Chapter 3

Abstraction of the sampling problem

To analyze sampling and reconstruction, we formulate these operations in the context of vector spaces. We consider signals as elements of a normed linear space. Sampling and reconstruction are then operators, which are mappings between vector spaces. This abstraction proves to be very useful to generalize the problem and to find expressions for alternative sampling and reconstruction operators. The vector space setting was already proposed by Shannon and thereafter widely adopted in the mathematical literature, but is not commonly known in the engineering community.

In Section 3.1, we introduce vector spaces, and we discuss the abstract vector space of signals. The theory of normed linear spaces and Hilbert spaces is presented in Sections 3.2 and 3.3, respectively. This material is rather basic but included to make this thesis self-contained. Using the concepts of vector spaces, we formulate sampling and reconstruction as abstract operators and the composite operator as a projection on a certain subspace of the signal space, which we discuss in Section 3.4. Then, in Section 3.5, we formulate the most general operators that do the job, and reduce the set of candidate operators guided by practical considerations. This leads to a general sampling scheme, which models the composition of sampling and reconstruction. The scheme has two filters which must serve additional criteria to achieve a good approximation. In Section 3.6, we state some results from approximation theory, which may help us to find good filters.

3.1 Signals as vectors

A vector space over a scalar field $K$ ($K = \mathbb{R}$ or $\mathbb{C}$) consists of a set $X$ which is closed for two dyadic operations: vector addition and multiplication by a scalar. The addition is an operation $+: X \times X \to X$, and the scalar multiplication is an operation $\cdot: K \times X \to X$, such that the following axioms hold $(x, y, z \in X, \ a, b \in K)$:

- $(x + y) + z = x + (y + z)$ associativity
- $x + y = y + x$ commutativity
- $x + 0 = x$ zero element
- $x + (-x) = 0$ inverse element
\[ a(bx) = (ab)x \quad \text{associativity} \]
\[ a(x + y) = ax + ay \quad \text{distributivity} \]
\[ (a + b)x = ax + bx \quad \text{distributivity} \]
\[ 1x = x \quad \text{identity} \]

The space is denoted as \( \{X, +, \cdot\} \), or just \( X \). The Euclidian space \( \mathbb{R}^n \), with vector addition \( x + y = (x_1 + y_1, \ldots, x_n + y_n) \) and multiplication \( ax = (ax_1, \ldots, ax_n) \) is the most common example of a vector space.

In Chapter 2 we defined a signal as a mapping \( x : T \to A \). Such signals can be considered as vectors in a linear space, defining vector addition \( (x+y)(t) = x(t) + y(t) \) and scalar multiplication as \( (ax)(t) = a \cdot x(t) \). In order to guarantee closedness for addition and scalar multiplication, it is sufficient\(^1\) that signals are either real or complex valued, i.e., the signal range \( A = \mathbb{R} \) or \( A = \mathbb{C} \). It is easy to verify that the axioms given above indeed hold.

In practice, a signal is observed during a finite time interval, say \([0, T]\). Moreover, we assume that a real-world signal can be adequately modelled by a continuous and bounded function. The natural space to consider therefore appears to be \( C_b[0, T] \), the space of real-valued, continuous and bounded functions defined on \([0, T]\).

### 3.2 Normed linear spaces

In this section, we give an overview of normed linear spaces, i.e., vector spaces furnished with a norm. We start with the basic concepts of normed spaces and then address the theory of linear operators defined on them.

#### 3.2.1 Basic concepts

A **norm** generalizes the length of a vector in \( \mathbb{R}^n \). It is an operation \( \| \cdot \| : X \to \mathbb{R} \) with the following properties:

- \( \|x\| \geq 0 \)
- \( \|x\| = 0 \iff x = 0 \)
- \( \|ax\| = |a| \cdot \|x\|, \ a \in K \)
- \( \|x + y\| \leq \|x\| + \|y\| \quad \text{(triangle inequality)} \)

The pair \((X, \| \cdot \|)\) is called a **normed linear space**. A norm induces a **metric** \( d(x,y) = \|x - y\| \). A metric is a generalization of the distance in \( \mathbb{R}^n \), and satisfies the following properties:

- \( d(x,y) \geq 0 \), real valued and finite
- \( d(x,y) = 0 \iff x = y \)

\(^1\)A necessary condition is that the signal range itself is a linear space.
3.2 Normed linear spaces

- $d(x, y) = d(y, x)$ (symmetry)
- $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality)

It is clear that the metric induced by a norm satisfies these properties, and, hence, all normed linear spaces are metric spaces.

Examples of normed linear spaces are $\mathbb{R}^n$ with norm $\|x\|_2 = \left(\sum_k |x_k|^2\right)^{1/2}$, or $\mathbb{R}^n$ with norm $\|x\|_\infty = \max|x_k|$. Note that the latter norm does not conform with our usual notion of vector length. However, it is a valid norm. An example of an abstract space is $C[a, b]$, the space of all continuous functions defined on the interval $[a, b]$, with norm $\|x\|_\infty = \sup_{t \in [a, b]} |x(t)|$.

Another important space is $l^\infty$, whose elements are bounded sequences $x = (x_k)_{k \in \mathbb{K}}$, for which $x_k < \infty$. Here $\mathbb{K}$ is a countable index set (e.g., $\mathbb{K} = \mathbb{N}$ or $\mathbb{K} = \mathbb{Z}$). The corresponding norm is $\|x\|_\infty = \sup_{k \in \mathbb{K}} |x_k|$. Note that the elements of $l^\infty$ can be seen as bounded discrete-time signals.

A normed linear space $X$ is called complete if every Cauchy sequence converges in $X$, i.e., has its limit in $X$. We recall the definition of a Cauchy sequence:

$$(\forall \varepsilon > 0) \ (\exists N = N(\varepsilon)) \ (\forall m, n > N) \ (\|x_m - x_n\| < \varepsilon).$$

Complete normed linear spaces are called Banach spaces. The examples given above are all Banach spaces. A normed space which is not complete can be completed. An example of an incomplete space is $C[0, 1]$ with the norm $\|x\|_1 = \int_0^1 |x(t)| \, dt$.

To illustrate this, consider the functions $x_m$ shown in Figure 3.1. They form a Cauchy sequence, since $d(x_m, x_n) < \varepsilon$ when $m, n > 1/\varepsilon$. However, the functions converge to a step function, which does not belong to $C[0, 1]$. Hence, the space $(C[0, 1], \| \cdot \|_1)$ is not complete.

![Figure 3.1: A sequence of continuous functions $x_m$ converging to a discontinuous function.](image)

If a subset $Y$ of a vector space $X$ is closed for vector addition and scalar multiplication, it is called a linear subspace. A subspace is not necessarily complete. If $X$ is a Banach space, a subspace $Y$ is complete if and only if the set $Y$ is closed in $X$. Often, subspaces are defined in terms of linear combinations of vectors. Consider a finite, non-empty subset $S$ of the space $X$: $S = \{x_1, x_2, \ldots, x_n\}$. A linear combination is given by

$$x = \sum_{k=1}^n c_k x_k, \ c_k \in \mathbb{C}. \quad (3.1)$$

The set of all linear combinations is a subspace of $X$, as it is closed for $+$ and $\cdot$. This subspace is the span of the vectors and denoted by $\text{span}\{x_1, \ldots, x_n\}$. 
A set of vectors $x_1, \ldots, x_n$ is called \textit{linearly dependent} if there is a finite linear combination with scalars $c_i$ not all zero, such that

$$\sum_{k=1}^{n} c_k x_k = 0.$$ 

If such a combination cannot be found, the set is called \textit{linearly independent}. If a space is spanned by $n$ linearly independent vectors, the space is called \textit{finite dimensional} and has dimension $n$. Those $n$ vectors form a \textit{basis} of $X$. Any vector can be uniquely represented by a linear combination of the basis vectors

$$x = \sum_{k=1}^{n} c_k x_k.$$ 

A finite dimensional subspace of a normed space $X$ is always closed in $X$ and hence complete. It can be shown that every vector space $X \neq \{0\}$ has a basis. For the case of infinite dimensional spaces, suppose we can find a sequence $\{x_k\}_{k=1}^{n}$, such that

$$\lim_{n \to \infty} \|x - \sum_{k=1}^{n} a_k x_k\| = 0$$

with a unique combination of scalars $a_k$, then $\{x_k\}_{k=1}^{n}$ is called a \textit{Schauder basis}. In a Schauder basis, the ordering of the $x_k$ may be important for convergence of the sum. A basis for which the ordering does not matter is called an \textit{unconditional basis}. Not all Banach spaces have unconditional bases, e.g., $L^1(\mathbb{R})$ and $L^\infty(\mathbb{R})$ do not.

A space $X$ is \textit{separable} if there is a countable subset which is dense\footnote{A subset $S$ is said to be dense in a space $X$ if its closure equals $X$.} in $X$. If a normed space has a Schauder basis, then it is separable, but the opposite is not true [Kre78].

To end this section we list some commonly used Banach spaces:

- $l^p$
  The spaces $l^p$ with $1 \leq p < \infty$ consists of real- or complex-valued sequences $\{x_k\}_{k \in \mathbb{N}}$ such that $\sum_{k \in \mathbb{N}} |x_k|^p < \infty$. The corresponding norm is

$$\|x\|_p = \left( \sum_{k \in \mathbb{N}} |x_k|^p \right)^{1/p}.$$ 

The space $l^2$ is a direct generalization of the Euclidian space $\mathbb{R}^n$. All $l^p$ spaces are separable. Note that, in general, the $l^p$ spaces can be used as spaces of discrete-time signals.

- $l^\infty$
  The space $l^\infty$ was already discussed. Its norm is

$$\|x\|_\infty = \sup_{k \in \mathbb{N}} |x_k|.$$ 

The space $l^\infty$ is not separable [Kre78].
The spaces $L^p(\mathbb{R})$ with $1 \leq p < \infty$ are other examples of function spaces. Their elements are functions $x$ defined over the reals with $\left( \int_{-\infty}^{\infty} |x(t)|^p \, dt \right)^{1/p} < \infty$. The norms are:

$$
\|x\|_p = \left( \int_{-\infty}^{\infty} |x(t)|^p \, dt \right)^{1/p}.
$$

The $L^p$ spaces are formally defined within the Lebesgue theory\(^3\).

The space $L^\infty(\mathbb{R})$ consists of bounded functions with norm

$$
\|x\|_\infty = \sup_{t \in \mathbb{R}} |x(t)|.
$$

This norm is also referred to as the supremum norm.

The spaces $L^p[a, b]$ consist of functions defined on the interval $[a, b]$ with norm

$$
\|x\|_p = \left( \int_a^b |x(t)|^p \, dt \right)^{1/p}.
$$

Similarly we can define $L^\infty[a, b]$ with norm $\|x\|_\infty = \sup_{t \in [a, b]} |x(t)|$.

In Section 3.1 we motivated the choice of the space $C_b[0, T]$ of continuous and bounded functions on $[0, T]$ as signal space. This space is not complete. If a complete space is required for any reason, the space must be completed. Depending on the choice of the norm, we have different completions: $L^\infty[0, T]$ for the supremum norm and $L^p[0, T]$ for the $\| \cdot \|_p$ norm.

### 3.2.2 Linear operators and functionals

An operator is a mapping between two spaces. Typical operators in the Euclidian space $\mathbb{R}^2$ are rotation and projection. Of special interest are linear operators:

**Definition 3.1 (Linear operator)** A linear operator\(^4\) $T$ is a mapping such that:

1. the domain $\mathcal{D}(T)$ of $T$ is a vector space over a scalar field $K$, and the range $\mathcal{R}(T)$ lies in a vector space over the same field,

2. for all $x, y \in \mathcal{D}(T)$ and scalars $a$ holds:

$$
\begin{align*}
\{ \quad & T(x + y) = Tx + Ty \\
& T(ax) = aTx \quad \}
\end{align*}
$$

\(^3\) $L^p(\mathbb{R})$ consists of Lebesgue measurable functions, such that the Lebesgue integral $\left( \int_{-\infty}^{\infty} |x(t)|^p \, dt \right)^{1/p}$ is finite. The formal definition is required to prove completeness, but in most cases one can treat the integrals as Riemann integrals.

\(^4\) Operators will be denoted in boldface capitals.
An example is the identity operator $I : X \to X$ which gives $Ix = x$ for all $x \in X$. Denoting by $\mathcal{P}[a,b]$ the space of all polynomials on $[a,b]$, we can define the differentiation operator $D : \mathcal{P}[a,b] \to \mathcal{P}[a,b]$, $Dx(t) = x'(t)$.

An operator is called bounded if there is a real number $c$ such that

$$\|Tx\| \leq c\|x\|. \quad (3.2)$$

It is obvious that the identity operator is bounded. The differentiation operator is not bounded, e.g., let $x_n \in \mathcal{P}[0,1]$ with norm $\|x\| = \max_{t \in [0,1]} x(t)$, $x_n(t) = t^n$ where $n \in \mathbb{N}$. Then $\|x_n\| = 1$ and $\|Dx_n\| = n$. Since $n$ is arbitrary, there is no real number $c$ such that (3.2) holds.

An operator is continuous in $x_0$ if for every sequence $x_n$ tending to $x_0$ follows that $Tx_n \to Tx_0$. The operator is continuous if it is continuous for every $x_0 \in \mathcal{D}(T)$. For linear operators, continuity and boundedness coincide.

A functional $^5$ is an operator from a vector space $X$ into the reals $\mathbb{R}$ or the complex numbers $\mathbb{C}$.

For example, the point evaluation functional on the space $C[a,b]$ is defined as:

$$f_t : C[a,b] \to \mathbb{R}, \quad f_t(x) = x(t), \quad t \in [a,b].$$

It is a bounded linear functional. Another example is the norm $\|\cdot\| : X \to \mathbb{R}$, which is not linear.

### 3.3 Hilbert spaces

In a special class of linear spaces, we can define an inner product, which is a generalization of the scalar product in the Euclidian space. The inner product is denoted as $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$, and satisfies the following properties:

- $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- $\langle ax, y \rangle = a\langle x, y \rangle$
- $\langle x, y \rangle = \overline{\langle y, x \rangle}$
- $\langle x, x \rangle \geq 0$, $\langle x, x \rangle = 0 \iff x = 0$

Here, the bar denotes complex conjugation. For real vector spaces, the third property reduces to the symmetry property: $\langle x, y \rangle = \langle y, x \rangle$. The combination $(X, \langle \cdot, \cdot \rangle)$ is called an inner product space or pre-Hilbert space. A complete inner product space is called a Hilbert space. The inner product induces a norm:

$$\|x\| = \langle x, x \rangle^{1/2}.$$ 

Hence, Hilbert spaces are special cases of Banach spaces.

An example of an inner product space is the space $\mathbb{R}^n$ with inner product $\langle x, y \rangle = \sum_{k=1}^n x_k y_k$.

---

$^5$Functionals will be denoted in lowercase: $f : X \to \mathbb{R}$. 
In the Euclidian spaces $\mathbb{R}^2$ or $\mathbb{R}^3$ the inner product of two vectors is related to the angle between them. In particular, two vectors are parallel if their angle is zero, and perpendicular if the angle is 90 degrees. This can be generalized to an inner product space: Two vectors $x$ and $y$ are said to be perpendicular or orthogonal if their inner product is zero:

$$x \perp y \iff \langle x, y \rangle = 0.$$ 

Vectors in an inner product space satisfy the so-called Cauchy-Schwarz inequality

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

Cauchy-Schwarz inequality.

More examples of Hilbert spaces are:

- the sequence space $l^2$ with inner-product $\langle x, y \rangle = \sum_{k=-\infty}^{\infty} x_k \overline{y_k}$,
- the function space $L^2(\mathbb{R})$ with inner product $\langle x, y \rangle = \int_{-\infty}^{\infty} x(t) \overline{y(t)} \, dt$,
- and $L^2[a, b]$ with inner product $\langle x, y \rangle = \int_{a}^{b} x(t) \overline{y(t)} \, dt$.

### 3.3.1 Orthogonal systems and bases

Suppose we have a countable set of vectors $\{x_k\}_{k \in \mathbb{K}}$ in a Hilbert space $\mathcal{H}$. Here $\mathbb{K}$ is a countable index set, e.g., $\mathbb{N}$. The set $\{x_k\}_{k \in \mathbb{K}}$ is called a system (or indexed set, family). An orthogonal system $\{e_k\}$ is a system for which $e_m \perp e_n$, $m \neq n$. If all $e_k$ have norm 1, then the system is orthonormal. In this case we have $\langle e_m, e_n \rangle = \delta_{mn}$. For example, the vectors $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$ are an orthonormal system in $\mathbb{R}^3$.

We give some more definitions:

- A system $\{x_k\}$ is linearly independent if none of the vectors $x_k$ is contained in the closed linear span of the others. Any orthonormal system is linearly independent.

- A system is complete $^{6}$ (or total), if its span is dense in a normed space $X$, i.e.:

$$\text{clos}(\text{span}\{x_k\}) = X.$$ 

- A system $\{x_k\}$ forms a basis for a Hilbert space $\mathcal{H}$ if every vector $x \in \mathcal{H}$ can be uniquely written as

$$x = \sum_{k \in \mathbb{K}} c_k x_k, \quad c_k \in \mathbb{C}. \quad (3.3)$$

$^{6}$This notion of completeness is not to be confused with completeness of normed linear spaces.
• A system \( \{x_k\} \) is a Riesz basis (or unconditional basis\(^7\)) if it can be mapped onto an orthonormal basis \( \{e_k\} \) by a bounded invertible linear operator \( T \). In addition to the definition of a basis above, the coefficients \( c_k \) must satisfy:

\[
A \|x\|^2 \leq \sum_k |c_k|^2 \leq B \|x\|^2,
\]

for constants \( A \) and \( B \) with \( 0 \leq A \leq B < \infty \). Recall that the ordering of the \( x_k \) does not matter.

If a system is a basis, it is independent and complete. The opposite is not true, see [Zwa91, p. 12] for a counterexample. However, a complete orthonormal sequence is always a basis. Every separable Hilbert space has a complete orthonormal system\(^8\).

For any orthonormal system in an inner product space \( X \), we have

\[
\sum_{k \in \mathbb{N}} |\langle x, e_k \rangle|^2 \leq \|x\|^2 \quad \text{Bessel inequality},
\]

which turns into equality for a complete orthonormal sequence:

\[
\sum_{k \in \mathbb{N}} |\langle x, e_k \rangle|^2 = \|x\|^2 \quad \text{Parseval relation}.
\]

Given a vector \( x \) and an orthonormal basis \( \{e_k\} \), the corresponding coefficients in (3.3) can be obtained by simply taking inner products with the \( e_k \):

\[
x = \sum_{k \in \mathbb{N}} \langle x, e_k \rangle e_k. \tag{3.6}
\]

Note that, by the Parseval relation, every orthonormal basis is a Riesz basis with \( A = B = 1 \).

**Example 3.1** The space of bandlimited functions \( B_{\omega_{\text{max}}} \), consisting of the functions whose Fourier transform is zero outside the interval \( [-\omega_{\text{max}}, \omega_{\text{max}}] \), is a separable Hilbert space. The functions

\[
e_k(t) = \frac{\sin \omega_{\text{max}} (t - k\Delta)}{\omega_{\text{max}}(t - k\Delta)}
\]

form a complete orthogonal system in this space.

According to the Paley-Wiener theorem [Pap84], a bandlimited signal is an entire\(^9\) function of exponential type, i.e., there exist two constants \( A \) and \( \sigma \) such that

\[
x(t) < Ae^{\sigma|t|}.
\]

For an entire function its Taylor expansion converges for all \( t \). Hence, we have two different series expansions which fully determine the signal.

---

\(^7\)The name Riesz basis is reserved for Hilbert spaces.

\(^8\)Every Hilbert space (not necessarily separable) has a complete orthonormal set, but for non-separable spaces this set is not countable, hence it is not a system.

\(^9\)An entire function \( x(t) \) is (complex) analytic for all \( t \).
To find the coefficients to a non-orthonormal basis \( \{x_k\} \in \mathcal{H} \), the situation is not so simple. However, a relation similar to (3.6) can be found using the so-called bi-orthogonal basis (also dual or reciprocal basis). Two systems \( \{x_k\} \) and \( \{\hat{x}_k\} \) are called bi-orthogonal if

\[
\langle x_m, \hat{x}_n \rangle = \delta_{mn}, \quad m, n \in \mathbb{K}.
\]

If \( \{x_k\} \) is a basis, then \( \{\hat{x}_k\} \) is a basis, too, and any \( x \in \mathcal{H} \) can be written as

\[
x = \sum_{k \in \mathbb{K}} \langle x, x_k \rangle \hat{x}_k
\]
or

\[
x = \sum_{k \in \mathbb{K}} \langle x, \hat{x}_k \rangle x_k.
\]

The bi-orthogonal system can be calculated from the original system. Introducing the Gram-matrix\(^{10}\):

\[
G_{mn} = \langle x_m, x_n \rangle,
\]

the bi-orthogonal system is expressed by

\[
\hat{x}_m = \sum_k (G^{-1})_{mn} x_n.
\]

### 3.3.2 Representation of functionals on Hilbert spaces

A feature distinguishing Hilbert spaces over ordinary Banach spaces is that bounded linear functionals can be represented as inner products:

**Theorem 3.2 (Riesz’s theorem)** For every bounded linear functional \( f \) on a Hilbert space \( \mathcal{H} \) there is a unique element \( z \in \mathcal{H} \) such that \( f \) can be represented in terms of the inner product:

\[
f(x) = \langle x, z \rangle,
\]

where \( z \) depends on \( f \), and has the norm

\[
\|z\| = \|f\|.
\]

### 3.3.3 Projections

The general approximation problem of finding an element in a subspace \( Y \) of \( X \), which is closest to a given element \( x \in X \), is of major importance. In case of \( X \) being a Hilbert space \( \mathcal{H} \), the problem is solved by orthogonal projection of this element \( x \) on the subspace \( Y \). We now discuss projection.

**Definition 3.2** A vector space \( X \) is the direct sum of two subspaces \( Y \) and \( Z \) if each \( x \in X \) has a unique representation

\[
x = y + z, \quad y \in Y, \ z \in Z.
\]

The direct sum is written as

\[
X = Y + Z.
\]

\(^{10}\)In statistics, the Gram-matrix is known as the covariance matrix.
A special representation of a Hilbert space is the decomposition in a closed subspace $Y$ and its orthogonal complement:

$$\mathcal{H} = Y \oplus Y^\perp$$

where $Y^\perp$ is the orthogonal complement of $Y$ in $\mathcal{H}$:

$$Y^\perp = \{ z \in \mathcal{H} \mid z \perp Y \}. \quad (3.7)$$

Here, the vector $z$ is orthogonal to the subspace $Y$, denoted by $z \perp Y$, if it is orthogonal to all elements $y \in Y$.

The mapping $P : \mathcal{H} \to Y$ which yields an $y \in Y$ such that $x = y + z, z \in Y^\perp$, is an orthogonal projection of $\mathcal{H}$ onto $Y$. The projection operator is idempotent, i.e. $P^2 = P$. The null-space of the orthogonal projection is $Y^\perp$.

Let an element $x \in \mathcal{H}$ be projected on the subspace $Y$. Using (3.7), we have the following relation:

$$\langle x - y, y \rangle = 0, \quad (3.8)$$

where $y$ is the image of the projection. If $Y$ is a finite dimensional space, and $\{y_k\}$ is a basis in $Y$, then we have the unique representation:

$$y = a_1 y_1 + a_2 y_2 + \cdots + a_n y_n.$$  

The condition (3.8) leads to $n$ equations:

$$\langle y_m, x - y \rangle = \langle y_m, x - \sum a_k y_k \rangle = 0.$$  

which can be solved for $a_k$. The infinite dimensional case is similar and will be discussed in section 3.6.2.

### 3.3.4 Reproducing kernel Hilbert spaces

In this section we discuss a special class of Hilbert spaces called reproducing kernel Hilbert spaces. These spaces are completely characterized by a single quantity: the reproducing kernel. The reason for our attention is that Shannon’s sampling theorem, as well as the generalized theorems, can be formulated in this context.

**Definition 3.3 (Reproducing kernel Hilbert space)** A Hilbert space $\mathcal{H}$ of functions $x$ defined on the set $T$ (time axis) is called a reproducing kernel Hilbert space (RKHS) if all point evaluation functionals $f_t(x)$ are continuous (bounded).

According to Riesz’s theorem there exists a unique element $r_t \in \mathcal{H}$ such that

$$f_t(x) = x(t) = \langle x, r_t \rangle.$$  

Similarly, let $r_s \in \mathcal{H}$ be the element corresponding to $f_s(x)$. Every RKHS has a function $R(s,t) := \langle r_t, r_s \rangle$, called the reproducing kernel, such that

---

11The null-space of an operator $T$ is the set of all $x \in D(T)$ such that $Tx = 0$. It is denoted by $\mathcal{N}(T)$. 

1. $R(\cdot, t) \equiv r_t \in \mathcal{H}$

2. $x(t) = \langle x, R(\cdot, t) \rangle^{12}$

Note that $r_t(s) = R(s, t)$. This means that all evaluation functionals are determined by the kernel $R(s, t)$.

The reproducing kernel is unique, Hermitian, i.e., $R(s, t) = \overline{R(t, s)}$, and positive definite. On the other hand, for every Hermitian and positive definite $R$, there is a unique RKHS $= \text{span}\{R(\cdot, t)\}_{t \in \mathbb{T}}$, denoted by $\mathcal{H}_R$ [Mát89].

A (trivial) example of an RHKS is the space $l^2$. Its kernel is $R(k, n) = \delta_{kn}$. It is easy to verify that the space $L^2[a, b]$ is not an RHKS. A more interesting example is the Paley-Wiener space $\mathcal{B}_{\omega_{\text{max}}}$ of bandlimited signals. This is an RHKS with kernel

$$R(s, t) = \frac{\sin \omega_{\text{max}}(s - t)}{\omega_{\text{max}}(s - t)}. \quad (3.9)$$

An RKHS has the following properties:

- If $R(t, t)$ is a bounded function, then every $x \in \mathcal{H}_R$ is a continuous function.

- If $R$ is a continuous function, then $\mathcal{H}_R$ is separable. Hence, $\mathcal{H}_R$ is generated by a countable subset $\{R(\cdot, t)\}_{t \in \mathbb{T}}$ with rational coordinates.

- Every closed subspace $M$ of an RKHS is also an RKHS.

Construction of the kernel is simple if we have a complete orthonormal system in the space. In particular, if $\{e_k\}_{k \in \mathbb{N}}$ is a complete orthonormal system in the space $\mathcal{H}_R$, then the kernel has the form

$$R(s, t) = \sum_{k=1}^{\infty} e_k(s) \overline{e_k(t)}. \quad (3.10)$$

Moreover, we have the following projection theorem:

**Theorem 3.3 (Projection on RKHS)** If $\mathcal{H}$ is a Hilbert space and $M$ is an RKHS subspace of $\mathcal{H}$, then for every $x \in \mathcal{H}$

$$\text{Proj}_M x(t) = \langle R_M(\cdot, t), x \rangle$$

(3.11)

where $\text{Proj}_M$ is the projection operator onto $M$ and $R_M$ is the reproducing kernel of $M$.

For example, projection on the space $\mathcal{B}_{\omega_{\text{max}}}$ takes the following form:

$$\left(\text{Proj}_{\mathcal{B}_{\omega_{\text{max}}}} x\right)(t) = \left\langle \frac{\sin \omega_{\text{max}}(\cdot - t)}{\omega_{\text{max}}(\cdot - t)}, x \right\rangle = \int_{-\infty}^{\infty} x(s) \frac{\sin \omega_{\text{max}}(s - t)}{\omega_{\text{max}}(s - t)} \, ds \quad (3.12)$$

12 The dot is used here to denote a dummy variable, e.g., for the $L^2(\mathbb{R})$ inner product $\langle x, R(\cdot, t) \rangle = \int_{-\infty}^{\infty} x(s) R(s, t) \, ds$. 

Abstraction of the sampling problem

Also the sampling theorem itself can be formulated in an RKHS context. We know that \( \{R(\cdot, t_k); t_k = k\Delta; k \in \mathbb{Z}\} \) is a complete orthogonal system. Using (3.6) any \( x \in \mathcal{B}_{\text{b}_{\text{max}}} \) can be written as

\[
x(t) = \sum_{k \in \mathbb{Z}} \langle x, R(\cdot, t_k) \rangle R(t, t_k)
\]

which is the sampling theorem, since \( \langle x, R(\cdot, t_k) \rangle = x(t_k) \).

The sampling theorem can be generalized to other RKHS [Mát89, NW91]:

**Theorem 3.4** Let \( R(t, s) \) be a Hermitian, positive definite function and let \( \{t_k\} \) be a sequence of points such that

1. \( \{R(\cdot, t_k)\} \) generates the space \( \mathcal{H}_R \),
2. \( R(t_m, t_n) = 0 \) for \( m \neq n \), and \( R(t_m, t_m) > 0 \).

Then every \( x \in \mathcal{H}_R \) can be written as

\[
x(t) = \sum_{k} x(t_k) \frac{R(t, t_k)}{R(t_k, t_k)}.
\]

In [NW91] examples are given for Kramer’s generalized sampling theorem and for non-orthogonal sampling expansions. Moreover, the authors provide bounds on the different sampling-errors in the \( L^2 \)-sense.

### 3.4 A geometrical perspective of signal sampling

In the previous sections we gave a brief introduction to vector spaces. With this knowledge, we now discuss sampling and reconstruction in this context, which will lead to a generalized notion of both operations.

Sampling a signal \( x \in \mathcal{C}_b[0, T] \) means extracting a finite number, say \( N \), of real-valued parameters from it (the samples). Now consider a collection of \( N \) independent vectors \( \{\phi_1, \ldots, \phi_N\}, \phi_k \in \mathcal{C}_b[0, T] \). The vectors span an \( N \)-dimensional subspace of \( \mathcal{C}_b[0, T] \). Using the parameters as weights to these vectors, we obtain a vector in this subspace

\[
\hat{x} = \sum_{k=1}^{N} c_k \phi_k, \quad c_k \in \mathbb{R}.
\]

This vector \( \hat{x} \) serves as a reconstruction of the signal. Sampling plus reconstruction, considered as a single operator, is therefore an approximation in an \( N \)-dimensional subspace \( V \) of \( \mathcal{C}_b[0, T] \). The quality of the overall operator is the difference \( d(x, \hat{x}) = \|x - \hat{x}\| \), see Figure 3.2.

Perfect reconstruction can only be achieved if the signal \( x \) lies in the subspace. In this case the composition of the sampling/reconstruction operator should not affect \( x \). Hence, the composite operator must be a projection.

The construction of a general sampling operator involves three steps:
3.5 The general sampling scheme

In this section we formulate requirements for the sampling and reconstruction operator, which are motivated by practice. We search for the most general operators that satisfy these requirements. The composition of the general sampling operator and the general reconstruction operator forms the general sampling scheme.
3.5.1 Requirements for the sampling operator

In standard sampling, \( N \) samples are obtained by taking point measurements at \( N \) equidistant points in the interval \([0, T]\). However, there is no reason to leave other methods aside. A ‘sample’ can be obtained by applying any functional:

\[
f : C_b[0, T] \rightarrow \mathbb{R}
\]

This statement is very general. For example, we could consider the first \( N \) Fourier coefficients as samples of a signal. However, this can hardly be seen as a sampling operator. In order to select a suitable functional, we now reduce the set of possible functionals as guided by practical considerations. We therefore assume the following requirements for the sampling operator, and thereby for the \( N \) different functionals \( \{f_k\}_{k=1,...,N} \):

- **Linearity**
  The functionals should be linear. This requirement has its origin in the fact that the sampled signal is subject to further (digital) processing. In order to have a transparent relation between the signal and its samples it is best to establish a linear relationship between them.

- **Boundedness**
  A bounded signal must give bounded samples. This implies that the functionals should be bounded, i.e., there exists a real number \( c \) such that:

  \[
  |f(x)| \leq c \|x\|, \quad \text{for all } x \in X.
  \]

- **Independence**
  The \( N \) functionals should be independent. If the functionals are dependent, then at least one of them, say \( f_m \), can be written as a linear combination of the others:

  \[
  f_m = \sum_{k \neq m} c_k f_k, \quad c_k \in \mathbb{R}.
  \]

  If this is the case, the corresponding sample is redundant, because it does not provide new information about the signal.

- **Time-invariance**
  It is not economical to have a different procedure for each of the \( N \) samples. This would require \( N \) different hardware or software modules. Instead, it is recommended to implement the sampling operator with a single procedure, initiated at \( N \) different times.

- **Equidistant evaluation**
  In order to prevent additional timing data to be stored with the samples, the procedure should be repeated at equidistant time instants. This implies that

  \[
  t_k = k \Delta, \quad \Delta = \frac{T}{N}.
  \]
We realize that the last two requirements are major limitations to the sampling operator. Of course, there are many situations where non-uniform sampling or adaptive (time-variant) sampling will do a better job. However, in the fields of data acquisition and process control, we require simple and robust devices and, therefore, we assume the requirements above.

In the next sections, we will analyze the requirements and finally come to the general sampling operator, i.e., the most general operator satisfying the requirements.

### 3.5.2 Bounded linear functionals

The set of all linear functionals defined on a vector space $X$ can itself be made into a vector space. This space is denoted by $X^*$ and is called the algebraic dual space of $X$. The sum of two functionals is defined as

$$(f_1 + f_2)(x) = f_1(x) + f_2(x),$$

and the product of a scalar $\alpha$ and the functional $f$ as

$$(\alpha f)(x) = \alpha f(x).$$

The set of all bounded linear functionals on a normed linear space $X$ constitutes a normed space $X'$ with norm defined by

$$
\|f\| = \sup_{x \in X, x \neq 0} \frac{|f(x)|}{\|x\|} = \sup_{x \in X, \|x\| = 1} |f(x)|.
$$

This space $X'$ is called the dual space. It is always a Banach space, whether or not the original space $X$ is complete. Since we are looking for a bounded linear functional, we have to examine the space $X'$ further. Depending on the type of the original space, we can be more explicit about the character of the elements in $X'$.

If $X$ is Hilbert space, Riesz’s theorem states that every bounded linear functional $f$ on $X$ can be represented in terms of the inner product. Hence, all elements of $X'$ are of the following form

$$f(x) = \langle x, \lambda \rangle. \quad (3.13)$$

If $X$ is a Banach space, the situation is less straightforward. However, for the case $X = C[a, b]$ there is another theorem by Riesz. First, we define:

**Definition 3.4 (Bounded variation)** A function $x$ defined on $[a, b]$ is of bounded variation on $[a, b]$ if its total variation $\text{Var}(x)$ on $[a, b]$ is finite, where

$$\text{Var}(x) = \sup \sum_{k=1}^{n} |x(t_k) - x(t_{k-1})|,$$

and the supremum is over all partitions:

$$a = t_0 < t_1 < \cdots < t_n = b.$$
Theorem 3.5 (Riesz’s theorem (functionals on C[a,b])) Every bounded linear functional $f$ on $C[a,b]$ can be represented by a Riemann-Stieltjes integral:

$$f(x) = \int_a^b x(t) \, dw(t)$$ (3.14)

where $w$ is of bounded variation on $[a,b]$ and has the total variation

$$\text{Var}(w) = \|f\|.$$ 

In case the function $w$ is differentiable, (3.14) reduces to

$$f(x) = \int_a^b x(t) \, w'(t) \, dt.$$ (3.15)

Note the similarity between (3.15) and (3.13). Motivated by these two special situations we restrict ourselves to functionals that can be written as

$$f(x) = \int_a^b x(t) \, \lambda(t) \, dt,$$ (3.16)

where $\lambda$ is a sum of a piecewise continuous function and at most countably many Dirac distributions. The Dirac distributions are included to cover point functionals in the general expression. Recall that, $\delta$ being the Dirac distribution,

$$x(t) = \int_a^b x(s) \, \delta(t) \, dt.$$ 

Functionals that are not covered by this representation (3.16) will no longer be considered, because they are not interesting for practical purposes.

3.5.3 Time-invariance

According to representation (3.16), each functional $f_k$ is fully determined by a function $\lambda_k$. Requiring a time-invariant sampling operator, the $k$th function $\lambda_k$ must be a time-shifted copy of a standard function $\lambda$:

$$\lambda_k(t) = \lambda(t - t_k).$$

Defining the reflection $x^\circ$ of a function $x$ as

$$x^\circ(t) = x(-t),$$

and introducing the function $x_e$:

$$x_e(t) = \begin{cases} x(t) & \text{if } t \in [0,T) \\ 0 & \text{otherwise} \end{cases}$$

14This representation (3.16) can be proven to hold in the space of distributions as well, with $\lambda$ being a distribution.
we can express the $k^{\text{th}}$ functional as
\[
 f_k(x) = \int_0^T x(t) \lambda_k(t) \, dt \\
= \int_{-\infty}^{\infty} x(t) \lambda_k(t) \, dt \\
= \int_{-\infty}^{\infty} x(t) \lambda(t - t_k) \, dt \\
= \int_{-\infty}^{\infty} x(t) \lambda^\nu(t_k - t) \, dt \\
= (x \ast \lambda^\nu)(t_k) \quad (3.17)
\]
This implies that we can obtain all functionals by filtering the signal with a filter having impulse response $\lambda^\nu$, and sampling the output at the times $t_k$. Equidistant evaluation is obtained by setting
\[
t_k = k\Delta, \quad k = 1 \ldots N.
\]

### 3.5.4 The general sampling operator

The general sampling operator meeting our requirements can be regarded as a filter, followed by an ideal sampler, sampling at equidistant times, see Figure 3.3.

\[\begin{array}{c}
x(t) \\
\text{PRE} \\
f(t) \\
\text{ } \\
f(k\Delta)
\end{array}\]

**Figure 3.3**: The general sampling operator. The filter PRE has the impulse response $\lambda^\nu$. The filter yields a continuous-time signal $f$. Its samples $f(k\Delta)$ give the functionals $f_k$.

### 3.5.5 The reconstruction operator

Reconstruction will be obtained by using the $N$ samples taken according with the functionals defined in (3.17) as coefficients to a set of vectors $V = \{\phi_1, \ldots, \phi_N\}$ in the space $C_b[0, T]$:
\[
z(t) = \sum_{k=1}^{N} f_k \phi_k(t). \quad (3.18)
\]
The question remains which functions $\phi_k$ to use.

For reconstruction, we again have a list of requirements:

- **Linearity**
  The reconstruction operator should be linear. Like in the case of a sampling operator, it is preferred to have a linear relationship between samples and reconstructed signal.

- **Boundedness**
  Bounded samples must give a bounded reconstruction.
• **Independence**
  In order to avoid redundancy, the $N$ reconstruction functions should be independent.

• **Time-invariance**
  For economical reasons, the reconstruction operation must be implemented as a single procedure, initiated at $N$ different times.

• **Equidistant evaluation**
  Equidistant activation of the procedure avoids storage of timing data.

By using a linear combination of $N$ functions in $C_b[0, T]$, we automatically have a linear relationship between samples and reconstruction. Moreover, the linear combination will lie in the space $C_b[0, T]$, hence it is bounded.

The time-invariance implies that the reconstruction functions are time-shifted copies of a function $\phi$, i.e., we can express the $k$th reconstruction function as:

$$\phi_k(t) = \phi(t - t_k)$$

and again, for equidistant evaluation, $t_k = k\Delta$.

The problem now reduces to finding conditions on $\phi$ such that $V = \{\phi_1, \ldots, \phi_N\}$ constitutes a basis in a certain subspace of $C_b[0, T]$. Note that by choosing $\phi$, the subspace is implicitly defined.

The general reconstruction method can also be interpreted as a filter. Considering the sequence of functionals $f_k$, and introducing the weighted $\delta$-sequence:

$$f^*(t) = \sum_{k=1}^{N} f_k \delta(t - k\Delta) = f(k\Delta) \cdot I_\Delta(t)$$

where $I_\Delta(t)$ is the impulse train as defined in (2.10), the reconstruction (3.18) can be written as:

$$z(t) = \sum_{k=1}^{N} f_k \phi(t - k\Delta) = (f^* \ast \phi)(t)$$

(3.19)

The general reconstruction operator is shown in Figure 3.4.

![Figure 3.4: The general reconstruction operator. The filter POST has the impulse response $\phi$. Note that by choosing $\phi$, the subspace is implicitly defined.]
3.5.6 The general sampling scheme

In Figure 3.5 we have combined the sampling operator of Figure 3.3 with the reconstruction operator of Figure 3.4. The ideal switch is omitted here, since direct multiplication with the impulse train has the same effect:

\[ f^*(t) = I_\Delta(t) \cdot f(k\Delta) = I_\Delta(t) \cdot f(t). \]

We will refer to this scheme as the *general sampling scheme* (GSS). Note the resemblance to the scheme of the sampling theorem (Figure 2.6 in Chapter 2). The choices left are the sampling rate \( \Delta \) and the two filters PRE and POST.

![Diagram of sampling and reconstruction operator](image)

**Figure 3.5:** A combination of sampling and reconstruction operator. This scheme will be referred to as the general sampling scheme. The subspace is determined by the choice of the post-filter.

Note the similarity between this scheme and Shannon’s sampling with anti-alias filtering, shown in Figure 2.6 of Section 2.2.4. There, the pre-filter and the post-filter are both ideal low-pass filters. We will come back to this similarity in Section 3.6.4.

3.6 Approximation property of the general sampling scheme

The filters in the sampling and reconstruction operators of the previous section must be tuned in such a way that the combined operator constitutes a projection on the space spanned by the \( N \) basis functions \( \phi_k \). How can this be achieved? For this question to be answered, we have to focus on approximation theory.

3.6.1 Approximation theory

The general approximation problem in a normed space can be formulated as: Let \( X = (X, \| \cdot \|) \) be a normed space and \( Y \) a subspace. A given \( x \in X \) is to be approximated by a \( y \in Y \).

The distance between \( x \) and the subspace \( Y \) is by definition:

\[ d(x, Y) = \inf_{y \in Y} \| x - y \| \]

**Definition 3.5 (Best approximation)** An element \( y_0 \in Y \) is called a best approximation of \( x \) if

\[ \| x - y_0 \| = d(x, Y) \]

This best approximation does not always exist. Moreover, it is not unique in general. However, the existence is guaranteed in the finite dimensional case:
Theorem 3.6 If \( Y \) is a finite dimensional subspace of a normed space \( X \), then for each \( x \in X \) there exists a best approximation in \( Y \).

The problem of uniqueness is related to convexity. A subset \( A \) of a vector space \( X \) is called convex if all points of the ‘straight line’ between two points \( a \) and \( b \) \((a, b \in A)\) are in \( A \), i.e., for every \( a, b \in A \), the set \( L := \{\lambda a + (1 - \lambda)b \mid 0 \leq \lambda \leq 1\} \) is a subset of \( A \), see Figure 3.6.

![Figure 3.6: An example of a convex set (a) and a non-convex set (b).](image)

Definition 3.6 (Strictly convex norm) A strictly convex norm is a norm such that for all \( x, y \) of norm 1,

\[
\|x + y\| < 2, \quad x \neq y.
\]

A normed space with such a norm is called a strictly convex normed space. For example, any Hilbert space is strictly convex, whereas the space \( C[a, b] \) is not [Pow81]. We have the following theorem:

Theorem 3.7 (Uniqueness theorem) In a strictly convex normed space \( X \) there is at most one best approximation to a given point \( x \) in the subspace \( Y \).

In the following sections, we consider two types of approximations:

- uniform approximation in \( C[a, b] \) (Chebyshev approximation),
- least squares approximation in Hilbert spaces.

3.6.2 Approximation in Hilbert spaces

Since Hilbert spaces are strictly convex, there is at most one best approximation. The uniqueness is therefore guaranteed. The existence is given by the following theorem:

Theorem 3.8 For every \( x \) in a Hilbert space \( \mathcal{H} \) and closed subspace \( Y \) of \( \mathcal{H} \), there is a unique best approximation: \( y = \text{Proj}_Y x \), where \( \text{Proj}_Y \) is the orthogonal projection onto \( Y \).

How to determine the best approximation? For the case that \( Y \) is a reproducing kernel Hilbert space, the projection was given by (3.11). Given an orthonormal basis \( \{e_k\}_{k \in \mathbb{K}} \) of the subspace \( Y \), this projection is easily calculated. We have

\[
\text{Proj}_Y x = \sum_{k \in \mathbb{K}} \langle x, e_k \rangle e_k.
\]
In the general case of a Riesz-basis \( \{ y_k \}_{k \in \mathbb{R}} \), the projection is given by

\[
\text{Proj}_Y x = \sum_{k \in \mathbb{R}} \langle x, \hat{y}_k \rangle y_k,
\]

where \( \{ \hat{y}_k \}_{k \in \mathbb{R}} \) is the dual basis.

Approximation in a Hilbert space is also called mean square approximation. It minimizes the distance induced by the norm, which in turn is induced by the inner product. For example, approximation in the space \( L^2(\mathbb{R}) \) minimizes the distance

\[
d(x, Y) = \inf_{y \in Y} \left( \int_{-\infty}^{\infty} |x(t) - y(t)|^2 \, dt \right)^{1/2}.
\]

### 3.6.3 Approximation in \( C[a, b] \)

Approximation in \( C[a, b] \) with the sup-norm is also called minimax approximation. Indeed, the distance between two elements of \( C[a, b] \) is given by the maximum difference for all \( t \in [a, b] \). The existence of a best approximation in a general subspace of \( C[a, b] \) cannot be proven. For finite dimensional subspaces, however, the existence is guaranteed by Theorem 3.6. A sufficient condition for uniqueness is Haar’s condition [Pow81]:

**Definition 3.7 (Haar’s condition)** A finite dimensional subspace \( Y \) of the real space \( C[a, b] \) is said to satisfy Haar’s condition if every \( y \in Y, y \neq 0 \), has at most \( n - 1 \) zeros in \([a, b]\), where \( n = \dim Y \).

This condition is used in the following uniqueness theorem:

**Theorem 3.9 (Haar’s uniqueness theorem)** Let \( Y \) be a finite dimensional subspace of the space \( C[a, b] \). Then the best approximation in \( Y \) of every \( x \in C[a, b] \) is unique if and only if \( Y \) satisfies Haar’s condition.

A way to find a good approximation is to start with a trial approximation. This trial approximation can be improved iteratively, via the so-called exchange algorithm. If \( Y \) satisfies Haar’s condition, convergence is guaranteed.

The exchange algorithm yields a sequence of approximations. We can consider the approximation as an operator \( \text{Appr} \), so that we have a sequence \( \{ \text{Appr}_n x \}_{n \in \mathbb{N}} \). This operator lacks favorable properties. It may be non-linear and does not always have the projection property [Pow81].

We have to conclude that in the case of approximations on \( C[a, b] \), a systematic search for the best operator lies beyond the scope of this work. We can, however, evaluate the approximation properties of operators that are obtained by other means.

### 3.6.4 Example: Shannon’s sampling scheme

Recalling that the space of bandlimited functions \( B_{\omega_{\text{max}}} \) is a separable Hilbert space, spanned by the orthonormal basis

\[
e_k(t) = \frac{\sin \omega_{\text{max}}(t - k\Delta)}{\omega_{\text{max}}(t - k\Delta)},
\]
the orthogonal projection onto this space becomes

\[
\text{Proj}_{B_{\omega_{\text{max}}}}x(t) = \sum_k \langle x, e_k \rangle e_k(t). \tag{3.20}
\]

The basis has the special property that it is shift-invariant, i.e., \(e_k(t) = e(t - k\Delta)\). Therefore, following a similar reasoning as in (3.17), the inner product can be interpreted as a convolution evaluated at \(t = k\Delta\):

\[
\langle x, e_k \rangle = (x * e)(k\Delta),
\]

where the reflection symbol is left out, since \(e\) is symmetric \((e(-t) = e(t))\). The inner products will be produced by the system of Figure 3.3 with the pre-filter being the ideal low-pass filter. Also the reconstruction can be regarded as a convolution (see (3.19)), where the post-filter is the ideal low-pass filter. Hence, (3.20) fits into the general sampling scheme of Figure 3.5 and, moreover, is equal to Shannon’s scheme with anti-alias filter as given in Figure 2.6. We conclude that the latter scheme constitutes an orthogonal projection onto the space \(B_{\omega_{\text{max}}}\).

In Section 3.3.4 we encountered another expression for projection on \(B_{\omega_{\text{max}}}\), namely:

\[
\text{Proj}_{B_{\omega_{\text{max}}}}x(t) = \left\langle \frac{\sin \omega_{\text{max}}(\cdot - t)}{\omega_{\text{max}}(\cdot - t)}, x(s) \right\rangle = \int_{-\infty}^{\infty} x(s) \frac{\sin \omega_{\text{max}}(s - t)}{\omega_{\text{max}}(s - t)} \, ds.
\]

This expression has the form of a convolution:

\[
\text{Proj}_{B_{\omega_{\text{max}}}}x(t) = (x * \lambda)(t),
\]

with

\[
\lambda(t) = \frac{\sin \omega_{\text{max}}t}{\omega_{\text{max}}t}.
\]

Hence, pre-filtering with the ideal low-pass filter already gives the projection on \(B_{\omega_{\text{max}}}\), in other words the output \(y\) of the pre-filter is equal to the output \(z\) of the general sampling scheme (see Figure 3.5). From the engineer’s point of view this is nothing new: the filter removes all high-frequency components, yielding a bandlimited signal. And from the sampling theorem we know that a bandlimited signal is reconstructed without error. However, the fact that \(y = z\) is rather unique. In the general case it holds \(y \neq z\).

**Summary**

Sampling and reconstruction were formulated as abstract operations, mapping signals of a vector space onto another vector space. Samples are taken by applying functionals on a signal. The outcomes of these functionals serve as coefficients in the expansion with respect to a certain basis of a subspace. Suitable functionals were selected based on practical requirements. These functionals can all be modelled by a filter followed by point measurements. For reconstruction we have a similar outcome. Time-invariance implies that the reconstruction functions, which together span the subspace, must be time-shifted copies of a single function. The reconstruction process can also be seen as a filter with impulse response \(\phi\). This filter characterizes the subspace. In order to guarantee that samples represent a signal well, we require that the composition of sampling and reconstruction yields a good approximation of the signal. If sampling is studied in the context of Hilbert spaces, this approximation problem can be solved uniquely.
In other cases we did not find a systematic method for constructing an optimal approximation operator.

The approach we followed is similar to that of Bertero et al. [BdMP85, BdMP88]. They consider the so-called linear inverse problem with discrete data: Given a class $X$ of functions, a set $\{f_n; 1 \leq n \leq N\}$ of linear functionals defined on $X$, and a set $\{g_n; n = 1 \leq n \leq N\}$ of values of such functionals, then find in $X$ an object $x$ such that

$$f_n(x) = g_n, \quad n = 1, \ldots, N.$$ 

This setting includes interpolation of sampled data, where the functionals are the point evaluation functionals, but also recovery of a function from its first $N$ Fourier coefficients. Their work was restricted to Hilbert spaces.