies turn into positive aliases accordingly. The phenomenon is shown on Figure 2.4.
Figure 2.3: Aliasing by subsampling
Figure 2.4: Aliasing by discretization of Square Wave
Not that in the current case, the fundamental frequency divides the Sampling rate, and the aliases become quasi-harmonics. In contrast, if we choose \( F = 800 \text{Hz} \), the aliases will occur in frequencies that are not related not the fundamental frequencies, see Figure 2.5.

While aliases should be avoided in usual sampling procedure, intentional aliasing can lead to interesting sound effects.

2.0.8 The mathematics of aliasing: sampling is periodization in the Fourier domain

In Section 1.2.5, we looked at the finite DFT of a Dirac comb and noted that it exhibits periodicity. Now, we will prove the continuous analog of this observation: if a continuous function is sampled, its Fourier transform is periodic.

**Theorem 1.** The Fourier transform of the discrete signal \( f_d \) obtained by sampling a continuous signal \( f \) at a sampling interval \( T \) is

\[
\hat{f_d}(\omega) = \frac{1}{T} \sum_{k \in \mathbb{Z}} \hat{f}(\omega - \frac{k}{T}).
\]

**Proof.** First note that \( f_d[n] = f(nT)\delta(t - nT) = f(t)\delta(t - nT) \), since the delta has no non-zero values outside \( 0 \). In other words:

\[
f_d[n] = f(t) \sum_{n \in \mathbb{Z}} \delta(t - nT) = f(t) \cdot \Pi_T(t).
\]

Now we invoke the convolution theorem for the Fourier transform: if \( h(t) = f(t) \cdot g(t) \), then \( \hat{h}(\omega) = [\hat{f} \ast \hat{g}](\omega) \), hence

\[
\hat{f_d}(\omega) = (\hat{f} \ast \Pi_T)(\omega).
\]

Thus we need to compute \( \Pi_T \) in a next step. First observe:

\[
\Pi_T(\omega) = \int_t \sum_{n \in \mathbb{Z}} \delta(t - nT)e^{-2\pi i t\omega} dt
\]

\[
= \sum_{n \in \mathbb{Z}} \int_t \delta(t - nT)e^{-2\pi i t\omega} dt = \sum_{n \in \mathbb{Z}} e^{-2\pi i nT\omega}
\]

We observe that the last expression is a sum of \( \frac{1}{T} \)-periodic functions, hence is \( \frac{1}{T} \)-periodic. Motivated by our observations in the finite discrete we conjecture, that

\[
\sum_{n \in \mathbb{Z}} e^{-2\pi i nT\omega} = \frac{1}{T} \sum_{k \in \mathbb{Z}} \delta(\omega - \frac{k}{T}).
\]
Figure 2.5: Aliasing by discretization of Square Wave
Since the sum \( \sum_{k \in \mathbb{Z}} \delta(\omega - \frac{k}{T}) \) is \( \frac{1}{T} \)-periodic, it can be written as \( \sum_n c_n e^{-2\pi i n T \omega} \) and we can compute its Fourier coefficients \( c_n \) by:

\[
c_n = T \int_{-1/2T}^{1/2T} \sum_{k \in \mathbb{Z}} \delta(\omega - \frac{k}{T}) e^{-2\pi i n T \omega} d\omega
= T \int_{-1/2T}^{1/2T} \delta(\omega) e^{-2\pi i n T \omega} d\omega = T \cdot 1 = T.
\]

Therefore \( \sum_{k \in \mathbb{Z}} \delta(\omega - \frac{k}{T}) = T \sum_n e^{-2\pi i n T \omega} \), hence

\[\widehat{\Pi_T}(\omega) = \frac{1}{T} \delta(\omega - \frac{k}{T})\]

and

\[\hat{f}_d(\omega) = (\hat{f} \ast \widehat{\Pi_T})(\omega) = \frac{1}{T} \sum_{k \in \mathbb{Z}} \hat{f}(\omega - \frac{k}{T}).\]

Let us now recollect, what we have observed so far: if we are given a continuous signal \( f \), and we sample it with a sampling rate of \( 1/T \) samples per second: \( f_d[n] = f(Tn) \). Then, the corresponding Fourier transform \( \hat{f}_d \) is a periodized version of the original \( \hat{f} \), with period equal to the sampling rate, i.e. \( 1/T \). Now, under the assumption, that the original signal was effectively bandlimited to the interval \([-\frac{1}{2T}, \frac{1}{2T}]\) it is now quite obvious what needs to be done to recover the continuous waveform: we have to cut off all the unnecessary copies of the spectrum.

### 2.0.9 Reconstruction from samples

We will now see that the elimination of all frequencies above \( \frac{1}{2T} \) leads to interpolation.

**Theorem 2** (Shannon Sampling Theorem). *Assume that the Fourier transform of a continuous signal \( f \) is contained in the interval \([-\frac{1}{2T}, \frac{1}{2T}]\), then \( f \) can be perfectly reconstructed from its samples at \( nT \), i.e. from \( f_d = f \ast \Pi_T \) as follows*

\[f(t) = \sum_{n \in \mathbb{Z}} f(nT) \operatorname{sinc}(\frac{t-nT}{T}) \quad (2.2)\]

**Proof.** Note that for \( n \neq 0 \) the support of \( \hat{f}(\omega - \frac{n}{T}) \) does not intersect with the support of \( \hat{f}(\omega) \), since \( \hat{f}(\omega) = 0 \) for \( |\omega| \geq \frac{1}{2T} \). Therefore

\[\hat{f}_d(\omega) = \frac{1}{T} \hat{f}(\omega) \text{ for } |\omega| \leq \frac{1}{2T}\]

We then have \( \hat{f}(\omega) = [T \cdot \Pi_T \cdot \hat{f}_d](\omega) \) and thus

\[f(t) = \mathcal{F}^{-1}(T \cdot \Pi_T \cdot \hat{f}_d)(t)\]
2 How does the Music end up on a CD? Sampling and Filtering

and \( \mathcal{F}^{-1} \Pi_T = \hat{\Pi}_T \), since \( \Pi_T \) is symmetric, hence

\[
f(t) = (T\hat{\Pi}_T * \hat{f}_d)(t) = T\hat{\Pi}_T \sum_{n \in \mathbb{Z}} f(nT)\delta(t - nT) \\
= \sum_{n \in \mathbb{Z}} f(nT)T\hat{\Pi}_T(t - nT)
\]

Recall that \( \hat{\Pi}(x) = \text{sinc}(x) \) and, since \( \hat{\Pi}_T(x) = \hat{D}_T \hat{\Pi}(x) = \frac{1}{T} D_T \hat{\Pi}(x) \), we find

\[
f(t) = \sum_{n \in \mathbb{Z}} f(nT)\text{sinc}(\frac{t}{T} - n).
\]

\( \square \)

2.0.10 Prefiltering of analog signals

Anti-alias Filter: The pre-filtering of an analog signal, before it is digitized or sampled, to remove or substantially attenuate the undesired aliasing components, i.e. those components that have higher frequency than half the sampling rate.
3 Digging the audio structure: Frames and Redundancy

3.1 Introduction - Uncertainty principle and time-frequency molecules

So far, we have been looking at a signal as being described either in time or in frequency - with the Fourier transform as a means of transformation from one domain to the other. Many signals of interest, however - above all: music and speech - are not stationary over time, they are time-variant, and we will be more interested in finding the frequency-content at a particular point in time, rather than knowing which frequencies comprise the entire signal. With this question we enter the important realm of time-frequency analysis (TF-Analysis). The desire to have complete insight in the local time-frequency structure, however, is impaired by a central fact of both TF-analysis and - more famously - quantum mechanics, called uncertainty principle. Loosely speaking, it states that, the more concentrated \( f(t) \) is, the wider its Fourier transform \( \hat{f}(\omega) \) must be. In other words, the scaling property of the Fourier transform may be seen as saying: if we "squeeze" a function in \( t \) (time), its Fourier transform "stretches out" in \( \omega \) (frequency). It is not possible to arbitrarily concentrate both a function and its Fourier transform.

**Example 17 (Gaussian Window).** The Fourier transform of the normalized Gaussian window \( \phi_0(t) = e^{-\pi t^2} \) is given by \( \hat{\phi}_0(\omega) = e^{-\pi \omega^2} \), in other words: the Gaussian is invariant under Fourier transform. Then, the dilated Gaussian \( \phi_a(t) = e^{-\pi t^2/a} \) has the Fourier transform \( \hat{\phi}_a(\omega) = \sqrt{a} e^{-\pi a \omega^2} \).

Suppose \( f \) is a piece-wise continuous, square-integrable function. Without loss of generality, assume that \( f \) is normalized:

\[
\int_{-\infty}^{\infty} |f(t)|^2 \, dt = 1.
\]

It follows from the Plancherel theorem that \( \hat{f}(\omega) \) is also normalized.

The standard deviation around \( t = 0 \) may be measured by

\[
\sigma_t(f) = \int_{-\infty}^{\infty} t^2 |f(t)|^2 \, dt.
\]

Similarly:

\[
\sigma_\omega(f) = \int_{-\infty}^{\infty} \omega^2 |\hat{f}(\omega)|^2 \, d\omega.
\]
The Uncertainty principle states that

\[
(\sigma_t(f) \cdot \sigma_\omega(f))^{1/2} \geq \frac{1}{4\pi}
\]

and equality holds only if \( f(t) = C\varphi_\omega(t) \) where \( C \) is an arbitrary constant. The above inequality implies that

\[
\left( \int_{-\infty}^{\infty} (t - t_0)^2 |f(t)|^2 \, dx \right)^{1/2} \left( \int_{-\infty}^{\infty} (\omega - \omega_0)^2 |\hat{f}(\omega)|^2 \, d\omega \right)^{1/2} \geq \frac{1}{4\pi}
\]

for all \( t_0, \omega_0 \in \mathbb{R} \).

In quantum mechanics, time and frequency are replaced by momentum and position and the inequality above is the famous statement of the Heisenberg uncertainty principle.

### 3.2 Approximation

#### 3.2.1 Least squares method

The method of least squares is a standard approach to the approximate solution of overdetermined systems, i.e., sets of equations in which there are more equations than unknowns. "Least squares" means that the overall solution minimizes the sum of the squares of the errors made in solving every single equation.

We first recall the following system of linear equations:

\[
Ax = b,
\]

where \( A \) is a \( k \times n \) matrix, \( x \in \mathbb{R}^n \) and \( b \in \mathbb{R}^k \). \( A \) is the matrix of a linear map from \( \mathbb{R}^n \) to \( \mathbb{R}^k \) with respect to some basis. The system of linear equations (3.1) can be solved with a unique solution, if and only if \( k = n \) and \( A \) is invertible. In this case, the rank of the matrix is \( n \) and so is the dimension of its range (column space \( C(A) \)), while the dimension of its kernel (null-space \( N(A) \)) is 0.\(^1\) This is the almost trivial, and in this section, we will be more interested in the cases when \( k \neq n \) and in particular, when \( k > n \), in which case the problem (3.1) is over-determined and we will usually only be able to find approximate solutions. Talking in the context of the subspaces mentioned so far: the range of our matrix \( A \) is a proper subspace of \( \mathbb{R}^k \).

On the other hand, if \( k < n \), we have less equations than parameters and usually (3.1) is thus under determined. In this case, the rank of the matrix, and hence the dimension of its range, is at most \( k \), so that there must be a non-trivial null-space \( N(A) \), i.e. there are \( y \neq 0 \in \mathbb{R}^n \), such that \( Ay = 0 \). Assume now, that \( x \) solves (3.1), then, for any scalar \( c \) and \( x + cy \) we have

\[
A(x + cy) = Ax + cAy = Ax = b
\]

3.2 Approximation

and hence there are always infinitely many solutions. Use Figure 3.1 to get a good orientation in the four subspaces involved in linear maps\(^2\).

**Projection onto a one-dimensional subspace**

Let us assume that we want to project a vector \(b \in \mathbb{R}^n\) onto a vector \(a \in \mathbb{R}^n\). The corresponding \(n \times n\)-matrix \(P_a\) is then given by \(a \cdot a^T\), since

\[
Pb = a \cdot \frac{\langle b, a, \rangle}{\|a\|^2} = a \cdot \frac{a^T \cdot b}{\|a\|^2}.
\]

**Example 18.** We want to project the vector

\[
b = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} + \begin{pmatrix} 2 \\ -2 \\ 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 2 \\ -3 \end{pmatrix}
\]

onto (a) \(a_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\), (b) \(a_2 = \begin{pmatrix} 1 \\ i \end{pmatrix}\).

In the first case, \(P_{a_1} = \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}\) and the resulting projection is the 0-vector.

This is the obvious result, if we try to approximate a vector (here \(b\)) by a vector that is orthogonal (perpendicular) to \(b\). On the other hand, \(P_{a_2} = \frac{1}{4} \begin{pmatrix} 1 & -1 & -1 & i \\ -1 & 1 & i & -1 \\ -i & 1 & 1 & i \\ i & -1 & -i & 1 \end{pmatrix}\) and

\(^2\)In Figure 3.1, \(m\) is used instead of \(k\), since this plot is taken from a book of G.Strang.


3 Digging the audio structure: Frames and Redundancy

\[ P_{u_2} \cdot b = \begin{pmatrix} -i \\ -1 \\ i \\ 1 \end{pmatrix}. \]

Note that the projection onto a vector \( b \) (a one-dimensional subspace) is equivalent to the approximation by \( b \). We next generalize this to the approximation by several vectors.

**Projection onto a subspace of higher dimension**

We now assume that we are given \( n \) vectors \( a_j \in \mathbb{R}^k \), and we consider the projection of a vector \( b \in \mathbb{R}^k \) onto the subspace \( \mathcal{A} = \text{span}(a_1, \ldots, a_n) \), in other words, we want to approximate \( b \) by an arbitrary linear combination of the given vectors. Since we assume \( k > n \), we can also assume that the \( n \) vectors are linearly independent. If \( k \) is significantly larger than \( n \) we cannot expect to find an exact solution, so we will try to minimize the following error:

\[ e(\hat{x}) = \|\hat{x}_1 a_1 + \ldots + \hat{x}_n a_n - b\|_2^2, \]

so we are looking for the coefficient vector \( \hat{x} \in \mathbb{R}^n \) such that \( e(\hat{x}) = \min_{\hat{x} \in \mathbb{R}^n} \|\hat{x}_1 a_1 + \ldots + \hat{x}_n a_n - b\|_2^2 \), or \( \hat{x} = \text{argmin} e(x) \).

Since the error \( e = b - A\hat{x} \), where \( A \) is the \( k \times n \) matrix with the \( a_j \) as \( n \) columns, must be orthogonal to \( a_1, \ldots, a_n \), we obtain the new set of linear equations

\[ A^T (b - A\hat{x}) = 0 \iff A^T A\hat{x} = A^T b \]

and since the vectors \( a_j \) where supposed to be linearly independent, the \( n \times n \) matrix \( A^T A \) is invertible.

Hence, the coefficients of the best approximation are given by

\[ \hat{x} = (A^T A)^{-1} A^T b \]

and the best approximation, or orthogonal projection onto \( \mathcal{A} \) is then

\[ p = P_{\mathcal{A}} b = A \cdot (A^T A)^{-1} A^T b. \]

**Example 19.** Determine the projection of \( b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} \) onto \( \text{span} \left( \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right) \). Result:

\[ \hat{x} = (5, -3), \quad p = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}. \]

**Proposition 8.** A matrix \( A^T A \) is invertible if and only if the columns of \( A \) are linearly independent.
3.2 Approximation

Proof. We show that \( \ker(A^T A) = \ker(A) \), which is equivalent to the statement of the theorem (argue, why!)

(I) Let \( x \in \ker(A) \), then \( Ax = 0 \), hence, by linearity, \( A^T Ax = A^T 0 = 0 \).

(II) Let \( x \in \ker(A^T A) \), i.e. \( A^T Ax = 0 \), then \( x^T A^T Ax = 0 \), hence \( (Ax)^T (Ax) = ||Ax||^2 = 0 \) and therefore \( Ax = 0 \) and \( x \) is also in the kernel of \( A \).

\( \square \)

Corollary 2. Let \( A \) be an \( k \times n \) matrix. If \( n > k \), then \( A^T A \) cannot be invertible.

This follows immediately from the proposition above. Argue, why!

An important application: data fitting with polynomials

In engineering and many other applications, it is often necessary to fit a line to a set of data. A line is a first degree polynomial:

\[ s = ct + d \]

in other words, a line with slope \( c \). Our goal is to determine the coefficients \( c \) and \( d \) of the polynomial that lead to the "best fit" of a line to the data. Now assume that we are given data \((t_i, s_i), i = 1, \ldots, k\), i.e. \( k \) measurements \( s_i \) at time points \( t_i \). We then want to minimize the error

\[ e(c, d) = \sum_{i=1}^{k} |s_i - (ct_i + d)|^2, \]

and, setting

\[ \hat{x} = \begin{pmatrix} d \\ c \end{pmatrix}; \quad A = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_k \end{pmatrix}; \quad s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{pmatrix} \]

we can obtain the solution by solving the overdetermined linear system \( A\hat{x} = s \).

Why is the best solution to this system of linear equations also the minimizer of \( e(c, d) \)?

Example 20. Find the line that best approximates the three measurements \((0, 6); (1, 0); (2, 0)\).

Result: \( s = -3t + 5 \).

The fitting process can be generalized to determine the coefficients of the \( N \)th-order polynomial that best fits \( N + 1 \) or more data points. The determination of the coefficients can be done in MATLAB by the function polyfit.

For example, for \( N = 3 \):

\[ s = c_0 + c_1 t + c_2 t^2 \]

This will exactly fit a simple curve to three points, as we see in the following example.

Example 21. Fit a polynomial of degree 2 to the data points from Example 20.

Now we use

\[ \hat{x} = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix}; \quad A = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \end{pmatrix}; \quad s = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} \]
and obtain the result \( s = 6 - 9t + 3t^2 \), i.e., the coefficient-vector is given by \( \hat{x} = \begin{pmatrix} 6 \\ -9 \\ 3 \end{pmatrix} \).

In this case, the "projection"-matrix onto the range of \( A \ (CS(A), \ SR(A)) \) is the identity, since the columns of \( A \) span all of \( \mathbb{R}^3 \).

Remark 7. In the lecture we discussed that there are three ways to interpret the method of least squares: geometrically (looking for the point in the hyperplane spanned by the columns of \( A \), that is closest to \( b \)), algebraically, by removing the part of \( b \), that is orthogonal to the range of \( A \) (i.e. to all the columns of \( A \)), and solve \( Ax = p \) instead of \( Ax = b = p + e \). The removed part \( e \) is the inevitable error due to the non-empty null-space of \( A^T \). Lastly, there is the analytic interpretation obtained by taking partial derivatives with respect to the unknown coefficients in \( \hat{x} \) and deriving the minimizer thereof.

### 3.2.2 Eigenvalues and singular values

We first recall some facts about matrices and diagonalization.

A square matrix \( A \) is called diagonalizable if it is similar to a diagonal matrix, i.e., if there exists an invertible matrix \( P \) such that \( P^{-1}AP \) is a diagonal matrix.

The finite-dimensional spectral theorem says that any symmetric matrix whose entries are real can be diagonalized by an *orthogonal matrix*. More explicitly: For every symmetric real matrix \( A \) there exists a real orthogonal matrix \( Q \) such that \( D = Q^T AQ \) is a diagonal matrix.

Here, \( Q' \) is the transpose of \( Q \) for real matrices and the conjugate transpose (Hermitian transpose or adjoint matrix) for a complex matrix: \( A' = (A)^T = A^\dagger \). Every real symmetric matrix is Hermitian, and therefore all its eigenvalues are real. As a consequence, since \( Q^{-1} = Q' \) for unitary (or orthogonal) matrices\(^3\), inversion of symmetric real matrices is straightforward once its decomposition \( QDQ' = A \) is known: \( A^{-1} = (QDQ')^{-1} = QD^{-1}Q' \).

Example 22. The matrix \( M_C = \begin{pmatrix} 0.1 & 0 & 0.1 & 1 \\ 1 & 0.1 & 0 & 0.1 \\ 0.1 & 1 & 0.1 & 0 \\ 0 & 0.1 & 1 & 0.1 \end{pmatrix} \) is the matrix corresponding to finite discrete convolution with the vector \( k_C = (0.1, 1, 0.1, 0) \). Check that \( M_C \cdot \delta = k_C \).

The eigenvectors of this matrix are therefore given by the vectors

\n
\[ s_l[n] = e^{2\pi iln\pi}, \quad l = 0, \ldots, 3. \]

We can easily compute the eigenvalues - and therefore the inverse of \( M_C \): since the eigenvectors \( s_l \) constitute exactly the matrix of the finite discrete Fourier transform, we have, for any vector \( v \in \mathbb{R}^4 \):

\[ M_Cv = \mathcal{F}D\mathcal{F}'v \]

\(^3\)A unitary matrix in which all entries are real is an orthogonal matrix.
3.2 Approximation

and this is, once more, the convolution relation for the Fourier transform: instead of convolving two vectors, we can take their Fourier transforms and apply pointwise multiplication. The action of pointwise multiplication written by means of a matrix is the multiplication with a diagonal matrix:

\[ \mathcal{F}'(M_C v) = D (\mathcal{F}' v). \]

The entries of the diagonal matrix \( D \) are, of course, the eigenvalues of the convolution, and are therefore given by the Fourier transform of the convolution kernel \( k_C \):

\[ \mathcal{F}'(v * k_C) = \hat{v} : \hat{k}_C = D (\mathcal{F}' v). \]

In our concrete example, we have \( \hat{k}_C = (1.2, -i, -0.8, i) \) and therefore

\[
M_C = \mathcal{F} D \mathcal{F}' = \begin{pmatrix}
0.5 & 0.5 & 0.5 & 0.5 \\
0.5 & 0.5i & -0.5 & -0.5i \\
0.5 & -0.5 & 0.5 & -0.5 \\
0.5 & -0.5i & -0.5 & 0.5i \\
\end{pmatrix} \cdot \begin{pmatrix}
1.2 & 0 & 0 & 0 \\
0 & -i & 0 & 0 \\
0 & 0 & -0.8 & 0 \\
0 & 0 & 0 & i \\
\end{pmatrix} \cdot \begin{pmatrix}
0.5 & 0.5 & 0.5 & 0.5 \\
0.5 & -0.5i & -0.5 & 0.5i \\
0.5 & -0.5 & 0.5 & -0.5 \\
0.5 & 0.5i & -0.5 & -0.5i \\
\end{pmatrix}.
\]

The inverse of \( M_C \) is then

\[
M_C^{-1} = \mathcal{F} D^{-1} \mathcal{F}' = \begin{pmatrix}
0.5 & 0.5 & 0.5 & 0.5 \\
0.5 & 0.5i & -0.5 & -0.5i \\
0.5 & -0.5 & 0.5 & -0.5 \\
0.5 & -0.5i & -0.5 & 0.5i \\
\end{pmatrix} \cdot \begin{pmatrix}
5/6 & 0 & 0 & 0 \\
0 & i & 0 & 0 \\
0 & 0 & -5/4 & 0 \\
0 & 0 & 0 & -i \\
\end{pmatrix} \cdot \begin{pmatrix}
0.5 & 0.5 & 0.5 & 0.5 \\
0.5 & -0.5i & -0.5 & 0.5i \\
0.5 & -0.5 & 0.5 & -0.5 \\
0.5 & 0.5i & -0.5 & -0.5i \\
\end{pmatrix}.
\]

Not all square matrices are invertible, let alone any rectangular matrices. We saw in the section on least squares approximation that we might still be interested in an inversion of the action of \( A \), "where it is possible", in other words, on the range of \( A \). The pseudoinverse does exactly that: it inverts the action of \( A \) mapping the row-space to the column-space. Before we look at this new inversion, we have to introduce a generalization of the eigenvalues, which are the singular values. Again, any matrix has a singular value decomposition (SVD)!

**Singular value decomposition**

The main feature of diagonalization of symmetric matrices is the fact, that the action of \( A \) can be written as a diagonal matrix by means of a change of basis. As we saw above, even inversion is then easily realized.

We are seeking a similar representation for all matrices, in particular for \( k \times n \)-matrices, when \( n \neq k \). In the general setting, however, we will have to work with two ONBs: one for the domain space (\( \mathbb{R}^n \)), one for the range space (\( \mathbb{R}^k \)).

The SVD is a factorization of a real or complex \( k \times n \) matrix \( A \) with rank \( r \) of the form

\[ A = U \Sigma V' \]

where \( U \) is a \( k \times k \) unitary matrix , \( \Sigma \) is a \( k \times n \) rectangular diagonal matrix with nonnegative real numbers on the diagonal, and \( V \) is an \( n \times n \) unitary matrix and \( V' \) denotes the complex transposition (adjoint) of \( A \), or just transposition in the real
The diagonal entries $\Sigma$ are the singular values of $A$. The columns of $U$ are the left singular vectors and form an ONB of $\mathbb{R}^k$ and the columns of $V$ are the right singular vectors of $A$ and form an ONB of $\mathbb{R}^n$. The SVD has many useful applications in signal processing and statistics.

The singular value decomposition and the eigen-decomposition are closely related, since

- the left singular vectors of $A$ are eigenvectors of $AA'$,
- the right singular vectors of $A$ are eigenvectors of $A'A$,
- the non-zero singular values of $A$ are the square roots of the non-zero eigenvalues of $AA'$ or $A'A$.

**Theorem 3 (SVD).** Let $A : \mathbb{R}^n \to \mathbb{R}^k$ with rank $r$. There exist an ONB $V = \{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ and an ONB $U = \{u_1, \ldots, u_k\}$ of $\mathbb{R}^k$ such that

$$Av_i = s_iu_i; A'u_i = s_iv_i \text{ and } A'Av_i = s_i^2u_i, \text{ with } s_i > 0 \text{ for } i \leq r.$$  

**Proof.** (1) $A'A$ is symmetric and can therefore by diagonalized by an ONB $V = \{v_1, \ldots, v_n\}$. Recall that $\text{rank}(A'A) = \text{rank}(A) = r$, hence, we may order the eigenvalues such that

$$\lambda_1 \geq \ldots \geq \lambda_r > = 0 = \lambda_{r+1} \ldots \lambda_n$$

and the vectors $v_1, \ldots, v_r$ form an ONB for the range of $A'$ (row-space), while $v_{r+1}, \ldots, v_n$ form an ONB for the kernel of $A$.

(2) Set $s_i = \sqrt{\lambda_i}$ and $u_i = \frac{1}{\sqrt{\lambda_i}}Av_i$ for $i = 1, \ldots, r$. Then, the $\{u_i, i = 1, \ldots, r\}$ span the range of $A$ and they form an ONB, since

$$\langle u_i, u_j \rangle = \frac{1}{\sqrt{\lambda_i}} \frac{1}{\sqrt{\lambda_j}} \langle Av_i, Av_j \rangle = \frac{1}{\sqrt{\lambda_i \lambda_j}} \langle A'Av_i, v_j \rangle = \frac{\lambda_i}{\sqrt{\lambda_i \lambda_j}} \langle v_i, v_j \rangle = \frac{1}{\sqrt{\lambda_j}} \langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else,} \end{cases}$$

since $V$ is an ONB. $\{u_i, i = 1, \ldots, r\}$ is an ONB for the range (column space) of $A$ and, since the kernel of $A'$ is orthogonal to the range of $A$, can be extended to an ONB $U$ of all $\mathbb{R}^k$ by adding an ONB of the kernel of $A'$. \qed

**Corollary 3 (SVD).** Every $k \times n$ matrix $A$ with rank $r$ can be decomposed by $A = U\Sigma V'$, with $U, V$ unitary (orthogonal) $k \times k$ and $n \times n$ matrices, respectively, and

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix},$$

is a $k \times n$ matrix in which $D$ is an $r \times r$ diagonal matrix with the positive singular values $s_i, i = 1, \ldots, r$ of $A$ in the diagonal.
3.2 Approximation

Proof. This decomposition follows directly from the construction of Theorem 3, by letting $V$ be the unitary $n \times n$ matrix with the vectors of the ONB $V$ as its columns and $U$ the $k \times k$ matrix with the vectors of the ONB $U$ as its columns. Then

$$AV = (s_1 u_1 \cdots s_r u_r 0 \cdots 0) = U \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

and multiplication with $A^t$ from the right completes the proof. \qed

Example 23. The SVD of

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix}$$

is given by

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Example 24. The SVD of

$$A = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}$$

is given by

$$A = \begin{pmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & -2/\sqrt{5} \end{pmatrix} \begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{pmatrix}$$

Example 25. The matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix}$$

cannot be diagonalized by means of an orthonormal basis! However, we can find its SVD, given by:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{10} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1/\sqrt{5} & 2/\sqrt{5} \\ 1 & 0 & 0 \\ 0 & 2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix}$$

3.2.3 Pseudoinverse: a generalization of matrix inversion

The singular value decomposition can be used for computing the pseudoinverse (PINV) of a matrix. PINV is a generalized inverse and is constructed according to the following ideas:

Given a $k \times n$ matrix $A: \mathbb{R}^n \to \mathbb{R}^k$, we consider the injective mapping given by restricting $A$ to the orthogonal complement of its kernel $N(A)$, which is the row-space of $A$, or column space of $A^t: C(A^t)$: cp. Figure 3.1. So, we consider the injective mapping

$$\tilde{A} : C(A^t) \to \mathbb{R}^k.$$
Now, $A$ and $\tilde{A}$ have the same range, which is $C(A)$ and $\tilde{A}$ considered as a mapping $C(A') \to C(A)$ has an inverse:

$$\tilde{A}^{-1} : C(A) \to C(A'),$$

and we expect that, since $Av_i = s_iu_i$ for the members of the ONBs of $C(A')$ and $C(A)$, $i = 1, \ldots, r$, that

$$\tilde{A}^{-1}u_i = \frac{v_i}{s_i}.$$  

The mapping $\tilde{A}^{-1}$ can be extended to an operator $A^+ : \mathbb{R}^k \to \mathbb{R}^n$ by defining

$$A^+(u_1 + u_3) = \tilde{A}^{-1}(u_1) \text{ for } u_1 \in C(A), u_3 \in N(A') = C(A)^\perp.$$

In other words, the part of the vector $u = u_1 + u_3$ that is orthogonal to the range of $A$ and is thus in the kernel of $A'$, is set to 0 by the pseudoinverse. Have a look at Figure 3.2. It is immediately clear, that $A^+$ fulfills the desired property $AA^+u = u$ if $u \in C(A)$, i.e. the product is the identity on the range of $A$, compare this to the corresponding property of an invertible matrix!

Now we will see that the work we did in the previous section was not in vain! Indeed, the pseudoinverse of the matrix $A$ with singular value decomposition $A = UV\Sigma'$ is

$$A^+ = V\Sigma^+U',$$  

(3.2)

where $\Sigma^+$ is the pseudoinverse of $\Sigma$, and $\Sigma^+$ is formed by replacing every nonzero diagonal entry by its reciprocal and transposing the resulting matrix. In other words, using the notation of Corollary 3, we have

$$\Sigma^+ = \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$  

Example 26. The PINV of $A$ from Example 23 is given by

$$A^+ = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{5} & 0 \\ 0 & 1/\sqrt{5} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$  

Example 27. The PINV of $A$ from Example 24 is given by

$$A^+ = \begin{pmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & -2/\sqrt{5} \end{pmatrix} \begin{pmatrix} 1/\sqrt{10} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{pmatrix}.$$  

Example 28. The PINV of $A$ from Example 25 is given by:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1/\sqrt{5} & 0 & 2/\sqrt{5} \\ 2/\sqrt{5} & 0 & -1/\sqrt{5} \end{pmatrix} \cdot \begin{pmatrix} 1/\sqrt{10} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}.$$
3.2 Approximation

Pseudoinverse \[ A^+ v_i = \frac{v_i}{\sigma_i} \text{ for } i = 1, \ldots, r. \]

Figure 3.2: Orthonormal bases that diagonalize \( A \) (3 by 4) and \( A^+ \) (4 by 3), matrices with rank 2.
Proposition 9. Let \( A \) be a \( k \times n \) matrix and let \( A^+ = V \Sigma^+ U' \) be its PINV as defined in (3.2).

(i) \( A^+ \) maps the range of \( A \), i.e. \( C(A) \) onto the row space of \( A \), which is \( C(A') \). The kernel of \( A' \), i.e., the orthogonal complement of the range of \( A \) is mapped to 0.

(ii) \( A^+ \) is the unique \( n \times k \) matrix for which \( AA^+ \) is the orthogonal projection onto the range of \( A \) (\( C(A) \)) and \( A^+ A \) is the orthogonal projection onto the range of \( A' \) (\( C(A') \)).

Proof. (i) Recall that the columns of \( U \) are the members of an ONB \( \mathcal{U} \) for \( \mathbb{R}^k \), in which the first \( r \) vectors \( u_1, \ldots, u_r \) span the range of \( A \), \( C(A) \). Hence, the range of \( A \) consists of all linear combinations \( \sum_{j=1}^{r} c_j u_j = v \). Then

\[
U' \cdot v = \begin{pmatrix}
\langle \sum_{j=1}^{r} c_j u_j, u_1 \rangle \\
\langle \sum_{j=1}^{r} c_j u_j, u_2 \rangle \\
\vdots \\
\langle \sum_{j=1}^{r} c_j u_j, u_r \rangle \\
\langle \sum_{j=1}^{r} c_j u_j, u_{r+1} \rangle \\
\vdots \\
\langle \sum_{j=1}^{r} c_j u_j, u_k \rangle 
\end{pmatrix} = \begin{pmatrix}
c_1 \\
c_2 \\
\vdots \\
c_r \\
0 \\
0 \\
0
\end{pmatrix},
\]

by orthonormality of \( \mathcal{U} \). Consequently, the \( n \times 1 \)-vector \( \Sigma^+ \cdot U' \cdot v \) is given by

\[
\Sigma^+ \cdot U' \cdot v = \begin{pmatrix}
\frac{c_1}{\sigma_1} \\
\vdots \\
\frac{c_r}{\sigma_r} \\
0 \\
\vdots \\
0
\end{pmatrix},
\]

and therefore

\[
V \Sigma^+ \cdot U' \cdot v = \sum_{j=1}^{r} \frac{c_j}{\sigma_j} v_j,
\]

which is in the row space \( C(A') \) of \( A \), since \( v_1, \ldots, v_r \) span \( C(A') \). From the same derivation it is apparent that the kernel of \( A' \), for which \( u_{r+1}, \ldots, u_k \) form an ONB, i.e., the orthogonal complement of the range of \( A \), is mapped to 0.

(ii) To prove the second statement, first note that \( AA^+ \) is an orthogonal projection, since, by unitarity of \( V \)

\[
AA^+ = U \cdot \Sigma \cdot V' \cdot V \cdot \Sigma' \cdot U' = U \cdot \Sigma \cdot \Sigma' \cdot U'
\]

and

\[
\Sigma \cdot \Sigma' = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},
\]

\[44\]
such that \((AA^+)^2 = AA^+\) and also \((AA^+)' = AA^+\).

From \(AA^+ = U \cdot \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \cdot U'\) it is easy to see that, for \(\nu = \sum_{j=1}^{k} c_j u_j \in \mathbb{R}^k\), we have

\[ AA^+ \nu = \sum_{j=1}^{r} \langle \nu, u_j \rangle u_j, \]

because all coefficients corresponding to indices above \(r\) are set to zero by the diagonal operator \(\Sigma \cdot \Sigma'\). Thus, \(AA^+\) is the orthogonal projection onto the range of \(A\). The statement for \(A^+A\) is shown completely analogously.

\[ \nabla \]

### 3.2.4 Pseudoinverse and least squares

The pseudoinverse is one way to solve linear least squares problems.

**Proposition 10.** Let \(A\) be a \(k \times n\) matrix. Then \(x^+ = A^+b\) is a least squares solution of \(Ax = b\) and for any other least squares solution \(\hat{x}\), we have \(\|\hat{x}\|_2 \leq \|x^+\|_2\), so \(\hat{x}\) has minimal norm.

**Proof.** Recall that \(\hat{x}\) of a least-squares solution of \(Ax = b\) is and only if it solves \(A'Ax = A'b\). First note that \(x^+\) is a least squares solution by observing that \(Ax^+ - b = AA^+b - b\) and since \(AA^+b\) is the orthogonal projection of \(b\) onto the range of \(A\), \(AA^+b - b\) is in the orthogonal complement of the range, hence in the kernel of \(A\) and therefore \(A'(AA^+b - b) = 0\).

We still have to show that for any other least squares solution \(\hat{x}\) of \(Ax = b\), we have \(\|\hat{x}\| \leq \|x^+\|\). To see this, note that

\[ A'(A\hat{x} - b) = A'(Ax^+ - b) = 0 \Rightarrow A'A\hat{x} = A'Ax^+ \]

and thus \(\hat{x} - x^+ \in N(A')\) and also \(\hat{x} - x^+ \in N(A)\) (by Proposition 8). However, \(x^+ \perp N(A)\) and we can estimate:

\[
\|\hat{x}\|^2 = \|x^+ - x^+ + \hat{x}\|^2 \\
= \|x^+\|^2 + \|\hat{x} - x^+\|^2 + 2\text{Re} \langle x^+, \hat{x} - x^+ \rangle \\
= \|x^+\|^2 + \|\hat{x} - x^+\|^2 \geq \|x^+\|^2
\]

\[ \nabla \]

### 3.3 The method of frames

Just as the inversion of non-singular square matrices is not the end of the story of matrix inversion, bases, let alone orthogonal bases, are not the end of the story of the expansion of functions, or signals. In particular, when we are interested in the analysis of music signals, we face insurmountable difficulties if we stick with the idea of analyzing or signals by using (orthonormal) bases. In this section, we will encounter an important generalization of bases, the so-called concept of frames. We will motivate this new idea with the most important analysis method used in audio signals processing.
3.3.1 Analysis of a time-variant signal: Short-time Fourier transform

The short-time Fourier transform (STFT) transform is used to determine the frequency content of local sections of a signal that changes over time. The function \( f \) of interest is multiplied by a window function which is nonzero for only a short period of time and the Fourier transform of the resulting, localized signal is computed. As the window is translated along the time axis, a two-dimensional representation of the signal is.

Mathematically, we define:

\[
\text{STFT} \{ f \} = S_\varphi f(\tau, \omega) = \int_{-\infty}^{\infty} f(t) \varphi(t-\tau) e^{-2\pi i \omega t} dt
\]

where \( \varphi \) is the window function and \( f \) is the signal to be analysed. \( S_\varphi f \) is essentially the Fourier Transform of \( f \cdot \varphi \), a complex function representing the phase and magnitude of the signal over time and frequency.

**Instantaneous Frequencies and some examples**

To gain some intuitive understanding of the kind of signal representation the STFT provides, let us consider the concept of "instantaneous frequency", which may be defined for signals with slowly varying frequency. Assume that a signal is given by \( f(t) = \sin(\theta(t)) \), with \( \theta(t) \) smooth, so that it can be approximated by its Taylor polynomial in the vicinity of time \( t \):

\[
\theta(t + \tau) \approx \theta(t) + \theta'(t) \cdot \tau,
\]

then, for small \( \tau \)

\[
\sin(\theta(t + \tau)) \approx \sin(\theta(t) + 2\pi \cdot \frac{\theta'(t)}{2\pi} \cdot \tau),
\]

and \( \nu_{\text{loc}} = \frac{\theta'(t)}{2\pi} \) can be interpreted as "local frequency" at time \( t + \tau \).

**Example 29.** Consider the three signals

\[
\begin{align*}
  f_1 &= \sin(2\pi \cdot 500 \cdot t) \\
  f_2 &= \sin(2\pi \cdot (500 \cdot t + (50/\pi) \sin 2\pi t)) \\
  f_3 &= \sin(2\pi \cdot (500 \cdot t + (125/2)t^2)),
\end{align*}
\]

then \( \nu_{\text{loc}} = 500Hz \), \( \nu_{\text{loc}}^2 = (500 + 100 \cos 2\pi t)Hz \) and \( \nu_{\text{loc}}^3 = (500 + 125t)Hz \). Compare these findings with the spectrograms shown in Figure 3.3.

**Motivation: the deficiencies of bases in audio analysis**

It is quite obvious that a reasonable analysis of time-variant signals such as music (or speech) requires a transformation that is local in both time and frequency. In other words, it doesn’t help a lot to have the frequency information for an entire music piece: we would like to know which frequency sounds at which time. However, why can’t we just cut our - probably sampled and thus discrete - signal into pieces of finite length and
3.3 The method of frames

Figure 3.3: Spectrograms of slowly time-variant signals. The instantaneous frequency, as computed in Example 29 is visible.
3 Digging the audio structure: Frames and Redundancy

Figure 3.4: A box function as a window and a Hanning window. The lower plots show the respective Fourier transforms.
Figure 3.5: Comparing the respective Fourier transforms of a simple signal and its windowed versions: once with a box function, once with a hanning window.
3 Digging the audio structure: Frames and Redundancy

Figure 3.6: Spectrograms resulting from the usage of different windows.
analyze each of them separately?
The answer to this question can be seen in Figure 3.4 and Figure 3.6 and is related to
the behavior of the Fourier transform of the box function. Indeed, "cutting the signal
into pieces" is equivalent to multiplying it with a number of shifted box functions, one
pieces would then be given by \( f_{\text{loc}} = f \cdot \Pi \), hence \( \hat{f}_{\text{loc}} = \hat{f} \ast \Pi = \hat{f} \ast \text{sinc} \). Then, each
piece can be easily recovered by means of the inverse Fourier transform and in fact, it
is even straightforward to see that we obtain an ONB if the shifted box-functions are
chosen accordingly. So far so good, but look at Figure 3.6, where the spectrogram (the
magnitude squared of the STFT) of such an analysis of a very simple signal is shown.
The problem with the cutting approach is the fact that sinc decays extremely slowly, and
that means that, as a result of \( f_{\text{loc}} = f \ast \text{sinc} \), the frequency information appears totally
smeared. Note that the signal shown is extremely simple, just two sinusoids multiplied
by an envelope, but even for this signal it would be hard to separate the two components.
On the other hand, have a look at the other proposed window, a smoother window, called
Hanning window, which is very popular in audio signal processing. This window's Fourier
transform has a much better decay, look at Figure 3.4, and as a result, the frequency
resolution in the STFT obtained by application of this window is a lot better. However,
since the Hanning window decays towards 0 smoothly, in order not to lose information,
we have to assure some overlap in the translation of the windows. In that way, we are
not able to obtain orthogonal bases.

Frames

A set \( \Phi = \{ \varphi_j \}_{j \in J} \) in a Hilbert space \( \mathcal{H} \) (Hilbert space!, i.e. a vector space with inner
product. For simplicity think of the euclidean spaces \( \mathbb{R}^n \) or \( \mathbb{C}^n \).) is complete if every
element in \( \mathcal{H} \) can be approximated arbitrarily well (in norm) by finite linear combinations
of elements in \( \Phi \). For a finite-dimensional vector space such as \( \mathbb{R}^n \) or \( \mathbb{C}^n \), this simply
means that any vector in \( \mathcal{H} \) can be written as a linear combination of the \( \{ \varphi_j \} \), in other
words, that \( \Phi \) spans \( \mathcal{H} \).

A complete set is overcomplete, if removal of one element of the set still results in a
complete system. In signal processing, overcompleteness can help to achieve a more
stable, more robust, or more compact decomposition than the usage of a basis which
implies uniqueness. Frames are an interesting generalization of bases and are widely
used in mathematics, computer science, engineering, and statistics.

**Definition 10** (Frames). Let \( \mathcal{H} \) be a Hilbert space (a finite-dimensional vector space with
inner product. A frame is defined to be a countable family of non-zero vectors \( \{ \varphi_j \}_{j \in J} \)
in \( \mathcal{H} \), such that for arbitrary \( f \in \mathcal{H} \),

\[
C_l \| f \|^2 \leq \sum_{j \in J} |\langle f, \varphi_j \rangle|^2 \leq C_u \| f \|^2
\]

where \( \langle \cdot, \cdot \rangle \) denotes the inner product, \( C_l \) and \( C_u \) are positive constants called lower and
upper frame bound, respectively. If \( C_l, C_u \) can be chosen such that \( C_l = C_u \), the frame is
called a tight frame.
3 Digging the audio structure: Frames and Redundancy

Note that the above inequality can be understood as an "approximate Plancherel formula" and in that sense, an ONB is a special case of a (tight) frame. Recall, that any ONB \( \{ \psi_j \}_{j \in J} \) implies a convenient signal representation for all \( f \in H \), given by

\[
f = \sum_{j \in J} \langle f, \psi_j \rangle \psi_j \quad \text{with} \quad \sum_{j \in J} |\langle f, \psi_j \rangle|^2 = \| f \|^2_2.
\]

For frames, the situation is, obviously, slightly more complicated. In general, we can not expect the mapping \( S: f \mapsto \sum_{j \in J} \langle f, \varphi_j \rangle \varphi_j \) to be identity, but at least, this mapping will be invertible, a property that opens the door to the virtues of frames.

**Definition 11 (Frame operator).** Consider a vector space \( \mathcal{V} \) and a family of elements \( \{ \varphi_j \}_{j \in J} \) in \( \mathcal{V} \).

\[
Sf = \sum_{j \in J} \langle f, \varphi_j \rangle \varphi_j \quad (3.3)
\]

Note that

\[
\langle Sf, f \rangle = \sum_{j \in J} \langle f, \varphi_j \rangle \varphi_j, f = \sum_{j \in J} |\langle f, \varphi_j \rangle|^2.
\]

Also note that the frame operator (3.3) can be written as the composition of an analysis operator \( C: \mathcal{V} \rightarrow \mathbb{C}^k \), given by \( C: v \rightarrow \{ \langle v, \varphi_j \rangle \}_{j \in J} \) and the synthesis operator \( D: \mathbb{C}^k \rightarrow \mathcal{V} \), given by \( D: c \rightarrow \sum_{j \in J} c_j \varphi_j \). In fact, \( D \) is the adjoint operator, i.e. the transposed, complex conjugate of \( C \): \( D = C^\dagger = C^T \) and thus \( S^\dagger = (DC)^\dagger = (C^T)^\dagger = C^*C = S \) and \( S \) is self-adjoint.

**Proposition 11.** Let \( \mathcal{V} \) be a finite-dimensional vector space and \( \Phi \) a frame for \( H \).

(i) \( S \) is invertible and self-adjoint.

(ii) There exist dual frames \( \tilde{\varphi}_k \), allowing an expansion of \( f \) as:

\[
f = \sum_j \langle f, \varphi_j \rangle \tilde{\varphi}_j = \sum_j \langle f, \tilde{\varphi}_j \rangle \varphi_j \quad (3.4)
\]

In particular, the canonical dual frame is given by \( \tilde{\varphi}_k = S^{-1} \varphi_k \).

(iii) If the frame is not a basis, then the coefficients \( c_j = \langle f, \varphi_j \rangle \) are not unique, but optimal in the sense of minimizing \( \sum_j |c_j|^2 \).

**Remark 8.** Completely analogous statements hold for infinite dimensional Hilbert spaces, but the proof is beyond our scope.

**Proof.** (i) We show that \( S \) is injective: assume that \( Sf = 0 \) for \( f \in \mathcal{V} \), then

\[
0 = \langle Sf, f \rangle = \sum_{j=1}^k |\langle f, \varphi_j \rangle|^2 \geq C_1 \|f\|^2 \Rightarrow \|f\| = 0 \Rightarrow f = 0.
\]
(ii) Since $S$ is invertible, we can write

$$ f = S^{-1} S f = S \left( \sum_j \langle f, \varphi_j \rangle \varphi_j \right) = \sum_j \langle f, \varphi_j \rangle S^{-1} \varphi_j $$

and setting $\tilde{\varphi}_j = S^{-1} \varphi_j$ proves the first equality. Changing order: $f = SS^{-1} f$ and noting that self-adjointness of $S$ leads to self-adjointness of $S^{-1}$, we obtain

$$ f = \sum_j \langle S^{-1} f, \varphi_j \rangle \varphi_j = \sum_j \langle f, S^{-1} \varphi_j \rangle \varphi_j. $$

(iii) The proof of this statement is similar to the proof of Proposition 10. In fact, suppose that the coefficient vector $c \in \mathbb{C}^k$ fulfills $f = \sum_{j=1}^k c_j \varphi_j$, hence, since $c_j = c_j - \langle f, S^{-1} \varphi_j \rangle + \langle f, S^{-1} \varphi_j \rangle$ and $f = \sum_{j=1}^k \langle f, S^{-1} \varphi_j \rangle \varphi_j$, we have

$$ \sum_{j=1}^k (c_j - \langle f, S^{-1} \varphi_j \rangle) \varphi_j = 0. $$

Setting $d_j = c_j - \langle f, S^{-1} \varphi_j \rangle$, this can be written as $Dd = 0$ and thus the sequence $d$, given by $d_j = c_j - \langle f, S^{-1} \varphi_j \rangle$ is in the kernel of $D = C'$, which is orthogonal to the range of $C$.

On the other hand, the sequence $\{ \langle f, S^{-1} \varphi_j \rangle \}_{j=1}^k$ is in the range of $C$, because

$$ \{ \langle f, S^{-1} \varphi_j \rangle \}_{j=1}^k = \{ S^{-1} f, \varphi_j \}_{j=1}^k = C(S^{-1} f). $$

Therefore, as in the proof of Proposition 10:

$$ \sum_{j=1}^k |c_j|^2 = \sum_{j=1}^k |c_j - \langle f, S^{-1} \varphi_j \rangle + \langle f, S^{-1} \varphi_j \rangle|^2 $$

$$ = \sum_{j=1}^k |c_j - \langle f, S^{-1} \varphi_j \rangle|^2 + \sum_{j=1}^k |\langle f, S^{-1} \varphi_j \rangle|^2 + $$

$$ + \sum_{j=1}^k \langle c_j - \langle f, S^{-1} \varphi_j \rangle, \langle f, S^{-1} \varphi_j \rangle \rangle, $$

and since the last sum is 0 due the orthogonality of the range of $C$ and the kernel of $D$, we have

$$ \sum_{j=1}^k |c_j|^2 = \sum_{j=1}^k |c_j - \langle f, S^{-1} \varphi_j \rangle|^2 + \sum_{j=1}^k |\langle f, S^{-1} \varphi_j \rangle|^2 \geq \sum_{j=1}^k |\langle f, S^{-1} \varphi_j \rangle|^2. $$

\qed
Frames in \( \mathbb{C}^n \), matrices and PINV

We will now look more closely at frames for the Euclidean spaces and investigate the links to linear algebra.

Let \( \{ \varphi_j \}_{j=1}^k \) be a frame for \( \mathbb{C}^n \). First observe, that the frame operator (3.3) can be written as the composition of the analysis operator \( C : \mathbb{C}^n \to \mathbb{C}^k \), given by \( C : v \to \{ \langle v, \varphi_j \rangle \}_{j=1}^k \) and the synthesis operator \( D : \mathbb{C}^k \to \mathbb{C}^n \), given by \( D : c \to \sum_{j=1}^k c_j \varphi_j \). In fact, \( D \) is the adjoint operator, i.e. the transposed, complex conjugate of \( C : D = C^\prime \).

We can now directly write out the matrices corresponding to these "operators" (linear maps). \( C \), the analysis operator, is then a \( k \times n \) matrix with \( \varphi_j \) in its \( j \)-th row. It follows immediately, from \( D = C^\prime \), that \( D \) is the \( n \times k \) matrix with the vector \( \varphi_j \) in its \( j \)-th column, and \( S = C^\prime \cdot C = D \cdot C \) is then a selfadjoint map, hence a symmetric matrix, since \( S^\prime = (DC) = (C^\prime C) = DC = S \).

**Proposition 12.** Let \( A \) be a \( k \times n \) matrix and \( \nu \in \mathbb{R}^n \) a vector given by its entries \( v_l, l = 1, \ldots, n \). Then the following are equivalent:

(i) There exists a constant \( C_u \):

\[
C_u \sum_{l=1}^n |v_l|^2 \leq \| Av \|_2^2, \quad \forall \nu \in \mathbb{C}^n.
\]

(ii) The columns of \( A \) form a basis for their span in \( \mathbb{C}^k \).

(iii) The rows of \( A \) form a frame for \( \mathbb{C}^n \).

**Proof.** Recall that the \( \ell^2 \)-norm of \( Av \) is given by \( \| Av \|_2^2 = \sum_{l=1}^k |(Av)_l|^2 \). Let \( \varphi_j[l] \), \( l = 1, \ldots, n \) denote the \( j \)-th column of \( A^\prime \), then the columns of \( A \) are given by the \( n \) vectors \( \psi_l = \begin{pmatrix} \varphi_1[l] \\ \varphi_2[l] \\ \vdots \\ \varphi_k[l] \end{pmatrix} \).

(i)\(\Rightarrow\)(ii): (3.5) means that \( C_u \sum_{l=1}^n |v_l|^2 \leq \| \sum_{l=1}^n v_l \psi_l \|_2^2 \forall \nu \in \mathbb{C}^n \). Now assume that the \( \psi_l, l = 1, \ldots, n \) do not form a basis for their linear span, which means that they are linearly dependent, i.e. there exists a non-zero vector \( \nu \in \mathbb{R}^n \), such that \( 0 = \sum_{l=1}^n v_l \psi_l \), which contradicts (3.5).

(i)\(\Rightarrow\)(iii): (3.5) can also be written as

\[
C_u \sum_{l=1}^n |v_l|^2 \leq \sum_{l=1}^n |\langle \nu, \varphi_l \rangle|^2, \quad \forall \nu \in \mathbb{C}^n,
\]

from which the frame property of the columns follows, since the upper frame bound is automatically satisfied. (iii)\(\Rightarrow\)(ii) holds by the definition of a frame.

**Example 30.** Consider the matrix \( A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \). Obviously, the rows constitute a frame for \( \mathbb{R}^3 \), while the columns span a two-dimensional subspace of \( \mathbb{R}^3 \).
**Proposition 13.** Let \( \{\varphi_j\}_{j=1}^k \) be a frame for \( \mathbb{C}^n \) with analysis operator \( C : \mathbb{C}^n \to \mathbb{C}^k \) synthesis operator \( D : \mathbb{C}^k \to \mathbb{C}^n \) and frame operator \( S \). Then

\[
D^+ v = \{ \langle v, S^{-1} \varphi_j \rangle \}_{j=1}^k, \quad \forall v \in \mathbb{C}^n.
\]

In other words, the canonical dual frame is given by the columns of the PINV of \( C \).

**Proof.** We know that \( v = \sum_{j=1}^k \langle v, S^{-1} \varphi_j \rangle \varphi_j \) and that \( v = \sum_{j=1}^k c_j \varphi_j = Dc \) for the vector \( c \in \mathbb{R}^k \). From Proposition 10 we have that \( c = D^+ v \) is the least-squares solution of this problem with the smallest norm and in Proposition 11(iii) it was shown that this solution is given by the coefficients \( \{ \langle v, S^{-1} \varphi_j \rangle \}_{j=1}^k = v \cdot C \cdot (C^* C)^{-1} \).

**Example 29 continued:** The pseudoinverse of \( A \) in Example 30 is given by

\[
A^+ = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix}
\]

Consider the frame consisting of the rows of \( A \), its frame operator is given by \( A^* A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \) and the dual frame consists of the columns of \( (A^* A)^{-1} \cdot A^* = A^+ \).

### Frames and stability

Assume that data are being transmitted using (a) a basis representation (b) a frame. That means, a sender wants to transmit \( f \in \mathbb{C}^n \) by using (a) an invertible \( n \times n \) matrix \( B \) and sends \( \hat{c}_b = B \cdot f \). Another sender uses an \( n \times k \) matrix \( A \) with full rank and transmits \( c_f = A \cdot f \). Now assume that the data are corrupted by some noise during transmission, i.e. \( \hat{c}_b = c_b + n_b \) and \( \hat{c}_f = c_f + n_f \), respectively. In case (a), the mapping \( B \) is unitary, so any noise added in the transmission will be completely picked up by the receiver. On the other hand, if the frame coefficients \( c_f \) are corrupted by noise and \( \hat{c}_b \) is received then there is justified hope, that an essential part of the noise is from the orthogonal complement of the range of \( A \), i.e., in \( N(A^*) \) and will, therefore, be set to 0 in the reconstruction.

#### 3.3.2 Gabor frames

Now we will introduce a class of frames that has become the most important one for audio signal processing and is tightly linked with the STFT. Look at the definition of the STFT: the integral can be interpreted as an inner product between \( f \) and the window function \( \varphi \) that is shifted in time and modulated (shifted in frequency).

**Definition 12 (Time-frequency shifts).** Let \( f \in \mathbb{C}^L \) and consider this vector extended to its \( L \)-periodic version by \( f(k + lL) = f(k) \) wherever necessary. \( T_k f(t) := f(t - k) \) is called translation operator or time shift. \( M_l f(t) := e^{2\pi i l f(t)}, l \in \mathbb{Z} \) is called modulation operator or frequency shift. The composition of these operators, \( M_l T_k \), is a time-frequency shift operator.
Generally, we are not interested in calculating the inner product in every point of the time-frequency lattice. This would yield a redundancy of $L$, the length of the given signal. We down-sample in time by $a$ and in frequency by $b$, so that the redundancy reduces to $\frac{L}{ab}$. $a$ and $b$ are referred to time- and frequency-shift parameters. The family

$$g_{m,n} := M_mbT_{na}g$$

for $m = 0, \ldots, M - 1$ and $n = 0, \ldots, N - 1$, where $Na = Mb = L$, is called the set of **Gabor analysis functions**.

Let us assume the $g_{m,n}$ were an orthogonal basis for a moment. In this case, the inner products $\langle f, g_{m,n} \rangle$ uniquely determine $f$, each representing a single and unique coefficient in the expansion

$$f = \sum_m \sum_n \langle f, g_{m,n} \rangle g_{m,n}$$

Together with Plancherel’s formula $\|f\|^2 = \sum_m \sum_n |\langle f, g_{m,n} \rangle|^2$ this gives a beautiful split of $f$ in pieces, preserving the signal’s energy in the coefficients.

Of course there is a problem. Theory tells us, that the members of a basis of the above form can never be well-localised in both time and frequency. Therefore we have to play a trade-off between nice properties of the representation on the one hand and satisfactory mathematical properties, similar to those of a basis, on the other hand.

The theory of **frames** gives the appropriate framework! The special case $f_k = g_{m,n}$ is called Gabor or Weyl-Heisenberg frame.

### Framebounds

In the finite discrete case of $f \in \mathbb{C}^L$ a collection $\{g_{m,n}\} \in \mathbb{C}^L$ with $k = NM$ can only be a frame, if $L \leq k$ and if the matrix $G$, defined as the $k \times L$ matrix having $g_{m,n}$ as its $(m + nM) - 1$st row, has full rank. In this case the frame bounds are the maximum and minimum eigenvalues of $S$, respectively. They yield information about numerical stability. The closer the frame-bounds are, the closer the frame operator will be to a diagonal matrix. If the frame bounds differ too much, the inversion of the frame operator is numerically unstable. The inversion of the frame operator provides reconstruction, as

---

4 Calculating the inner product in every point of the time-frequency lattice would yield the **full short-time Fourier transform**, a representation of redundancy $L$. The term “short-time Fourier transform” is often used for sampled short-time Fourier transforms as well. The spectrogram, its modulus squared, is one of the most popular time-frequency representations.

5 The **Balian-Low theorem** is a key result in time-frequency analysis. It states that if a Gabor system in $L^2(\mathbb{R})$ forms an orthogonal basis for $L^2(\mathbb{R})$, then

$$ \left( \int_{-\infty}^{\infty} |g(t)|^2 dt \right) \left( \int_{-\infty}^{\infty} |\omega g(\omega)|^2 d\omega \right) = +\infty. $$

There exist several extensions, first of all to the case of **exact** frames, which are frames that cease to be complete when any element is deleted.

6 Pure sinusoid form an orthogonal basis and are a nice example for functions having unbounded support in time.
3.3 The method of frames

we saw in Proposition 11. The canonical dual frame \( \tilde{g}_{m,n} \), which yields reconstruction as in (3.4), is given by

\[
\tilde{g}_{m,n} = S^{-1}g_{m,n}
\]
as

\[
f = S^{-1}Sf = \sum \langle f, g_{m,n} \rangle S^{-1}g_{m,n} = \sum \langle f, g_{m,n} \rangle \tilde{g}_{m,n}
\]

In the case of a tight frame, we have \( S = C_uI \). (I denotes the identity operator) and therefore \( S^{-1} = \frac{1}{C_u}I \).

The next section introduces the special case arising from applications in audio signal processing. We will see that in this case the frame operator takes a simple form.

**Gabor frames for audio**

Let us from now on assume that we are given a signal \( f \in \mathbb{C}^L \). This signal represents a piece of music or a spoken sentence etc., which we are interested to investigate and/or modify. Modifications might aim to achieve noise reduction in old or degraded recordings. Another issue might be the extraction of certain features of the signal, for example single instrument components. Let us further assume that an engineer approaches the problem by using a Fourier transform of length \( l \) in a first step. This implies that the window used for cutting out the part of interest must have this length. Looking at the definition of the Gabor coefficients:

\[
c_{m,n} = \langle f, g_{m,n} \rangle = \sum_{j=0}^{L-1} f(j)g_{m,n}(j)
\]
as an inner product, which can be interpreted as correlation between the window and the respective part of the signal, we can see that the signals \( f \) and \( g_{m,n} \) must have the same length, at least theoretically. Practically, of course, as \( l \ll L \), most of the “theoretical” \( g \) would be zero. As we don’t tend to waste computation time on multiplying with 0, only the effective length of \( g \), here \( l \), is multiplied with the part of interest of \( f \). This procedure implicitly introduces a frequency lattice constant \( b = \frac{L}{l} \). The time constant \( a \) is related to what is often called overlap. If \( a = \frac{l}{2} \) or \( a = \frac{l}{4} \), the overlap is \( \frac{1}{2} \) and \( \frac{3}{4} \), respectively. The redundancy of the representation is thus given by \( \text{red} = \frac{L}{a} \), e.g., if the overlap is half the window length, we get twice as many data points as in the original signal. This is in accordance with the general case where

\[
\text{red} = \frac{L}{a} = \frac{L}{\alpha} = \frac{L}{\frac{L}{b} = \frac{L}{a\frac{l}{b} = \frac{l}{a}}}
\]

**Remark:**

The reduction of redundancy from \( L \) in the case of the full short-time Fourier transform to a reasonable amount of redundancy in the Gabor setting ensures a balance between computability on the one hand and sufficient localisation on the other hand. The choice of a reasonable window-length and overlap common in applications corresponds roughly
3 Digging the audio structure: Frames and Redundancy

to such a rather balanced situation in the Gabor setting. Gabor theory, though, allows for more general choices of lattices, concerning the redundancy as well as the distribution of the lattice-points. It also provides detailed knowledge about the dependence of results on the choice of analysis parameters. This is especially decisive in the case of modification of the synthesis coefficients, which are non-unique due to the

Let us now look at the calculation of the inner products $c_{m,n} = \langle f, g_{m,n} \rangle$ more closely. They can also be written as

$$ (c_{m,n})_{m=1,\ldots,M; n=1,\ldots,N} = G \cdot f $$

where $G$ is the operator (matrix) introduced in Section 3.3.2. $G$ consists of blocks

$$ G_n, \quad n = 0, \ldots, N - 1 $$

each corresponding to one time-position of the window $g$. If we define $g^l$ as the restriction of $g \in \mathbb{C}^L$ to its non-zero part of length $l$, we get the following. The block $G_n$ acts on the samples $f(na + 1), \ldots, f(na + l) =: f_{na}(t)$ by taking inner products of this slice $f_{na}$ of the signal with each of the $l$ modulated windows

$$ M_{mb}g^l(t) = e^{-\frac{2\pi imbL}{l}}g^l(t) $$

$$ m = 0, \ldots, M - 1 \quad \text{and} \quad t = 0, \ldots, l - 1 $$

The coefficients $e^{-\frac{2\pi imbL}{l}}$ are exactly the entries of the Fourier matrix $\mathcal{F}_l$ of the FFT of length $l$ with $\hat{f} = \mathcal{F}_l f$. Therefore

$$ G_n f_{na}(t) = \mathcal{F}_l(f_{na} \cdot g^l)(t) $$

$$ t = 0, \ldots, l - 1 \quad \text{and} \quad n = 0, \ldots, N - 1 $$

and the action of $G_n$ on $f_{na}$ corresponds to multiplying $f$ with $g$, skipping zero entries and taking the Fourier transform of the non-zero part.

Remarks:

1. Although for implementation in real-life situations, the FFT-approach is always preferred, it is useful to look at the expansion from an operator point of view. Many important theoretical issues, yielding better understanding also for the applications, can be investigated more easily.

2. As mentioned before, all operators in Gabor theory generally act on the whole signal length $L$. In the definition of the building blocks $g_{m,n}$, the modulation operator is therefore defined as

$$ M_{mb}g(t) = e^{-\frac{2\pi imbL}{L}}g(t) \quad \text{for} \quad m = 0, \ldots, N - 1 \quad \text{and} \quad t = 0, \ldots, L - 1 $$

The blocks $G_n$, as opposed to the situation arising from implementation as discussed above, will not have identical entries, as the zero entries are in different positions.
3.3 The method of frames

Example:
Let \( g \in \mathbb{C}^{32} \) with
\[
g(t) = \begin{cases} 
0 & \text{for } t = 0, \ldots, 7 \\
0 & \text{else}
\end{cases}
\]
Then (by assumption \( b = \frac{l}{T} \), so that \( e^{-2\piimb} = e^{-\frac{2\pi imt}{T}} \))
\[
M_{mb}g(t) = (g(0), e^{-\frac{2\pi im}{T}} g(1), e^{-\frac{2\pi im}{T}} g(2), \ldots, e^{-\frac{2\pi im}{T}} g(7), 0, \ldots, 0)
\]
whereas
\[
M_{mb}T_ag(t) = (0, \ldots, 0, e^{-\frac{2\pi ima}{T}} g(a), e^{-\frac{2\pi ima}{T}} g(a+1), \ldots, e^{-\frac{2\pi ima}{T}} g(a+7), 0, \ldots, 0)
\]
\( e^{-\frac{2\pi ima}{T}} \) is not necessarily 1, so that the blocks will differ by a phase factor.

3. The restriction that \( a \) be a divisor of \( l \) is also due to the usual choice of parameters in applications. Two common cases would be \( a = \frac{l}{2} \) and \( a = \frac{l}{4} \), in which cases the number of different kinds of blocks reduce to 2 and 4, respectively.

The difference only concerns the phase spectrum, which is usually not considered in further processing, except for reconstruction. The dual window does not depend on the phase factor in the case discussed in the theorem as will be seen below.

Mastering the frame operator - the Walnut representation

Let us now come back to the central question of how to find a set of windows \( \tilde{g}_{m,n} \) for reconstruction as in (3.4). If it is possible to find a window \( \tilde{g} \) which is smooth and similar to the original window \( g \) especially in decaying to zero and if the rest of the dual family can be deduced in analogy to the Gabor analysis function set by time-frequency shifts, this will make reconstruction in a kind of overlap-add process easier. In fact, all the above conditions can be fulfilled. Generally, the elements of the dual frame \( (\tilde{g}_{m,n}) \) are generated from a single function (the dual window \( \tilde{g} \)), just as the original family. This follows from the fact that \( S \) and \( S^{-1} \) (the frame operator and its inverse) commute with the modulation and translation operators \( M_{mb} \) and \( T_{ma} \), for \( m = 1, \ldots, M \) and \( n = 1, \ldots, N \).

The higher redundancy, the closer the shape of the dual window gets to the original window’s shape. As in applications redundancy of 2, 4 or even higher are common, well localised dual windows can be found. Even more is true. The special situation in which the effective length of the window \( g \) equals or is shorter\(^7\) than the FFT-length, the frame operator takes a surprisingly simple form.

From the definition of the frame operator
\[
Sf = \sum_{m,n} \langle f, g_{m,n} \rangle g_{m,n}
\]
\(^7\)E.g. in the case of zero padding.
we deduce that the single entries of $S$ are given by

$$S_{j,k} = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} M_mbT_{na}g(j)\overline{M_mbT_{na}g(k)}$$

Looking at the inner sum, note that $\sum_{m=0}^{M-1} e^{2\pi imb(j-k)/L} = 0$ if $(j - k)$ is not equal to 0 or a multiple of $M$. In these cases

$$\sum_{m=0}^{M-1} e^{2\pi imb(j-k)/L} = 0$$

This leads to the *Walnut representation* of the frame operator for the discrete case:

$$S_{j,k} = \begin{cases} M \sum_{n=0}^{N-1} T_{na}g(j)\overline{T_{na}g(k)} & \text{if } |j - k| \mod M = 0 \\ 0 & \text{else} \end{cases} \quad (3.6)$$

There will obviously be non-zero entries in the diagonal, $j = k$, but as $M = l$, i.e. the window-length, $j = k$ is in fact the only case for which $|j - k| \mod M = 0$ holds and $g(j)$ and $g(k)$ both have non-zero values. Therefore, the frame operator is diagonal and the dual window $\hat{g}$ is calculated as

$$\hat{g}(t) = g/(M \sum_{n=0}^{N-1} T_{na}|g(t)|^2)$$