Mathematical Methods

for Computer Graphics

Christian Lessig

‘‘It is through science that we prove,
but through intuition that we discover.’’

Henri Poincaré

Abstract

Mathematics plays an important role in computer graphics. These notes ...

We don't learn how to prove things.

Scientists approach: use mathematics to express ideas and model things ... and through this to solve the problems of interest.

What we do not talk about: - measure theory. - probabilistic techniques. - We also only discuss groups and Lie groups in passing.

Reproducible science: like in other science also in computer science experiments should be reproducible. Since we have only very few and highly standardized experimental setups, known as programming languages or environments, this is simpler than in any other science. For all experiments that are presented in the following source code is provided and the reader is encouraged to experiment with the experiments her self.[1]

[1]Source code is available at http://www.dgp.toronto.edu/people/lessig/teaching/math_for_cg/. See (Donoho, Maleki, Rahman, Shahram, and Stodden, "Reproducible Research in Computational Harmonic Analysis") for a more detailed discussion on reproducible computational science.

Contents

Chapter 1

Linear Algebra

Linear algebra is an elementary pillar of computer graphics, as is evident by the central place it takes in introductory computer graphics classes. In the following, however, we will consider some aspects that are usually not emphasized in computer graphics but that play important roles for us in subsequent chapters.

1.1 Linear Spaces

Definition 1.1.

1.2 Linear Spaces with Additional Structure

1.2.1 Norms

Banach space: Remark on separability.

1.2.2 Inner Products

Hilbert space: Remark on separability.

1.3 Linear Functionals

Theorem 1.1.

1.4 Bases for Linear Spaces

Define: - linear independence - span of a set of vectors. - dimension of a vector space.

1.4.1 Biorthogonal Bases

- Schauder basis. - Hamel basis.

1.4.2 Orthonormal Bases

1.4.3 Overcomplete Bases: Frames

Example 1. Mercedes Benz frame

Homework 1. *Generalize the Mercedes Benz frame to a frame for* $\mathbb{R}^3$*. Choose the normalization of the basis vectors such that one has a Parseval tight frame.*

Solution: The frame vectors of the Mercedes Benz frame in two dimensions are the vectors to the vertices of an equilateral triangle. The three dimensional analogue of the equilateral triangle is the regular tetrahedron.

Figure of regular tetrahedron (draw such that vertex position can be read off).

With the vectors to the vertices of the regular tetrahedron we obtain for the basis matrix

$$B = \begin{pmatrix} +1.0 & 0.0 & -1.0/\sqrt{2.0} \\ -1.0 & 0.0 & -1.0/\sqrt{2.0} \\ 0.0 & +1.0 & 1.0/\sqrt{2.0} \\ 0.0 & -1.0 & 1.0/\sqrt{2.0} \end{pmatrix}. \tag{1.1}$$

It is easily verified numerically that

$$2\,\mathrm{Id} = B^T\,B. \tag{1.2}$$

Hence, the Parceval tight frame is given by

$$B_{PTF} = \frac{1}{2}B. \tag{1.3}$$

That the vertices of the equilateral triangle and the regular tetrahedron yield tight frames can be seen as a special case of a result by Benedetto and Fickus[1] that characterizes tight frames as minimizer of a "force" that repels frame vectors and hence maximizes the distance between them.

1.5 Linear Maps

So far we only studied the elements of linear spaces. However, we are also interested in "transformations" of our vectors. The natural maps in a linear space are linear maps.

1.5.1 Fundamental Concepts

Definition 1.2 (Linear Maps)**.** *Let* V, W *be linear spaces. Then a mapping*

$$T : V \to W$$

is a ***linear map*** *when for* $u, v \in V$ *and* $a \in \mathbb{R}$*:*

i) $T(u+v) = T(u) + T(v)$;

ii) $T(av) = aT(v)$.

The ***domain*** *of* T *is* V *and the target of* T *is* W.

We hence see that linear maps are compatible with the essential linear structure of vector spaces.

Example 2. Let V be a linear space. Then the **identity map** $\mathrm{Id} : V \to \mathrm{V}$ is

$$\mathrm{Id}(v) = v \quad , \quad \forall v \in V.$$

Example 3. Let $V = \mathbb{R}^2$. The rotation operator $R(\theta)$ that rotates a vector by an angle θ is a linear operator.

Example 4. Let A_N be the space of sequences $a = (a_1, \cdots, a_N) \in A_N$ of length N. Then the left right operator $R : A_N \to A_N$ is

$$R(a) = (a_N, a_1, \cdots, a_{N-2}, a_{N-1}),$$

that is $(Ra)_i = a_{\mathrm{mod}(i+1,N)}$.

Example 5. Let us also consider some nonlinear operators, more will be provided in Chapter 2.5 when we consider linear maps on function spaces. The simplest examples are "nontrivial" functions on the real line. For example $\sin(x)$ and $\log(x)$ are nonlinear operators $N : \mathbb{R} \to \mathbb{R}$. On $\mathbb{R}^n$, operators of the form

$$N = \sum_{i=1}^{n} a_i v_i^{k_i} \tag{1.4}$$

acting on $v \in \mathbb{R}^n$ with coordinates v_i for real valued coefficients a_i and $k_i \geq 2$. Another common example of nonlinear operators, and once that often occur in physics, are

$$N = \sum_{i=1}^{n} a_i(v) v_i^{k_i} \tag{1.5}$$

where the coefficients $a_i(v)$ depend on the argument v and $k_i \geq 1$ now. Nonlinear operators between vector spaces are very general and hence many more examples could be constructed.

As is customary, we will often drop the brackets and write $Tv \equiv T(v)$.

Definition 1.3. *Let* $A : V \to W$ *and* $B : V \to W$ *be linear maps between linear spaces* V, W. *Then the* ***composition*** $B \circ A : V \to W$ *is the linear map*

$$(BA)(v) \equiv (B \circ A)(v) = B(A(v))$$

where $v \in V$.

One often writes AB without specifying a vector. When it is unclear what this means one should recall the definition in terms of the action on vectors. The following two concepts are important in many applications.

Definition 1.4 (kernel,range,)**.** *Let V, W be linear spaces and $T : V \to W$ be a linear map. Then the* ***kernel*** $\ker(T)$ *of T is*

$$\ker(T) = \{v \in V \mid T(v) = 0\}.$$

Then ***range*** $\mathrm{ran}(T)$ *(or image) is*

$$\mathrm{ran}(T) = \{w \in W \mid \exists v \in V : T(v) = w\}.$$

It is usually important to distinguish the target and range of an operator and they coincide only in special cases.

Theorem 1.2 (Rank-Nullity Theorem)**.** *Let V, W be linear spaces and $T : V \to W$ be a linear map. Then*

$$\dim(\ker(T)) + \dim(\mathrm{ran}(T)) = \dim(V).$$

The dimension $\dim(\mathrm{ran}(T))$ *of the range of T is the* ***rank of*** *T*

$$\mathrm{rank}(T) = \dim(\mathrm{ran}(T)).$$

We have established the essential properties of linear maps on vector spaces but not discussed how to work numerically with them. Without loss of generality, let V be a Hilbert space and $T : V \to V$ be a linear map. Furthermore, let $\{e_i\}_{i=1}^N$ be an orthonormal basis for V so that for any $v \in V$ we have

$$v = \sum_{i=1}^{N} v_i \, e_i = \sum_{i=1}^{N} \langle v, e_i \rangle . \tag{1.6}$$

Then for the linear map T applied to v we have

$$T(v) = T\left(\sum_{i=1}^{N} v_i \, e_i \right) \tag{1.7a}$$

and by linearity of T with respect to vector addition this equals

$$T(v) = \sum_{i=1}^{N} T(v_i \, e_i). \tag{1.7b}$$

Also using linearity with respect to scalar multiplication by the v_i we obtain

$$T(v) = \sum_{i=1}^{N} v_i \, T(e_i). \tag{1.7c}$$

Hence, the action of T on v is fully determined by the action on the basis vectors e_i. This should not come as a surprise since the e_i are, through their linear superposition, equivalent to any vector v.

$T(v)$ in the above form is not useful numerically. For this, we also need the coordinate representation of the image $\bar{v} = T(v)$. As we have seen before, we have $\bar{v}_j = \langle e_j, v \rangle$. Hence,

$$\langle e_j, T(v) \rangle = \left\langle e_j, \sum_{i=1}^{N} v_i \, T(e_i) \right\rangle \tag{1.8a}$$

and by the linearity of the inner product we have

$$\langle e_j, T(v) \rangle = \sum_{i=1}^{N} \langle e_j, v_i \, T(e_i) \rangle \tag{1.8b}$$

$$= \sum_{i=1}^{N} v_i \, \langle e_j, \, T(e_i) \rangle \tag{1.8c}$$

By defining

$$\langle e_j, T(v) \rangle = \sum_{i=1}^{N} v_i \underbrace{\langle e_j, \, T(e_i) \rangle}_{T_{ji}} \tag{1.8d}$$

we obtain

$$\langle e_j, T(v) \rangle = \sum_{i=1}^{N} T_{ji} v_i \tag{1.8e}$$

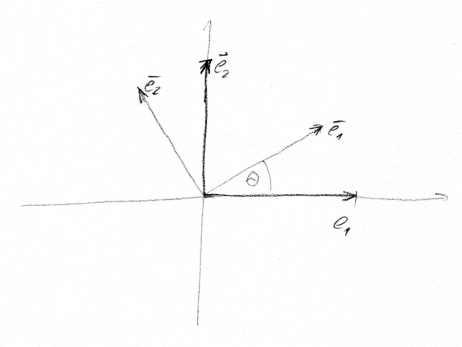
If we collect the coefficients T_{ji} in a two-dimensional "array" and the $\bar{v}_j$ and v_i into one-dimensional ones then we obtain

$$\begin{pmatrix} \bar{v}_1 \\ \vdots \\ \bar{v}_N \end{pmatrix} = \begin{pmatrix} T_{11} & \cdots & T_{1N} \\ \vdots & \ddots & \vdots \\ T_{N1} & \cdots & T_{NN} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}. \tag{1.9}$$

The foregoing derivation shows that matrices are the coordinate representation of linear operators. In computer science, matrices are often said to *be* linear operators. However, as we will see in the following it is useful, and at times important, to distinguish linear operators and their representations as matrices.

Homework 2. *Repeat the above derivation for the coordinate representation of a linear map for a biorthogonal basis* $(\{e_i\}_{i=1}^{N}, \{\tilde{e}_i\}_{i=1}^{N})$ *for a Hilbert space.*

Exercise 1. *Derive the matrix representation for a rotation by θ°, counter-clockwise, using the the above ansatz.*



For the above example of a rotation, we can easily determine the linear map that "inverts" the effect of the rotation, just use $-\theta^\circ$ instead of θ° and it is obvious that this is also a linear map. In general, the question when an inverse of a linear map exists is more subtle.

Proposition 1.1. *Let V, W be linear spaces and $T : V \to W$ be a linear map. Then T is **invertible** when it is bijective (one-to-one) and surjective (onto) onto W. T then has an **inverse** $T^{-1} : W \to V$ that is a linear operator and satisfies*

$$T^{-1}T = \mathrm{Id} \qquad TT^{-1} = \mathrm{Id}.$$

An important and non-trivial fact in the above proposition is that the inverse of a linear map $T : V \to W$ is also a linear map $T^{-1} : W \to V$ going in the "opposite direction" from W to V. At least intuitively it is easy to see that bijectivitiy (one-to-one) and surjectivity (onto) are necessary conditions for the existence of such a map. For example, if bijectivity would not hold and $T(v) = T(\bar{v}) = w \in W$ then it would not be clear if $T^{-1}(w)$ should map to v or $\bar{v}$. Similarly, if surjectivity would not hold then would be $T^{-1}(w)$ for a w not in the image of T. Unsurprisingly, the coordinate representation of the inverse T^{-1} of a linear map T is the inverse of a matrix. This provides a practical means to determine the inverse.[2]

Although an inverse in the above sense is commonly used, for an application it might be irrelevant which $v \in V$ we use as long as $T(v) = w$ for $w \in T$. Additionally, if multiple $v \in V$ exist then we might be able to enforce additional properties that are useful for an application at hand. The pseudo inverse provides such a relaxed notion of the inverse.[3]

[2] In practice, only very rarely the inverse T^{-1} is needed explicitly but one typically seeks for a given $w \in W$ a $v \in V$ such that $w = T(v)$. Then v should be determined by solving the linear system associated with $w = T(v)$.

[3] In the literature sometimes also the name 'generalized inverse' is used instead of 'pseudo inverse'.

Definition 1.5. *Let V, W be linear spaces and $T : V \to W$ be a linear map. A **left pseudo inverse** T_L^{-1} of T is a linear map $T_L^{-1} : W \to V$ such that*

$$T_L^{-1} T = \mathrm{Id}.$$

*A **right pseudo inverse** T_R^{-1} of T is a linear map $T_R^{-1} : W \to V$ such that*

$$T T_R^{-1} = \mathrm{Id}.$$

When T has an inverse in the sense of Proposition 1.1 then the left and right pseudo inverse coincide and they are equal to the inverse T^{-1}.

Note that in contrast to the inverse the pseudo inverse is in general not unique. For example two different right pseudo inverses T_R^{-1} and $\bar{T}_R^{-1}$ can yield $v = T_L^{-1}(w)$ and $\bar{v} = \bar{T}_L^{-1}(w)$ and both are valid as long as $T(T_L^{-1}(w)) = T(\bar{T}_L^{-1}(w)) = w$.

Example 6. Let V, W be Hilbert spaces and $T : V \to W$. Then the **Moore-Penrose pseudo inverse** T^+ of T is the linear operator $T^+ : W \to V$ such that $v = T^+(w)$ is the solution to $w = T(v)$ that has minimum L_2 norm.

We will see in Proposition 1.8 that the Moore-Penrose pseudo inverse can also efficiently be determined numerically. Alternative characterizations of the Moore-Penrose pseudo inverse exists. Those in Example 6 emphasizes one of the properties of particular practical importance since it yields a vector v that is most "regular".[4]

When an inner product is available, then next to the inverse another linear map is associated with a linear operator T.

Proposition 1.2. *Let V, W be Hilbert spaces and $T : V \to W$ a linear map. Then the **adjoint** T^* of T is a linear map $T^* : W \to V$ such that*

$$\langle T(v), w \rangle_W = \langle v, T^*(w) \rangle_V$$

*for all $v \in V$ and $w \in W$. A linear map $A : V \to V$ is **self-adjoint** when*

$$\langle T(u), v \rangle = \langle u, T(v) \rangle$$

for all $u, v \in V$.

Similar to the inverse, the adjoint T^* to a linear map T is a map in the opposite direction from W to V. However, for the adjoint only an equality in terms of the inner product, for example measurements, has to hold. Adjoint operators play an important role in physics, and in computer graphics they have found applications for example for light transport.[5]

[4] The notion of L_2 regularity that we used here is a weak. Nonetheless, as we will see in the following it plays a central in linear least squares problems.

[5] Christensen, "Adjoints and Importance in Rendering: an Overview".

Example 7. Let $A_t : V \to V$ be the linear operator describing the time evolution of a system by $f_t = A_t f_0$ where f_t is the system state at time t and f_0 the state at the initial time $t = 0$. For example, the system can be a pendulum for small time t, in which case A_t is a time dependent rotation matrix that determines the rotation of the pendulum, or a fluid with f_t being the fluid density at time t, in which case A_t is infinite dimensional and will be discussed in more detail in the next chapter. A measurement M of such a dynamical system, yielding a value that in principle can be determined using a physical measurement device, is then usually given by $M_t = \langle f_t, m \rangle$. We hence have

$$M_t = \langle f_t, m \rangle \tag{1.10a}$$
$$= \langle A_t f_0, m \rangle \tag{1.10b}$$
$$= \langle f_0, A_t^* m \rangle \tag{1.10c}$$
$$= \langle f_0, A_{-t} m \rangle \,. \tag{1.10d}$$

Can one work this out for pendulum or rotation?

The last equality can be seen as a definition of A_{-t} but for most physical systems one can indeed show that A_t^* is obtained by inverting the sign for time, an immediate consequence of time reversibility.

Numerically, again the coordinate representation of the adjoint is needed. Let V, W be Hilbert spaces and $T : V \to W$ be a linear map. From our previous discussion we already know that it suffices to study the effect of T on basis vector. Hence, let $\{e_i\}_{i=1}^n$ be a basis for V and $\{f_j\}_{j=1}^m$ a basis for W. Then

$$\langle Te_i, f_j \rangle_W = \langle e_i, T^* f_j \rangle_V \tag{1.11a}$$

and by introducing the matrix representations, analogous to Eq. 1.8, we have

$$T_{ji} = T_{ij}^* . \tag{1.11b}$$

Hence, the coordinate representation of the adjoint T^* is given by swapping the indices of T, that is by the transpose T^T of T, providing a practical means to obtain T^* numerically. The coordinate representation of a self-adjoint linear map is hence given by a symmetric matrix.

A self-adjoint operator enforces equality of the inner product when it is applied to one argument. A linear map that preserves the inner product when applied to both arguments is orthogonal.

Definition 1.6. *Let V be a Hilbert space and $T : V \to V$ be a linear operator. Then T is **unitary** or **orthogonal** when for all $v \in V$:*

$$\langle T(u), T(v) \rangle = \langle u, v \rangle \,.$$

In the infinite dimensional case or when the linear space V is defined over the complex numbers then a map satisfying the above condition is usually denoted as unitary, in the real and finite dimensional case it is called orthogonal.

Proposition 1.3. *Let V be a finite dimensional Hilbert space. The coordinate representation of an orthogonal linear map $T : V \to V$ is an* ***orthogonal matrix*** *$\bar{T}$ satisfying*

$$\bar{T}\bar{T}^{-1} = \bar{T}^{-1}\bar{T} = \mathrm{Id}.$$

Exercise 2. *Show that the above proposition holds.*

We have by definition of an orthogonal operator for an orthonormal basis $\{e_i\}_{i=1}^n$ for V that

$$\langle T(e_i), T(e_j)\rangle = \langle e_i, e_j\rangle \tag{1.12a}$$

and by the orthonormality of the basis this equals

$$\langle T(e_i), T(e_j)\rangle = \delta_{ij}. \tag{1.12b}$$

Writing the operators and the inner product in their coordinate representation we have

$$\sum_{a=1}^{n}\left(\sum_{b=1}^{n} T_{ab}\, e_b^i\right)\left(\sum_{c=1}^{n} T_{ac}\, e_c^j\right) = \delta_{ij} \tag{1.12c}$$

and taking the adjoint of T_{ab} and reordering the summations yields

$$\sum_{b=1}^{n}\sum_{c=1}^{n} e_b^i\, e_c^j\left(\sum_{a=1}^{n} T_{ba}\, T_{ac}\right) = \delta_{ij}. \tag{1.12d}$$

Since the last equation has to hold for all tupled (i, j) we have to have

$$\sum_{a=1}^{n} T_{ab}\, T_{ac} = T^T \cdot T = \mathrm{Id}. \tag{1.13}$$

Since V is finite dimensional $T^T \cdot T = \mathrm{Id}$ also implies $T \cdot T^T = \mathrm{Id}$.

Proposition 1.4. *An orthogonal matrix T has rows and columns that are orthonormal as vectors. For example, for the columns it holds*

$$\sum_{a=1}^{n} T_{ac}\, T_{bc} = \delta_{ab}. \tag{1.14}$$

Exercise 3. *Show that the above proposition holds.*

An alternative perspective on orthogonal matrices is as rotations. Indeed, we have both in $\mathbb{R}^2$ and $\mathbb{R}^3$ that rotation matrices, for example

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{1.15}$$

are orthogonal by the Pythagorean trigonometric identity.[6] The above definition using either the orthonormality of columns or that the inverse is given by the transpose are those that work for arbitrary dimensions, whereas formulas in terms of the rotation angles are only available in two or three dimensions. We will see yet another perspective on orthogonal matrices in the next section.

A classical example for linear maps, and one that is of considerable importance in practice, are change-of-basis maps. In fact, their matrix representation are the bases matrices we already encountered in the foregoing. It might seem that it does not make sense to abstractly talk about such maps without a coordinate representation but we will see that this is not the case in the following in Sec. 1.5.2.

Proposition 1.5. *Let V be a Hilbert space with orthonormal basis $\{e_i\}_{i=1}^n$, and let $\{f_i\}_{i=1}^m$ be an arbitrary frame for V. Then the* ***basis matrix*** *$B_{e_i}(f_j)$ of $\{f_i\}_{i=1}^m$ with respect to $\{e_i\}_{i=1}^n$ given by*

$$B_e(f) = \begin{pmatrix} \langle f_1, e_1\rangle & \cdots & \langle f_1, e_1\rangle \\ \vdots & \ddots & \vdots \\ \langle f_m, e_n\rangle & \cdots & \langle f_m, e_n\rangle \end{pmatrix} \in \mathbb{R}^{m\times n}.$$

provides the change-of-basis from $\{e_i\}_{i=1}^n$ to $\{f_i\}_{i=1}^m$ so that

$$\begin{pmatrix} v(f)_1 \\ \vdots \\ v(f)_m \end{pmatrix} = B_e(f) \begin{pmatrix} v(e)_1 \\ \vdots \\ v(e)_n \end{pmatrix}$$

where the $v(e)_i$ are the coefficients of $v \in V$ with respect to $\{e_i\}_{i=1}^n$ and the $v(f)_i$ are those with respect to $\{f_i\}_{i=1}^m$.

We have already derived part of this result before and the reader should recall how the matrix B can be derived. Proposition 1.5 shows that the basis matrix provides on the one hand the numerical representation of the basis, since every row is the basis expansion of one of the basis functions f_i, and at the same time a way to determine the expansion function coefficients for the basis or frame. This viewpoint further strengthened by the next result.

Proposition 1.6. *Under the assumptions of the foregoing proposition, the columns of a right pseudo inverse B_R^{-1} contain the basis expansion of the dual frame functions $\tilde{f}_i$ with respect to $\{e_i\}_{i=1}^n$ so that*

$$v = \sum_{i=1}^m \left\langle v, \tilde{f}_i \right\rangle f_i.$$

[6] By Euler's theorem, which states that every rotation in R^3 can be expressed as a rotation around a suitably chosen axis, it suffices to consider the two-dimensional case.

Note that the role of the $\tilde{f}_i$ is fixed when the right pseudo inverse B_R^{-1} is used to obtain them. When the f_i are used for analysis to obtain the coefficients v_i, that is

$$v = \sum_{i=1}^{m} \langle v, f_i \rangle \, \tilde{f}_i \tag{1.16}$$

then a left pseudo inverse B_L^{-1} has to be employed to obtain the $\tilde{f}_i$. Also recall from Proposition 1.1 that in the case B has a regular inverse then the pseudo inverse coincides with the inverse. This is the case when $\{f_i\}_{i=1}^m$ forms a basis and $m = n$ and the $\tilde{f}_i$ are then the biorthogonal dual basis functions. Since then $B^{-1} B = B B^{-1} = \mathrm{Id}$ for a biorthogonal basis the primary and dual functions can both be used for projection and reconstruction.

This concludes our discussion of the fundamental properties of linear operators. Such operators will play a central role in all subsequent chapters.

Homework 3. *Show that the space of linear maps $T : V \to W$ from a linear space V to a linear space W has itself the structure of a vector space. Begin by developing the linear structure for the space of $m \times n$ matrices.*

1.5.2 Eigen and Singular Value Decomposition

Operators and their representation as matrix are often abstract and unintuitive. The eigen and singular value decomposition provide important tools to analyze and understand operators. Let us begin by recalling what we mean by an eigenvector and an eigenvalue.

Definition 1.7. *Let V be a Hilbert space space and $T : V \to V$ a linear map on V. An **eigenvector** v of T satisfies*

$$Tv = \lambda v$$

*and λ is an **eigenvalue** of T.*

Eigenvectors and eigenvalues are also only defined for a linear map $T : V \to V$ that maps a linear space into itself. A generalization that is defined also when $T : W \to V$ is the singular value decomposition that will be introduced at the end of the section. Note that v does not have to be unique, even when we identify linear dependent vectors, and then a nontrivial subspace of V is associated with the eigenvalue λ. Typically one assumes that eigenvectors are normalized, that is $\|v\| = 1$, and then the only degree of freedom that is left is the sign of the eigenvalue, that is we have (λ, v) and $(-\lambda, -v)$ both represent the same eigenvalue-eigenvector pair.

Example 8. Let R be a rotation in $\mathbb{R}^3$. Then R has an eigenvalue $\lambda = 1$ so that $Rv = v$. The eigenvector v that is preserved under R is the rotation axis

and we see that any rotation in $\mathbb{R}^3$ can be considered as a two-dimensional rotation around a fixed axis. The result is known as Euler's theorem.[7]

Draw figure

Example 9. In Definition 1.4 we introduced the kernel of a linear map. By Definition 1.7 we can characterize the kernel of a map T as the subspace of V associated with the eigenvalue $\lambda = 0$. This characterization is computationally important since it enables to numerically determine the kernel of a linear map in its coordinate representation.[8] By Theorem 1.3 below, the eigenvectors associated with kernel thereby have to be chosen such that they are orthogonal to all remaining eigenvectors. In the finite dimensional case, this uniquely fixes the eigenvectors that span the kernel.

As a concrete example let us consider the projection P_x onto the x-axis in $\mathbb{R}^2$:

Draw figure for x-projection in $\mathbb{R}^2$.

The matrix representation of P_x is given by

$$P_x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{1.17}$$

and it is immediately apparent that the eigenvalues and eigenvectors are:

$$\begin{aligned} \lambda_1 = 1 \quad v_1 = (1.0, 0.0)^T \\ \lambda_2 = 0 \quad v_2 = (0.0, 1.0)^T \end{aligned} \quad . \tag{1.18}$$

It holds for arbitrary projection operators that the eigenvalues are all either zero or one. The example of P_x is a special case of a diagonal matrix which we will consider again in Example 10.

Theorem 1.3 (Spectral Theorem for Self-Adjoint Operators)**.** *Let H be a Hilbert space and T a compact, self-adjoint operator. Then the eigenvectors v_i of T provide an orthonormal basis for V and the eigenvalues of T are real.*

Obviously, the above result also holds in the finite dimensional case. T is then self-adjoint when it is symmetric and compactness is always satisfied. When T is not symmetric one will obtain complex-valued eigenvalues. In the infinite dimensional case various generalizations beyond the case of compact operators exist but since their treatment would require considerable additional technical machinery we restrict ourselves to the above result and exclusively consider the finite dimensional case in the following. However, we will consider a special instance of the infinite dimensional case in the next chapter.

[7] For a proof see for example (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, Chapter 9.2).

[8] Strictly speaking, due to the limited precision of computers, only the numerical kernel can be determined, say up to floating point precision.

Example 10. Let D be a diagonal matrix, that is a square matrix satisfying

$$D = \begin{pmatrix} d_1 & 0 & \cdots & & \\ 0 & d_2 & 0 & \cdots & \\ & \ddots & \ddots & \ddots & \\ & \cdots & 0 & d_{n-1} & 0 \\ & & \cdots & 0 & d_n \end{pmatrix} \tag{1.19}$$

Then the k^{th} eigenvalues and the associated eigenvector are given by

$$\lambda_k = d_k \qquad v_k = (0, \cdots, 1, \cdots, 0)^T \tag{1.20}$$

where the k^{th} element of v_k is non-zero.

Example 11. Let C be a circulant matrix, that is a square matrix satisfying

$$C = \begin{pmatrix} c_0 & c_{n-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & \cdots & c_3 & c_2 \\ \vdots & & \ddots & & \vdots \\ c_{n-2} & c_{n-3} & \cdots & c_0 & c_{n-1} \\ c_{n-1} & c_{n-2} & \cdots & c_1 & c_0 \end{pmatrix} \tag{1.21}$$

The eigenvalues and eigenvectors of C are complex since C is not symmetric. Through the special structure of C the eigenvectors and eigenvalues have an analytic form. The eigenvectors are given by

$$v_k = (1, \omega_k, \omega_k^2, \cdots, \omega_k^{n-1}) \quad , \quad \omega_k = e^{2\pi i k/n} \tag{1.22}$$

where i is the imaginary unit with $i = \sqrt{-1}$. The corresponding eigenvalues are

$$\lambda_k = c_0 + c_{n-1}\omega_k + \cdots + c_1\omega_k^{n-1}. \tag{1.23}$$

The v_i are related to discrete Fourier basis functions, which arise for example in the discrete Fourier transform. Their connection to the harmonics can be seen by Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$ so that

$$\omega_k = e^{2\pi i k/n} = \cos(2\pi i\, k/n) + i\sin(2\pi i\, k/n) \tag{1.24a}$$

$$= \cos(2\pi i\, \lambda_k) + i\sin(2\pi i\, \lambda_k) \tag{1.24b}$$

with the λ_k being equidistant samples on $[0, 1]$. Hence, the v_k are the discrete Fourier transform functions sampled at λ_k.

figure

This also implies that the v_k are orthogonal. In the following chapter, we will also see the infinite dimensional analogue of this example.

Except for special cases, we have to compute the eigenvalues and eigenvectors numerically using tools from numerical linear algebra,[9] a subject that goes beyond the present notes.

[9] For an introduction see for example (Golub and Van Loan, *Matrix Computations*).

Remark 1. An alternative perspective on the eigen decomposition of a matrix is to consider it as a factorization of the form

$$A = USU^T \tag{1.25a}$$

where U is the orthogonal matrix whose columns are the eigenvectors and S is diagonal with the eigenvalues being the nonzero entries, that is

$$A = \begin{pmatrix} & & \\ v_1 & \cdots & v_n \\ & & \end{pmatrix} \begin{pmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & 0 \\ & & \ddots & & \\ & 0 & & \lambda_{n-1} & \\ & & & & \lambda_n \end{pmatrix} \begin{pmatrix} & v_1 & \\ & \vdots & \\ & v_n & \end{pmatrix}. \tag{1.25b}$$

Since U is orthogonal its columns form an orthonormal basis, cf. Proposition 1.4, and U^T is the change of basis matrix into this basis. By Eq. 1.25a, the action $A(v)$ of A applied to v can also be understood as first transforming v using U^T into a basis where A is diagonal, then applying A in its diagonalized form by S, and then transforming back to the original basis so that the result of $(USU^T)(v)$ indeed equals $A(v)$. The eigen decomposition is hence also often denoted as the **diagonalization** of an operator.

Exercise 4. *Consider the linear map*

$$A = \begin{pmatrix} 1.6250 & 0.6495 \\ 0.6495 & 0.8750 \end{pmatrix}. \tag{1.26}$$

Determine and interpret its eigen decomposition.

Since A is symmetric it has a real eigen decomposition given by

$$A = USU^T = \begin{pmatrix} 0.87 & -0.50 \\ 0.50 & 0.87 \end{pmatrix} \begin{pmatrix} 2.0 & 0.0 \\ 0.0 & 0.5 \end{pmatrix} \begin{pmatrix} 0.87 & 0.50 \\ -0.50 & 0.87 \end{pmatrix}. \tag{1.27}$$

By construction, the change of basis matrix U is given by

$$U = R(30°) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{1.28}$$

Hence, $A(v)$ corresponds to a change of basis to a coordinate basis that is rotated by 30° with respect to the standard basis for $\mathbb{R}^2$, then a scaling in the rotated coordinate system, and finally a change of basis back to the standard basis for $\mathbb{R}^2$.

figure: draw the effect of A with a sketch of the rotated coordinate system

The eigen decomposition exists only when $T : V \to V$ is a linear map from a space onto itself. A generalization that is well defined also when $T : V \to W$ is the singular value decomposition.

Definition 1.8. *Let V, W be finite linear spaces and $T : V \to W$ be a linear map. Then $u \in W$ is a **left singular vector** and $v \in V$ a **right singular vector** of T when*

$$T\,v = \sigma\,u$$

*and σ is the **singular value** associated to (u, v).*

The singular value decomposition also exists in the infinite dimensional case but we will not consider it here.[10] The importance of the singular vectors and values stems from the following proposition.

Proposition 1.7. *Let V, W be Hilbert spaces of dimension n and m, respectively, and let $T : V \to W$ be a linear map. In matrix form the singular value decomposition is given by*

$$T = U\Sigma V^T.$$

where the columns of U are formed by left singular vectors u_i and the columns of V by right singular vectors v_i, and Σ is a quasi-diagonal matrix whose nonzero entries are the singular values. Moreover, the left singular vectors $\{u_i\}_{i=1}^n$ form an orthonormal basis for W and the right singular vectors $\{v_i\}_{i=1}^m$ an orthonormal basis for V.

It is customary to order the singular values in non-decreasing order so that $\sigma_i \geq \sigma_{i+1}$ and we will follow this convention in the following. It follows from the definition that only for T and $m \times n$ matrix there are at most $\min(m, n)$ nonzero singular values.

Remark 2. For a rectangular matrix $T \in \mathbb{R}^{m\times n}$ the matrix Σ is only quasi-diagonal. For $m < n$ this means

$$\Sigma = \begin{pmatrix} \sigma_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_m & 0 & \cdots & 0 \end{pmatrix} \tag{1.29}$$

and the matrix has the form of the $m \times m$ identity matrix with an $m \times (n - m)$ block that is a zero matrix adjoining on the right. Conversely, when $m > n$ then Σ has the form

$$\Sigma = \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}. \tag{1.30}$$

[10] For the infinite dimensional case see for example (Stakgold and Holst, *Green's Functions and Boundary Value Problems*) or Lax (Lax, *Functional Analysis*, Chapter 30).

Remark 3. The eigen decomposition is a special case of the singular value decomposition. Because of the significance for quantum mechanics, the eigen decomposition has been developed earlier and in more detail.

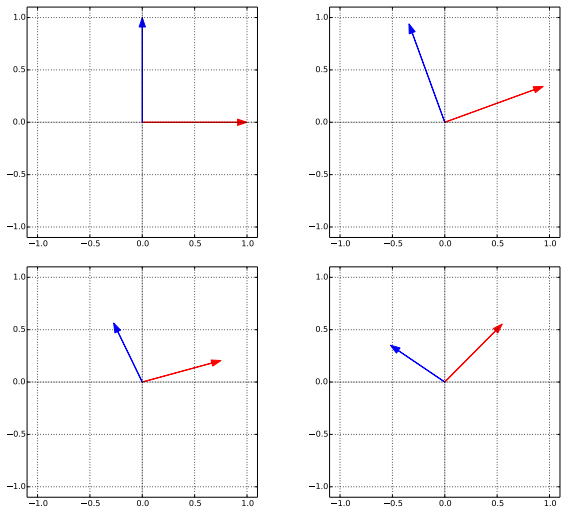
Example 12. Consider the linear map

$$A = \begin{pmatrix} 0.55 & -0.52 \\ 0.55 & 0.35 \end{pmatrix}. \tag{1.31}$$

Its singular value decomposition is given by

$$A = U\Sigma V^T = \begin{pmatrix} 0.87 & -0.50 \\ 0.50 & 0.87 \end{pmatrix} \begin{pmatrix} 0.80 & 0.0 \\ 0.0 & 0.60 \end{pmatrix} \begin{pmatrix} 0.94 & 0.34 \\ -0.34 & 0.94 \end{pmatrix} \tag{1.32}$$

The effect of the transformation, sequentially by first applying V^T, then Σ, and then U, one the canonical basis (e_1, e_2) is shown below (from left to right, top to bottom):



Analogous to the eigen decomposition, we can also interpret the singular value decomposition as a diagonalization of an operator. Since the domain and target of the operator do not coincide, two bases are required for this diagonalization, and this are the bases formed by the left and right eigenvectors. In the above example, we can interpret the singular value decomposition as representing A as a rotation by V^T, then a scaling by Σ along the rotated axes, followed by another rotation by U.

As an example for the manifold applications of the singular value decomposition we return to the pseudo inverse introduced in Definition 1.5 and Example 6.

Proposition 1.8. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map with singular value decomposition $T = U\Sigma V^T$. Then the Moore-Penrose pseudo-inverse T^+ of is given by*

$$T^+ = V\,\Sigma^+\,U^T \tag{1.33}$$

where Σ^+ is the diagonal matrix whose i^{th} diagonal element is given by $1/\sigma_i$, that is Σ^+ is obtained by inverting the diagonal entries.

The above proposition provides a practical means to compute the Moore-Penrose pseudo inverse. Note that as a special case one also obtains means to compute the inverse. This technique is however more expensive than using state-of-the-art numerical techniques.[11]

Exercise 5. *Show that the inverse of a square diagonal matrix*

$$D = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix} \tag{1.34}$$

is given by

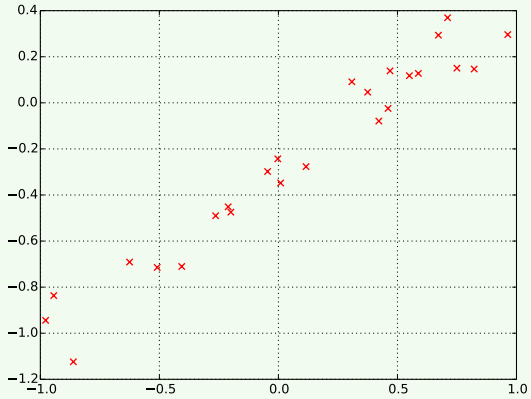
$$D^{-1} = \begin{pmatrix} 1/d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/d_n \end{pmatrix}. \tag{1.35}$$

Exercise 6. *As an application of the pseudo inverse we will consider linear least squares. The problem is to estimate the parameters of a function, for example of a polynomial*

$$p^k(x) = \sum_{i=0}^{k} a_i x^k \tag{1.36}$$

from m noisy measurements (x_i, y_i), such as

[11] As remarked before, in almost all applications one should not explicitly compute the matrix inverse but instead solve the associated linear system.



For the above data a reasonable model seems to be a first order polynomial, that is $p(x) = a_0 + a_1\, x$. *However, the data clearly does not perfectly lie on a line, since the measurements were contaminated with noise, and we also have far too many measurements to directly determine the two parameters* a_0 *and* a_1. *Clearly, we would like to use the fact that we have* $m >> 2$ *measurements to make our estimate robust and average out the contribution of the noise.*

The measurements should all satisfy the linear equation. Hence we have

$$y_1 = a_0 + a_1 x_1$$

$$y_2 = a_0 + a_1 x_2$$

$$\dots$$

$$y_m = a_0 + a_1 x_m$$

Since the a_0,a_1 *are identical for all equations we can write equivalently as matrix-vector equation*

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1.0 & x_1 \\ \vdots & \vdots \\ 1.0 & x_m \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = Xa. \tag{1.37}$$

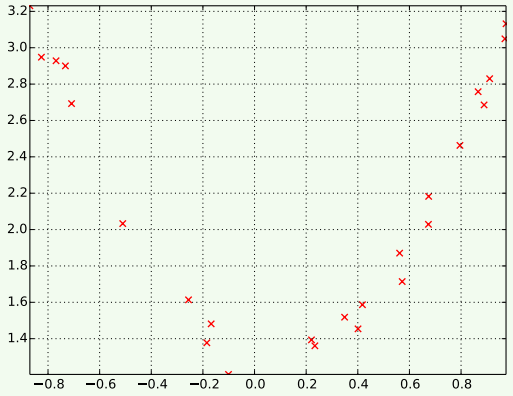
The above equation is an overdetermined linear system. Hence, a solution can be determined using a pseudo inverse as $a = P_L^{-1} y$. *But which pseudo inverse should be used? One can show that the Moore-Penrose, which can be determined using the singular value decomposition, yields the solution* $\bar{a}$ *that minimizes the quadratic error*[12]

$$E(a_0, a_1) = \sum_{i=1}^{n} \left\| y_i - p_{a_0,a_1}(x_i) \right\|^2 . \tag{1.38}$$

One can show that the minimizer can also be determined without the pseudo inverse by solving the normal equation

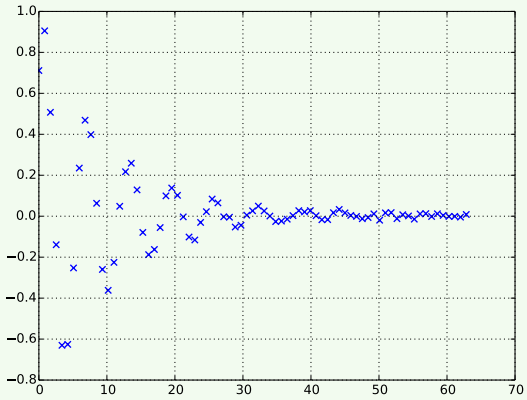
$$X^T y = X^T X a. \tag{1.39}$$

Now let us assume that we are given are the following measurements:



Choose a suitable polynomial to approximate the data and compute the solution of the least squares problem both using the pseudo inverse, determined from the pseudo inverse, and the normal equation.

What is linear about the above least squares problem? Or equivalently, what least squares problems can be solved efficiently? Linearity means, as in the general definition, that we consider "vectors", here the unknown polynomial of degree k, that are linear combinations of basis vectors, here the elementary polynomials x^i, with weights a_i. This is a first example of a vector space whose elements are functions. We will study such spaces in much greater detail in the next section. An example of a set of measurements where linear least squares is no longer sufficient is:



The data are measurements from a damped wave phenomenon

$$u(t) = \hat{u}\, e^{-\delta t} \sin(\omega\, t + \hat{\omega}) \tag{1.40}$$

with the harmonic part representing the wave and the exponential decay the linear damping. In Eq. 1.40 $\hat{u}$ is the amplitude of the wave, δ is the damping coefficient, $\omega = 2\pi f$ is the angular frequency, and $\hat{\omega}$ the phase shift. Determining the unknown parameters $\hat{u}$, δ, ω, and $\hat{\omega}$ from the measurements (u_i, t_i)

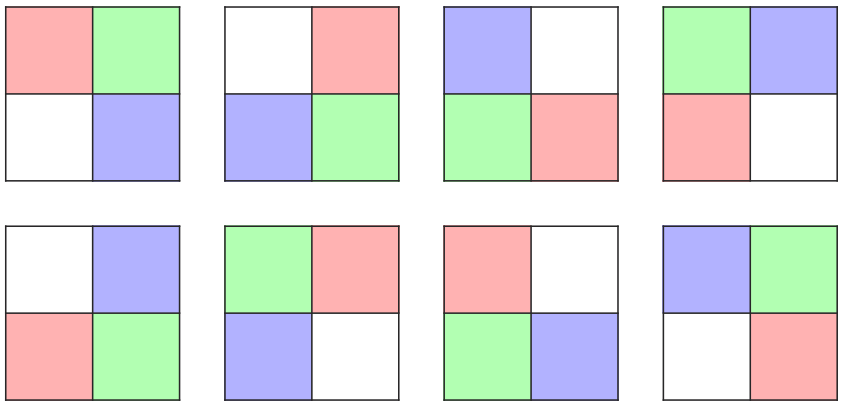


Figure 1.1: A group can be considered as the set of possible transformations of a set such as the possible arrangements, or permutations, of "color tiles" on a 2×2 panel.

such that the quadratic error

$$E(\hat{u}, \delta, \omega, \hat{\omega}) = \sum_{i=1}^{m} \|u_i - u_{\hat{u},\delta,\omega,\hat{\omega}}(t_i)\|^2 \tag{1.41}$$

is minimized is a nonlinear problem since parameters appear both in the exponent of the exponential and inside the sine term. Such a nonlinear least squares problem requires iterative methods such as the Gauss-Newton algorithm for its solution.

Solution: See `linear_least_squares.py`.

Homework 4. *Using the singular value decomposition describe the three cases that exist for the pseudo inverse with respect to the relationship of m and n for an $m \times n$ matrix.*

1.6 Linear Spaces as Groups

Groups and their continuous extension, known as Lie groups, are subjects we will only scratch on in this course. A classical example for Lie groups, and one of the original templates for the concept, are linear spaces. Lie groups will be discussed in Chapter 3.7 but we will already introduce the regular group structures of linear spaces at this point.

Definition 1.9. *A **group** G is a set with a binary **group multiplication***

$$g \circ h = gh : G \times G \to G, \quad g, h \in G$$

that is closed in the set and associative so that $f \circ (g \circ h) = (f \circ g) \circ h$. *The* ***identity element*** e *of a group is the unique element such that*

$$e \circ g = g \circ e = g$$

for all $g \in G$. *For every group element* $g \in G$ *there exists a unique* ***inverse element*** g^{-1} *such that*

$$g \circ g^{-1} = g^{-1} \circ g = e.$$

A group is ***Abelian*** *if group multiplication commutes and* $g \circ h = h \circ g$ *for all* $g, h \in G$.

Example 13. A classical example for a group are permutations, Fig. 1.1. A permutation σ is a bijection $(1, \ldots, k) \to (\sigma(1), \ldots, \sigma(k))$ that rearranges the k elements in its domain. It is often written as

$$\begin{pmatrix} 1 & \cdots & k \\ \sigma(1) & \cdots & \sigma(k) \end{pmatrix}$$

where $\sigma(i)$ denotes the i^{th} element of the permuted set. The set of all such bijections forms the permutation group S_k. Group multiplication for S_k is the sequential application of two re-orderings, the identity element e is the permutation that leaves all elements at their original place, and the inverse σ^{-1} of a permutation σ returns all elements to the position before σ was applied.

Some additional useful notions for permutations are as follows. A permutation σ is a *transposition* when exactly two elements are interchanged; it is *odd* when it consist of an odd number of transpositions, while it is *even* when it can be decomposed into an even number of transpositions, and through the group structure these notions are well defined. The sign $\operatorname{sgn}(\sigma)$ of a permutation is

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \sigma \text{ is even} \\ -1 & \sigma \text{ is odd} \end{cases}$$

Permutations will also play an important role in Chapter 3.3 and Chapter 3.4 for working with tensors and differential forms.

Let us now consider a linear space as a group, cf. Def. 1.1.

Exercise 7. *Show the group structure for a linear space* V.

Solution: The group multiplication of V is given by vector addition and since addition is commutative V is in fact an Abelian group. The identity element e of the group is the zero vector and the inverse of an element $v \in V$ is $-v$ since then $v + (-v) = 0$.

Many more objects we already encountered, such as linear maps, have a group structure, but we will return to the subject later in Chapter 3.7.

When one encounters groups in an unfamiliar context it often useful to recall the permutation group or a linear space, such as Euclidean space. The additional Lie group structure of a linear space corresponds intuitively to the possibility to smoothly change the group elements. For a permutation group such a possibility does not exist.

1.7 Affine Spaces

In the foregoing, we always used $\mathbb{R}^2$, or more generally Euclidean space, as an example of a vector space. However, this model can only represent vectors starting at the origin and not point and the translation of points using vectors. A model for Euclidean space that incorporates both points and vectors is an affine space. Another model, which subsumes the previous ones, follows in Chapter 3.

Definition 1.10. *An **affine space** is a set A together with a vector space V and a map*

$$t : V \times A \to A \,:\, v + a \to b$$

for $a, b \in A$ and $v, w \in V$, that satisfies

i) identity: $0 + a = a,\ \forall a \in A$;

ii) associativity: $(v + w) + a = v + (w + a)$;

iii) uniqueness: $v + a = b$ *is a bijection .*

The last property of the translation map '+' ensures the existence of well defined inverse which is denoted by '$-$' and which allows to combine two elements of A to form a vector, that is $v = a - b$. An affine space is sometimes denoted as a vector space where one forgot the origin since vectors are no longer required to be based on the origin but can start from any point in the plane.

Example 14. Euclidean space $\mathbb{E}^2$ is the two dimensional plane together with the vector space $\mathbb{R}^2$.

It is common to use $\mathbb{R}^n$ for both the vector space and the affine space $\mathbb{E}^n$.

Remark 4. The name 'affine' comes from affine combinations of the form

$$\lambda a + (1 - \lambda)b = o + \lambda(a - o) + (1 - \lambda)(b - o) \tag{1.42}$$

for $a, b \in A$ and $\lambda \in \mathbb{R}$. In an affine space only such combinations are independent of the (in an affine space arbitrary) origin o.

Exercise 8. *For the following example, show that only for affine combinations and not general linear combinations the addition of two points is independent of the origin in $\mathbb{R}^2$. With respect to the usual origin $o = (0.0, 0.0)$, let $a = (1.0, 1.5)$, $b = (-1.4, 1.2)$ and furthermore let a second origin be $\bar{o} = (1.0, 1.0)$. Also we use $\lambda = 0.75$, $\beta = 0.5$.*

Solution: For the affine combination with the origin being o we have

$$\lambda a + (1-\lambda)b = \begin{pmatrix} 0.75 \\ 1.125 \end{pmatrix} + \begin{pmatrix} -0.35 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 0.4 \\ 1.425 \end{pmatrix}. \tag{1.43a}$$

Moving the origin to $\bar{o} = (1.0, 1.0)$ we have $\bar{a} = (0.0, 0.5)$, $\bar{b} = (-2.4, 0.2)$. Then

$$\lambda \bar{a} + (1-\lambda)\bar{b} = \begin{pmatrix} 0.75 \\ 0.375 \end{pmatrix} + \begin{pmatrix} -0.6 \\ 0.05 \end{pmatrix} = \begin{pmatrix} -0.6 \\ 0.425 \end{pmatrix}. \tag{1.43b}$$

And translating the result back from $\bar{o}$ to o we see that the affine combination is indeed independent of the origin. In contrast,

$$\lambda a + \beta b = \begin{pmatrix} 0.0 \\ 1.125 \end{pmatrix} + \begin{pmatrix} -0.7 \\ 0.6 \end{pmatrix} = \begin{pmatrix} 0.05 \\ 1.725 \end{pmatrix} \tag{1.44a}$$

and

$$\lambda \bar{a} + \beta \bar{b} = \begin{pmatrix} 0.0 \\ 0.375 \end{pmatrix} + \begin{pmatrix} -1.2 \\ 0.1 \end{pmatrix} = \begin{pmatrix} -1.2 \\ 0.475 \end{pmatrix}. \tag{1.44b}$$

Translated back to the origin we thus have $(0.2, 1.475)$ which does not equal the result for o.

A central property of affine spaces is that the linear addition of a point by a vector yields again a point in the space. Manifolds, which will be studied in Chapter 3, are spaces that are not closed under linear addition but where nonlinear curves are needed to connect points in the space.

Remark 5. An affine space is an example of a group action, here the action of a linear space considered as a group, on another space or set. We will study this idea in more detail in Chapter 3.7.

TODO: Affine transformations, in affine space translations are defined, which is not case in a vector space; to represent it as a linear map one needs homogeneous coordinates, that is translations are not linear maps with respect to the vector space structure.

1.8 Further Reading

The abstract perspective on linear spaces is discussed for example by Lax.[13] Classical texts on matrix theory are those by Horn and Johnson[14] and by Golub and van Loan.[15]

[13] Lax, *Linear Algebra and Its Applications*.
[14] Horn and Johnson, *Matrix Analysis*.
[15] Golub and Van Loan, *Matrix Computations*.

Chapter 2

Signal Processing and Applied Functional Analysis

In this section, we will study one of the most important class of examples for linear spaces: linear spaces whose elements are continuous functions, so called function spaces. Elements in function spaces that correspond to a quantity or phenomenon in the real world are often called "signals", in particular in engineering and the sciences. In computer graphics, signal processing and applied functional analysis play important roles for example in rendering,[1] for the representation of signals and curves,[2] and for mesh processing.[3]

2.1 Functions as Vectors in a Linear Space

In this chapter, we will consider spaces of functions. Hence, we will begin by making precise what we mean by a function.

Definition 2.1. *Let X be a set. A **function** is a map*

$$f : X \to \mathbb{R}$$

into the real numbers $\mathbb{R}$.

[1] For example in precomputed radiance transfer, cf. (Lehtinen, "A Framework for Precomputed and Captured Light Transport"; Ramamoorthi, "Precomputation-Based Rendering"), for radiosity, e.g. (Zatz, "Galerkin Radiosity: A Higher Order Solution Method for Global Illumination"; Gortler, Schröder, Cohen, and Hanrahan, "Wavelet Radiosity"; Schröder, Gortler, Cohen, and Hanrahan, "Wavelet Projections for Radiosity"), or for the representation and interpolation of light intensity, e.g. (Lehtinen, Zwicker, Turquin, Kontkanen, Durand, Sillion, and Aila, "A Meshless Hierarchical Representation for Light Transport"; Mitchell, "Spectrally Optimal Sampling for Distribution Ray Tracing")

[2] For example, (Schröder and Sweldens, "Spherical Wavelets: Efficiently Representing Functions on the Sphere"; Schröder and Sweldens, "Spherical Wavelets: Texture Processing"; Finkelstein and Salesin, "Multiresolution Curves").

[3] For example (Taubin, "A Signal Processing Approach to Fair Surface Design"; Öztireli, Alexa, and Gross, "Spectral Sampling of Manifolds"; Sorkine, "Laplacian Mesh Processing").

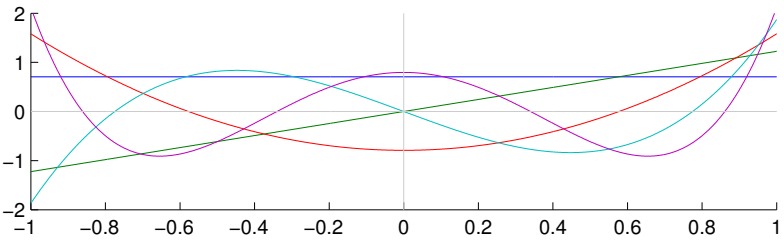


Figure 2.1: Our running example in this section will be the space $\mathcal{H}_4 \equiv \mathcal{H}_{\leq 4}([0,1])$ over $[0,1]$ spanned by the first five Legendre polynomials $P_i(x)$.

For us, a function is hence the special case of a map $A : X \to Y$ whose target (or codomain) is always the real numbers. We could also consider functions as mappings into the complex numbers but since the applications in computer graphics are very limited and it removes (maybe) much of the intuitive understanding we have about functions we will restrict us to the real-valued case. In Definition 2.1 we allowed functions to be defined over arbitrary sets. In this section we will usually consider $X = \mathbb{R}^n$, and for a few example also $X = S^2$, and in Chapter 3 we will study functions over arbitrary manifolds.

Definition 2.2. *A **function space** $F(X)$ is a linear space whose elements are functions $f : X \to \mathbb{R}$ with addition of elements of $F(X)$ defined by pointwise addition*

$$f + g\,|_x = f(x) + g(x)$$

and scalar multiplication by

$$a\,f|_x = a\,f(x)$$

for $f, g \in F(X)$ and $a \in \mathbb{R}$.

We see in the above definition that the linear operations of elements of a function space are reduced to scalar operations on real numbers, cf. Fig. 2.2. The properties that the addition operation in a linear space has to satisfy are hence trivially satisfied, cf. Def. 1.1. In the following, we will show how the different structures that we introduced in general in the last chapter are realized for function spaces. Let us begin by giving some examples of function spaces.

Example 15. The space P_k spanned by all polynomials up to degree k is a function space. As a running example we will consider in this chapter the space $\mathcal{H}_4 \equiv \mathcal{H}_{\leq 4}([-1,1])$ on $[-1,1]$ spanned by the first five Legendre polynomials $P_i(x)$, cf. Fig. 2.2. Explicitly, the P_i are given by

$$P_0(x) = 1 \tag{2.1a}$$

$$P_1(x) = x \tag{2.1b}$$

$$P_2(x) = \frac{1}{2}\left(3x^2 - 1\right) \tag{2.1c}$$

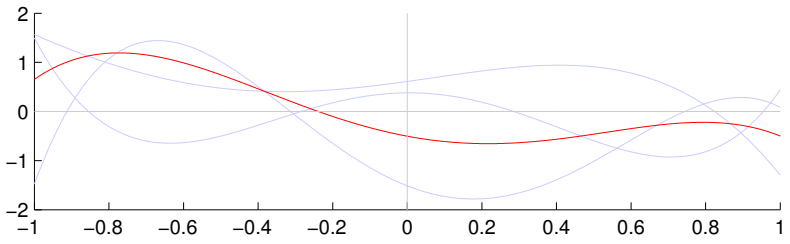


Figure 2.2: A function $f(x)$ that is the linear combination $f(x) = 0.5g_1(x) + 0.6g_2(x) + 0.7g_3(x)$ of three function $g_i(x)$ shown in the background.

$$P_3(x) = \frac{1}{2}\left(5x^3 - 3x\right) \tag{2.1d}$$

$$P_4(x) = \frac{1}{8}\left(35x^4 - 30x^2 + 3\right) \tag{2.1e}$$

and they form a basis for $P_k([-1, 1])$.

Exercise 9. *In physics one often has vector valued functions $\vec{f} : \mathbb{R}^3 \to \mathbb{R}^3$. We will consider applications of these functions and the underlying structure in detail in Sec. 3. Introduce a notion of vector addition for such functions such that one obtains a vector space of vector valued functions.*

Solution: TODO: should follow immediately since in fiber-wise vector addition satisfies the axioms.

Definition 2.3. *Let $X = \mathbb{R}^n$. The* Lebesgue space $L_p(\mathbb{R}^n)$ *for $1 \leq p \leq \infty$ is*

$$L_p(\mathbb{R}^n) = \{\|f\|_p \mid f : \mathbb{R}^n \to \mathbb{R}\}$$

where the L_p norm $\|\cdot\|_p$ for $1 \leq p < \infty$ is defined as

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

and for $p = \infty$ as

$$\|f\|_\infty = \{C \geq 0 \mid |f(x)| \leq C \ , \ \forall x \in \mathbb{R}^n\} \, .$$

The L_p-spaces are Banach spaces.

The L_p spaces above are the continuous analogues of the discrete ℓ_p that we have already encountered in Chapter 1.2.1. The above definition of L_p spaces can be generalized beyond $\mathbb{R}^n$ by either using measure spaces or with the notion of integration that is defined on manifolds. We will study the latter approach in Chapter 3.4.

Remark 6. For us, the L_p-spaces themselves are typically not of particular interest. The functions of interest will usually be in all reasonable function spaces, which, as we will see in the following, is a consequence of our restriction to finite computations. It often also follows from modelling real world phenomena. The more important question for us is what does the L_p norms, or other norms we can consider, measure and do they represent the quantity or behaviour of interest to us. The norms most commonly used thereby are the L_1, L_2, and L_∞ norm. Compared to the L_1 norm, the L_2 norm is more sensitive to regions of extreme values, since squaring amplifies these, but less sensitive to regions with small values, since squaring further diminishes their contribution in the norm. Obviously, which norm is most appropriate will depend on the application.

Definition 2.4. *For $X = \mathbb{R}^n$, the* Sobolev space $W^{k,p}(\mathbb{R}^n)$ *for integers $k \geq 0$ and $1 \leq p < \infty$ is the space*

$$W^{k,p}(X) = \{f \in L_p(X) \mid (D^\alpha f) \in L_p(X) \ , \ \forall |\alpha| \leq k\} . \tag{2.3}$$

Hence, for f to be in a Sobolev space all mixed derivatives D^α whose total order $|\alpha|$ is at most k have to lie in the Lebesgue space $L_p(X)$, where α is a multi-index. When suitably completed, the Sobolev space $W^{k,p}$ forms a Banach space whose norm is given by

$$\|f\|_{k,p} = \left(\sum_{|\alpha| \leq k} \left(\|D^\alpha f\|_p \right)^p \right)^{1/p} \tag{2.4}$$

where the summation is over all mixed derivatives of at most order k. For $p = \infty$ Sobolev spaces are defined analogously to the corresponding Lebesgue spaces.

In contrast to L_p spaces, the norm on Sobolev spaces also takes the derivative into account. It should then come at no surprise that Sobolev spaces play an important role in the theory of partial differential equations. As a last example for a Banach we consider another space that imposes a constraint on the derivative, and hence how wildly a function can vary.

Example 16. Let $X = \mathbb{R}$. The *variation* $V_a^b(f)$ of a function $f : \mathbb{R} \to \mathbb{R}$ is

$$V_a^b(f) = \int_a^b |f'(x)| \, dx. \tag{2.5}$$

The *space of functions with bounded variation* $\mathrm{BV}([a,b])$ is hence defined as

$$\mathrm{BV}([a,b]) = \left\{ f \in L_1([a,b]) \mid V_a^b(f) < \infty \right\} \tag{2.6}$$

and it is a linear subspace of $L_1([a,b])$. Moreover, with the norm

$$\|f\|_{\mathrm{BV}} = \|f\|_1 + V_a^b(f) \tag{2.7}$$

the space $\mathrm{BV}([a,b])$ is a Banach space. In higher dimensions and over more complex domains, the space of functions of bounded variation can be defined using distributional derivatives. It should be noted that the space of functions of bounded variation is not separable.

The space $BV([a,b])$ is used for example in integration theory and we will encounter it again in Chapter 2.4.4. As in the general case, the function spaces that are most useful are Hilbert spaces where an inner product is available. Recall that the Riesz representation theorem then enables to identify functionals $\alpha \in \tilde{\mathcal{H}}$ on the space, which map elements to real numbers and can be interpreted as general measurements, to functions $f \in \mathcal{H}$ in the space with the action being realized through the inner product

$$\alpha(f) = \langle g_\alpha, f \rangle \,. \tag{2.8}$$

As in the discrete case, the space $L_2(\mathbb{R}^n)$ is also in the case of function spaces a Hilbert space.

Example 17. The Lebesgue space $L_2(\mathbb{R}^n)$, cf. Example 2.3, is a Hilbert space with inner product

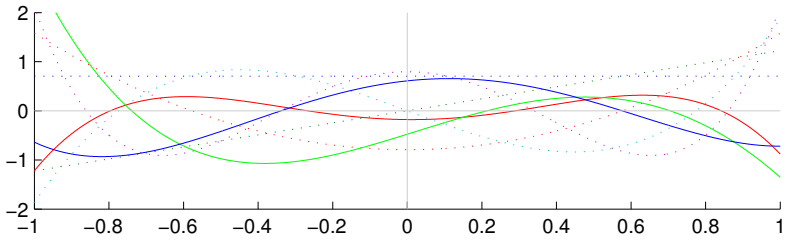
$$\langle f, g \rangle = \int_{R^n} f(x)\, g(x)\, dx. \tag{2.9}$$

The L_2-inner product is by far the most common inner product encountered for function spaces and unless mentioned otherwise we will in the following always assume this inner product for function spaces.

Example 18. Our example space $\mathcal{H}_4$ has inner product spanned by the first five Legendre polynomials is a Hilbert space with L_2-inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\, g(x)\, dx. \tag{2.10}$$

For example, the inner products of the three functions $g_1(x)$ (red), $g_2(x)$ (blue), $g_3(x)$ (green) with the Legendre basis functions (dotted, in the background),



are given by:

	$\langle \cdot, P_0 \rangle$	$\langle \cdot, P_1 \rangle$	$\langle \cdot, P_2 \rangle$	$\langle \cdot, P_3 \rangle$	$\langle \cdot, P_4 \rangle$
$g_1(x)$	-0.0137	0.0624	-0.1602	0.0498	-0.3708
$g_2(x)$	-0.2797	-0.2483	0.4522	-0.9506	0.1003
$g_3(x)$	-0.1566	0.2437	-0.6749	-0.1809	0.2355

Example 19. The Sobolev spaces $W^{k,2}(\mathbb{R}^n)$ modelled on the Hilbert spaces $L_2(X)$ are Hilbert spaces, cf. Example 2.4, and the inner product for the spaces is given by

$$\langle f, g\rangle = \sum_{|\alpha|\leq k} \langle D^\alpha f, D^\alpha g\rangle \tag{2.11}$$

where the summation is over all multi-indices α that have at most order k, and $\langle\,,\rangle$ is the L_2-inner product of Example 17. The spaces area usually denoted as Hilbert-Sobolev spaces $H^s(X) = W^{s,2}(X)$.

Exercise 10. *Let us return to the example of vector valued functions $\vec{f} : \mathbb{R}^3 \to \mathbb{R}^3$. Introduce an inner product for such functions so that an L_2 space can be defined.*

Solution: We have an inner product at each point, by taking the inner product of vectors, and by replacing scalar multiplication by this vector-valued multiplication we obtain:

$$\left\langle \vec{f}, \vec{g} \right\rangle = \int_{\mathbb{R}^n} \vec{f}(x) \cdot \vec{g}(x)\, dx \tag{2.12}$$

It needs to be shown that this inner product in fact satisfies all necessary properties, although this is at least reasonable since the pointwise inner product given by the dot product clearly satisfies them. The L_2-space of vector valued functions is then defined as usual as the space of all $\vec{f}$ that have finite L_2 norm.

In the remainder of the section we will almost exclusively discuss L_2-type Hilbert spaces. These are used most often in practice and their theory most accessible. A discussion of other Hilbert spaces, such as the Sobolev spaces H^2, and Banach spaces would require a more specialized course.

2.2 Bases and Numerical Computations

Since function spaces are special instances of vector spaces, bases and frames provide the principal means for performing numerical computations with continuous functions. Sometimes this is thought to be paradoxical: how can one perform computations with *continuous* functions on a *discrete* computer. The key is that a we want to perform computations on a finite machine. Hence, as long as our function spaces are finite dimensional we will be able to perform numerical computations with them. However, the signals of interest will not always be finite dimensional and hence we have to consider the question which finite dimensional space, or equivalently which basis, enables to approximate the signals. Approximation of functions is a subject we will not be able to discuss in detail but which we will at least introduce in Chapter 2.3.

Exercise 11. *What does it mean for a function to be finite dimensional? Try to develop a precise notion of the concept. Is it sufficient to consider one function?*

Solution: Every function f is an element of a one-dimensional function space: the space spanned by f by af with $a \in \mathbb{R}$. Hence, the notion of dimensionality only makes sense if we consider a family of functions. Typically, this are all the functions that can possibly and reasonably describe a phenomenon of interest. A useful example, and one that has been studied extensively in the literature, is the space of natural images.

2.2.1 Orthonormal Bases

- Repeat definition and most important properties of orthonormal bases.
- For finite dimensional Hilbert spaces $\mathcal{H}$: isomorphism from $\mathcal{H}$ and $\mathbb{R}^n$.
  - Continuous function is equivalent to vector in $\mathbb{R}^n$ with the basis function coefficients providing the coordinates.
  - Since we have an isomorphism, all operations in $\mathcal{H}$ are mapped to a corresponding operation in $\mathbb{R}^n$.
    - Addition.
    - Scalar multiplication.
    - ClassExercise: Inner product and L_2 norm.
    - HomeworkExercise: Parseval's identity.
    - How this applies to operators, which is called Galerkin projection, is discussed in Chapter 2.5.3.
  - ClassExercise: Verify addition for random signals in $\mathcal{H}_4$.
  - Remark: Fourier
    - Fourier series vs. Fourier transform: compactness of domain can have crucial difference.
    - Fourier transform has a continuum of basis functions. Summation has to be replaced by integration but all properties carry over by analogy.
    - Historically, Fourier and polynomials (there's a whole family of orthogonal polynomials) were the only bases, orthonormal or not, that were in general considered as practical. It was only beginning in the 1980s that it was realized that many more bases could be construct and that these had many useful properties, like localization in space, that are not available with classical bases.

Remark 7. We have seen that orthonormal bases provide a convenient and powerful way to work numerically with continuous functions. However, to recover the result we often need the value of the function, at least pointwise for some locations in the domain. This then requires to evaluate the basis functions. The requirement that the basis functions have to be evaluable accurately and efficiently yields a considerable constraint on the number of bases that are numerically practical.

In our examples we employ Legendre polynomials. In principle, we could evaluate the polynomials directly by implementing the formulas in Eq. 2.1 and for $\mathcal{H}_4$ this is indeed a viable option. However, already for H_{10} this approach would suffer from substantial inaccuracies due to the use of floating point numbers.[4] Instead, the recurrence relation

$$(n+1)P_{n+1}(x) = (2n+1)\,x\,P_n(x) - n\,P_{n-1}(x). \tag{2.13}$$

It is useful to keep in mind that even "elementary functions" such as $\sin(x)$ or $\log(x)$ cannot be evaluated directly on a computer but that approximation algorithms are employed that evaluate them up to machine precision (and fortunately these are implemented in hardware so that the computation takes only a few processor cycles).

Remark 8. TODO: Tensor product spaces and based for them. Not complicated but has to be mentioned.

2.2.2 Biorthogonal Bases and Frames

- Repeat definition of biorthogonal bases.
- Frames as overcomplete bases:
- Practical motivation:
  - Orthogonal bases are hard to construct and even if one can construct them they are restrictive. Biorthogonal bases are much easier to construct and they are flexible enough so that one can incorporate other desirable properties (although it is typically not easy to enforce these). What could be such properties? For example, the Legendre polynomials are symmetric (more precisely symmetric and anti-symmetric) with respect to the y-axis. Such symmetries are often desirable since they avoid that one has directional bias in the basis representation.
  - Frames have redundancy which makes them robust against errors and loss of information.

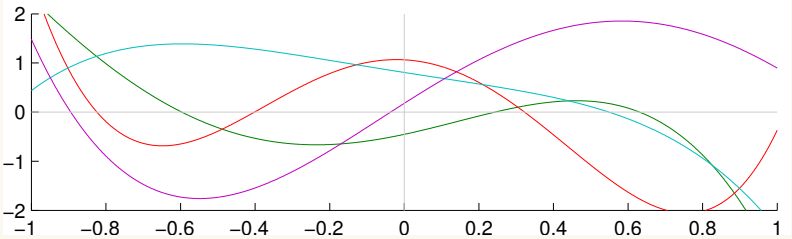
[4] The classical Numerical Recipes book writes on the subject: "Come the (computer) revolution, all persons found guilty of such criminal behavior [the evaluation of a polynomial by directly evaluating $p(x) = a_0 + a_1x + a_2x^2 + \cdots$] will be summarily executed, and their programs won't be!", (Press, Teukolsky, Vetterling, and Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Chapter 5.3).

- Numerical representation:
  - Basis matrix that is formed by basis expansion of $\psi_i(x)$ with respect to some reference basis, that is each row of B contains the basis function coefficients for one function $\psi_i(x)$.
  - The reference basis is necessary for example so that we can evaluate the functions.
  - Dual basis is constructed by inverting the basis matrix. For overcomplete frames the dual frame is not uniquely defined and numerically a pseudo inverse has to be employed to compute it; for example, we can employ the Moore-Penrose pseudo inverse that we discussed before and that can be computed using the SVD.

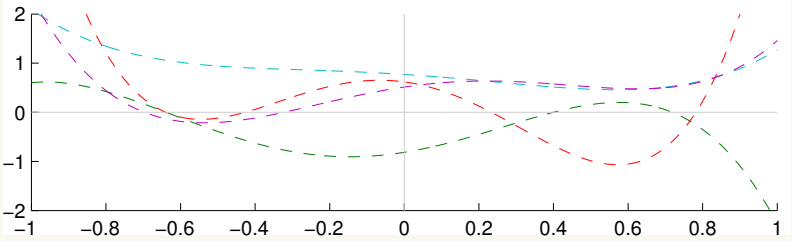
Exercise 12. *Construct a biorthogonal basis for H_4 by generating five random functions in the space. Verify that the vectors indeed span the space. Plot both the primary and dual basis functions.*

For an arbitrary signal given by its basis function coefficients with respect to the P_i determine the error that results when the signal is projected into your biorthogonal basis and then reconstructed. When is the error minimized? Suggest possibilities to improve the quality of your randomly generated biorthogonal basis.

Solution: See `legPlotBiorthoBasis.m`. Five randomly generated signals $\psi_i(x)$ in $\mathcal{H}_4$ are:



The associated dual basis functions $\tilde{\psi}_i(x)$ are given by



Since the biorthogonality condition is satisfied to good accuracy (6.938894×10^{-17}) the reconstruction error for a signal is also small: 1.028951×10^{-15}.

Homework 5. *In Example 1 we introduced the Mercedes Benz frame and in Exercise 1 we generalized it to $\mathbb{R}^3$. When we employ the Cartesian coordinates of the vectors as basis function coefficients for $\mathcal{H}_2$ with respect to the Legendre polynomials P_i then this yields a frame the function space. Is this frame again a tight frame? Plot the signal $f(x)$ corresponding to the expansions coefficients $f_1 = 0.13$, $f_2 = -0.56$, $f_3 = 0.87$.*

Solution: This is again a tight frame since the Legendre polynomials provide an isomorphism from $\mathcal{H}_2$ to $\mathbb{R}^3$.
TODO: Plot signal.

2.3 Approximation of Functions

In the foregoing we assumed that the function $f(x)$ we would like to represent in a basis or a frame lies in the space $\mathcal{F}(X)$ spanned by the representation, that is $g(x) \in \mathcal{F}(X)$. In practice, however, one often has a signal in some space $\overline{\mathcal{F}}(X)$ and wants or needs to represent it in a"smaller" space $\mathcal{F}(X)$. For example, $\overline{\mathcal{F}}(X)$ might be an infinite dimensional space and $\mathcal{F}(X)$ a finite dimensional approximation space spanned by a given basis.

- Distinguish linear vs. nonlinear.
  - Linear: fixed approximation space.
  - Nonlinear: approximation space is determined based on the signal. Under certain conditions this is can be achieved surprisingly effectively.

The distinction between linear and nonlinear approximation is equivalent to a stratification along other directions:

Linear Approximation	Nonlinear Approximation
classical approximation theory	modern approximation theory
Fourier / polynomial bases	wavelet bases
optimal for smooth functions with global regularity	optimal for functions with locally varying regularity and a finite number of discontinuities.

In the best case, one obtains even in the case of locally varying regularity and with singularities the same order or approximation as in the globally smooth case. The key to this result is an adaptation to the local properties of a function so that the approximation only becomes "finer" where it is necessary. As we will see in Chapter 2.3.3 only in $1D$, and to a good extent in $2D$, does this hold and in higher dimensions the question of how to effectively adapt to singularities is still not resolved.

2.3.1 Linear Approximation

- Typically use of Fourier basis and polynomials to approximate functions.
- Optimal for sufficiently smooth function whose smoothness is constant is essentially constant over the domain.
- Approximation by retaining the first k coefficients (if necessary we can reorder basis functions). Error is given by remainder term.

For example for function in the Sovolev spaces $H_2([0,1])$, cf. Example 19, we have the following result.[5]

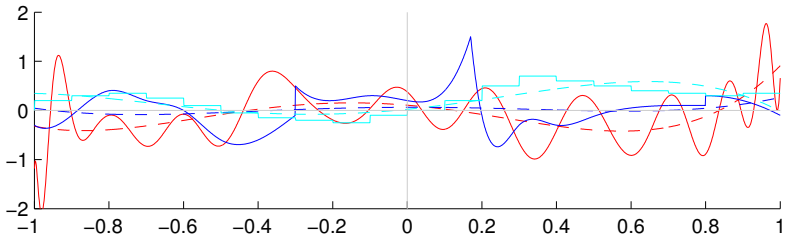
Theorem 2.1. *Let $f \in \mathcal{H}^s([0,1])$ Then the linear N-term approximation error*

$$\epsilon_l(N, f) = \|f - f_N\|^2$$

attained by approximating f in the Fourier basis over $[0,1]$ is $\epsilon_l(N, f) = o(N^{-2s})$ and this rate is asymptotically optimal.

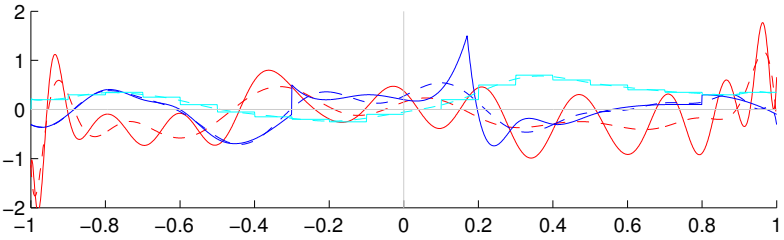
The Fourier basis over $[0,1]$ used in the above theorem is discussed in Sec. 2.5. There we also show how to define Sobolev spaces for non-integer s. Theorem 2.1 shows that the approximation rate increases as functions get smoother and that we have at least a quadratic convergence as long as our function is differentiable. However, the convergence rate becomes only linear when f has singularities. The same approximation rate of $\epsilon_l(N, f) = o(N^{-2s})$ can also be attained using wavelets when these have $q > s$ vanishing moments, see the next section for an introduction to wavelets. Results analogous to Theorem 2.1 can be shown for other global regularity classes.

Example 20. L_2-optimal approximations (dashed) of three test signals (full lines), a smooth signal in H^{30}_{Leg} (red), a smooth signal with singularities (blue), and a piecewise constant signal (cyan), in H^4_{Leg}:



The same signals approximated in the larger space H^{20}_{Leg}:

[5] Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Theorem 9.1 and 9.2.



We see how the approximation becomes better as the number of basis functions, and hence the dimensionality of the approximation space, becomes larger. We also see the different effectiveness of the Legendre polynomials for different signals. The smooth parts of the blue signal and the cyan signal are captured well by the first 20 Legendre polynomials. For neither of the three test signals the basis can capture the high frequency variations.

TODO: Implement approximation of blue signal with three different bases (should be in old course notes).

2.3.2 Nonlinear Approximation

- Derive that L_2 approximation is equivalent to retaining the largest coefficients: selects subspace based on the data and is hence a linear process. Also remainder in sum is error and convergence as $N \to \infty$.
  - Show that there is a linear (projection type) operator for linear approximation but not for nonlinear approximation (or that, in this case, it cannot satisfy the linearity axioms; should follow relatively easily by choosing a suitable example of two functions where completely different subspaces are selected.)
- Remark / Example: wavelets.
  - TODO: Add figures from adaptive wavelet sampling on local tree depth. Any additional figures from Mallat
  - Provide at least basic introduction in 1D.
  - Fast wavelet transform: efficient algorithms.
  - ClassExercise: Derive transform from basis representation. Why can the transform be computed efficiently (Solution: local support of the basis functions, sums are short)? What would be the structure of the basis matrix (sparse with only few nonzero entries)?
- Sparsity of representation and nonlinear approximation.
  - Inherent to nonlinear approximation: one obtains a vector that in some sense has many zeros or at least many very small elements.

  - Moreover, this sparsity reflects the essential signal properties, for example when wavelets are used it shows the spatial location of singularities.
  - Essential difference to linear approximation: in some sense blind to the signal (well, the basis function coefficients are not ... but N-term approximation).

- Language: more words allow to express ideas more succinctly. Similar, one wants to have highly sparse representations it is often useful to start with an overcomplete frame and employ only those basis functions that are useful to model the features of the signal.
  - Another motivation for using frames.
  - The prize that has to be paid is that there are no longer easy and efficient techniques to find the optimal set of basis functions that should be employed and one has to solve an optimization problem.
  - Common numerical techniques are basis pursuit and matching pursuit.[6]

To quantitatively analyze the effectiveness of approximations of nonlinear approximation one needs a mathematical model of functions with local regularity, that is the type of functions where we have seen that nonlinear approximation can be useful. One such model is Lipschitz regularity that can be seen as a generalization of the Taylor series to functions that do not have a classical derivative.

Definition 2.5. *A function $f : \mathbb{R} \to \mathbb{R}$ is **pointwise Lipschitz** $\alpha \geq 0$ at $\bar{x}$ if there exists a $K > 0$ and a polynomial $p_{\bar{x}}(x)$ of degree $m = \lfloor \alpha \rfloor$ such that*[7]

$$|f(x) - p_{\bar{x}}(x)| \leq K|x - \bar{x}|^{\alpha} \quad , \quad \forall x \in \mathbb{R}. \tag{2.14}$$

*A function is $f : \mathbb{R} \to \mathbb{R}$ is **uniformly Lipschitz** α over $[a, b]$ if the above Lipschitz condition is satisfied for all $x \in [a, b]$ for a K independent of x. The **Lipschitz regularity** of f is the supremum of the α such that f is Lipschitz α. The **homogeneous Hölder** α **norm** $\|f\|_{C^{\alpha}}$ of f is the infimum of the K that satisfy the Lipschitz condition in Eq. 2.14 for fixed α.*

The exponent α in the above definition is also known as Hölder exponent. The definition is motivated by the Taylor series for an m times differentiable function where one has

$$|f(x) - p_{\bar{x}}(x)| \leq |x - \bar{x}|^m \left(\frac{1}{m!} \sup_{u \in [\bar{x}-x, \bar{x}+x]} f^m(u) \right) \tag{2.15}$$

[6] See (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Chapter 12) for details.

[7] $\lfloor \alpha \rfloor$ is the largest integer smaller than α.

for all $x \in [\bar{x} - x, \bar{x} + x]$ where $p_{\bar{x}}(x)$ is the Taylor polynomial of order m,

$$p_{\bar{x}}(x) = \sum_{k=0}^{m-1} \frac{f^{(k)}(\bar{x})}{k!}(x - \bar{x})^k. \tag{2.16}$$

When f is m times continuous differentiable then $p_{\bar{x}}(x)$ is the Taylor expansion of f at $\bar{x}$. If $0 \leq \alpha < 1$ then $p_{\bar{x}}(x) = f(\bar{x})$, the "zero-th order Taylor approximation", and the Lipschitz condition becomes

$$|f(x) - f(\bar{x})| \leq K|x - \bar{x}|^\alpha. \tag{2.17}$$

Any $0 \leq \alpha < 1$ characterizes the singularity type at $\bar{x}$ and $\alpha = 0$ corresponds to a bounded but discontinuous function at $\bar{x}$. Using this notion of local regularity we can present a nonlinear analogue of Theorem 2.1:[8]

Theorem 2.2. *Let $f : [0, 1] \to \mathbb{R}$ with K discontinuities and uniform Lipschitz α between the discontinuities with $1/2 < \alpha < q$ then the linear approximation error for wavelet with q vanishing moments is*

$$\epsilon_l(N, f) = O\left(K\|f\|^2_{C^\alpha} M^{-1}\right)$$

and the error using nonlinear wavelet approximation is

$$\epsilon_n(N, f) = O\left(\|f\|^2_{C^\alpha} M^{-2\alpha}\right).$$

The discontinuities in the above definition are isolated points where $\alpha = 0$ and $\|\cdot\|_{C^\alpha}$ is the homogeneous Hölder norm defined in Def. 2.5. Theorem 2.2 shows that the convergence rate of nonlinear approximation is unaffected by the K singularities and one obtains still the same rate as if these were not present. This robustness to singularities is the most important practical advantage of nonlinear approximation.

TODO: Example of nonlinear approximation of local spline signal with singularities, analogous to Mallat (cf. Fig. 9.1 / 9.2).

The problem in practice is to find a "good" model class for a given application.

Remark 9 (Compressed Sensing)**.** In the foregoing we only considered optimal approximations in the L_2 norm. However, an alternative to the L_2 norm that has become very popular recently is the L_1 norm, mainly for applications in compressed sensing.[9] The idea of the method is recover a signal that requires m coefficients f_i in a sparse representation from little more than m "samples" of the form $\langle f, \gamma_i \rangle$ where the γ_i are suitably chosen vectors such that $\langle \gamma_i, \psi_i \rangle$ is nonzero for all ψ_i that are needed to represent f, at least with very high probability. The condition $\langle \gamma_i, \psi_i \rangle \neq 0$ is known as incoherence condition in the

[8] Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Theorem 9.12.

[9] (Candès, Romberg, and Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information"; Donoho, "Compressed Sensing"), see for example (Candès and Wakin, "An Introduction to Compressive Sampling") for an introduction.

compressed sensing literature. Minimizing the L_1 norm simultaneously finds the coefficients that are nonzero and their numerical value. The price to be paid is that one requires an expensive optimization procedure which so far typically outweighs any benefits. The remarkable aspect of compressed sensing is that the optimization is guaranteed to succeed with very high probability. Another application based on the same philosophy of minimizing the L_1 norm is matrix completion from partial observations.[10]

2.3.3 From One to Higher Dimensions

Unfortunately, the approximation of signals becomes much more complex as the dimension increases. Part of the phenomenon is known as *curse of dimensionality*. Additionally the structure of singularities becomes much more involved as the dimension increases.

Remark 10. Dimension of bases space vs. dimension of function space.

Curse of Dimensionality [11]

- Integral: with a tensor product one needs N^k quadrature points in k dimension: exponential dependence on dimension k
- Similar for approximation:

$$C_0 n^{-s/k} \leq \delta_k(\mathbb{B}(H^s))_{L_p(\Omega)} \leq C_1 n^{-s/k} \tag{2.18}$$

  where $\mathbb{B}(H^s)$ is the unit ball in the Hilbert-Sobolev space H^s and $\delta_n(\mathbb{B}(H^s))$, known as Kolmogorov width, corresponds to the nonlinear approximation error $\epsilon_n(N, f)$ we considered before. We see that effective approximations are only possible when $s \approx k$, that is when the smoothness of the functions increases proportional to the dimensionality. This typically unrealistic and we then have again an exponential dependence on the dimensionality.[12]
- For $\mathbb{R}^n$ or $[0,1]^n$ and functions with classical regularity, sparse and multigrid techniques provide a solution for moderate dimensionality up to about five.[13]
  - Finer notion of smoothness that allows for adaptivity.
- Very active area of research currently.

For more details see the survey by Donoho.[14]

[10] Candès and Recht, "Exact Matrix Completion via Convex Optimization".
[11] The term was coined by Bellmann (*Adaptive Control Processes: A Guided Tour*).
[12] See (R. DeVore, *Capturing Functions in High Dimensions*) for details.
[13] Bungartz and Griebel, "Sparse Grids".
[14] Donoho, "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality".

Singularities in Higher Dimensions

- In 2D studied extensively for natural images.
  - Singularities in 1D have one degree of freedom. In 2D singularities can be arbitrary curves. In higher dimensions this phenomenon continues. In nD one can have subsets of dimensions $1 \cdots 1n$.
  - ClassExercise: Singularities for functions $f : \mathbb{R} \to \mathbb{R}^3$. What new types of singularities arise and find an example where these play a role? (Solution: 2D sub-manifolds in $\mathbb{R}^3$, appears for example in elasticity when cracks occur.
- Curvelets, bandlets, and shearlets provide representations that are theoretically optimal, at least asymptotically, but not entirely practical.
  - Also one requires frames to be able to effectively adapt to singularities.
- High dimensional functions, $3D$ and beyond, with singularities largely open research problem.

2.4 Reproducing Kernels

In the last sections we ignored an important question: numerically central are the basis function coefficients, which provide the coordinates of a signal in $\mathbb{R}^n$ that can be used for numerical computations. However, the coefficients are given by

$$f_i = \langle f(x), \tilde{\psi}_i(x) \rangle = \int_X f(x)\, \tilde{\psi}_i(x)\, dx. \tag{2.19}$$

However, the integral is a continuous operation that cannot easily be evaluated on a computer. The only information that can typically easily be obtained in "computer problems" are function values $f(\bar{x})$ for arbitrary points $\bar{x}$ and in "real world problems" measurements of the form $m_\gamma = \langle f, \gamma \rangle$ for a measurement functional γ.[15] For example, the measurement of a pixel sensor on a CCD chip can be described by

$$m = \int_{P_{ij}} \int_{H^2_x} \ell(x, \omega)\, p_{ij}(x, \omega)\, d\omega\, dx \tag{2.20}$$

where P_{ij} is the area of the pixel, H^2_x the hemisphere above a point x on the pixel, $p_{ij}(x, \omega)$ is the pixel response function for x and direction ω, and $\ell(x, \omega)$ the incoming light intensity. By defining an appropriate L_2-inner product we can write the above measurement also as

$$m = \langle \ell, p_{ij} \rangle \tag{2.21}$$

[15] This availability of functionals $\langle f, \gamma_i \rangle$ in many "real world" applications is one of the reasons for the relevance of compressed sensing, cf. Remark 9.

When such measurements are available we can construct the basis matrix to obtain the signal representation in a suitable basis or employ compressed sensing when the signal is sparse and we are willing to perform an expensive optimization, cf. Remark 9.

Unfortunately, in computer graphics one has typically synthetic problems where only pointwise information $f(\bar{x})$ of the function of interest is available. In this section, we will discuss how to effectively work with signals when only function values $f(\bar{x}_i)$ are available.

2.4.1 Point Evaluation Functionals

In Sec. 1.3 we discussed linear functionals on linear spaces V, that is maps $\varphi : V \to \mathbb{R}$ that yield real number $\alpha(v)$ for elements $v \in V$ and that are linear in their arguments so that $\phi(av + bw) = a\,\varphi(v) + b\,\phi(w)$ for $v, w \in V$ and $a, b \in \mathbb{R}$. The space of all such functionals was the dual space of V, and we distinguished between the algebraic dual space $V*$ and the continuous dual space $\tilde{V}$. Also recall that for Hilbert spaces $\mathcal{H}$, the Riesz representation theorem enables to identify the continuous dual space $\tilde{H}$ with $\mathcal{H}$ such that

$$\varphi(v) = \langle w_\varphi, v \rangle \tag{2.22}$$

where $w_\varphi \in \mathcal{H}$ is a unique element in $\mathcal{H}$, cf. Theorem 1.1. We used this for example in the construction of biorthogonal bases where the dual basis function was the realization of the coordinate functionals in $\mathcal{H}$. An important functional on function spaces, and the one of central importance for the present section is the following.

Definition 2.6. *Let $\mathcal{F}(X)$ be a function space defined over a set X. The* ***point evaluation functional*** *$\Upsilon_{\bar{x}}$ on $\mathcal{F}(X)$ at $\bar{x} \in X$ is for all $f \in \mathcal{F}(X)$*

$$\Upsilon_{\bar{x}}[f] = f(\bar{x}).$$

Example 21. Let $C_0^\infty(\mathbb{R})$ be the set of compactly supported *test functions* on the real line. The point evaluation functional on $C_0^\infty(\mathbb{R})$ is the **Dirac delta** $\delta_{\bar{x}}$ so that

$$\Upsilon_{\bar{x}}[f] = \delta_{\bar{x}}[f] = f(\bar{x})$$

for all $f \in C_0^\infty(\mathbb{R})$. The Dirac delta $\delta_{\bar{x}}$ is the prototype for a *generalized function* or *distribution in the sense of Schwartz*. Such objects are useful for example to define weak derivatives for functions that do not have derivatives in the classical sense, such as a step function. The derivatives used in the definition of Sobolev spaces are typically such weak derivatives.

The above result can be extended to L_2 since the space of test functions $C_0^\infty(\mathbb{R})$ is dense in L_p for $1 \leq p < \infty$. This yields a definition of the Dirac delta also for $L_2(\mathbb{R})$ as it is commonly used in the engineering literature.

The above example shows that the point evaluation functional $\Upsilon_{\bar{x}}$ is in general not continuous. However, in many applications and for computations the spaces where $\Upsilon_{\bar{x}}$ is continuous are of particular.

Definition 2.7. *Let $\mathcal{H}(X)$ be a Hilbert space where the point evaluation functional $\Upsilon_{\bar{x}}$ is continuous. Then the* ***reproducing kernel*** *$k_{barx}(x)$ at $\bar{x}$ is the unique function $k_{\bar{x}}(x) \in \mathcal{H}(X)$ such that*

$$\Upsilon_{barx}[f] = \langle f(x), k_{\bar{x}}(x) \rangle \quad , \quad \forall f \in \mathcal{H}(X) \tag{2.23}$$

and $\mathcal{H}(X)$ is a ***reproducing kernel Hilbert space****.*

The remainder of the section will largely be concerned with studying the reproducing kernel Hilbert spaces introduced in the last section. Let us begin with some examples that show that reproducing kernel Hilbert spaces are indeed practically relevant.

Example 22. The Sobolev spaces $H^2(\mathbb{R}^n)$ with $s > n + 1/2$ are reproducing kernel Hilbert spaces.

Example 23. The *Paley-Wiener space* $\Omega_B(\mathbb{R})$ of B-Fourier bandlimited functions for which the Fourier transform is supported only on $[-B, B]$. The reproducing kernel is

$$\mathrm{sinc}_B(\bar{x} - x) = \frac{sin(B\,(\bar{x} - x))}{B\,x}. \tag{2.24}$$

Example 24. Every finite dimensional Hilbert space is a reproducing kernel Hilbert space. The reproducing kernel is given below in Eq. 2.26

The last example is crucial for us since it shows that all function spaces relevant for numerical computations are reproducing kernel Hilbert spaces. Before we continue, let us however consider an alternative perspective on the reproducing kernel. Let $\{\phi_i\}_{i=1}^n$ be an orthonormal basis for $\mathcal{H}(X)$ and $f \in \mathcal{H}(X)$ be an arbitrary function. Then

$$f(\bar{x}) = \sum_{i=1}^{n} f_i\, \phi_i(\bar{x}) \tag{2.25a}$$

$$= \sum_{i=1}^{n} \langle f(x), \phi_i(x) \rangle\, \phi_i(\bar{x}). \tag{2.25b}$$

Using linearity of the inner product and that $f(x)$ is independent of the summation we obtain

$$f(\bar{x}) = \sum_{i=1}^{n} \langle f(x), \phi_i(x)\, \phi_i(\bar{x}) \rangle \tag{2.25c}$$

$$= \left\langle f(x)\,,\, \sum_{i=1}^{n} \phi_i(x)\, \phi_i(\bar{x}) \right\rangle. \tag{2.25d}$$

Comparing Eq. 2.25d to Eq. 2.23 we see that we have to have

$$k_{\bar{x}}(x) = k(\bar{x}, \bar{x}) = \sum_{i=1}^{n} \phi_i(\bar{x})\, \phi_i(x). \tag{2.26}$$

Hence, the reproducing kernel acts by determining the basis function coefficients using the inner product and then reconstructing the function at $\bar{x}$. Eq. 2.26 can also be interpreted as the basis expansion of $k_{\bar{x}}(x) \in \mathcal{H}(X)$ with respect to $\phi_i(x)$ with basis function coefficients $\phi_i(\bar{x})$. The importance of Eq. 2.26 lies in the possibility to easily construct reproducing kernels for a space once an orthonormal basis is known; in practice this is usually not a impediment since we typically start with a space that is specified by a basis.

Homework 6. *Repeat the derivation in Eq. 2.25 for a biorthogonal instead of an orthonormal basis.*

A third perspective that verifies the reproducing property of reproducing k Using the interpretation of Eq. 2.26 as basis expansion of the reproducing kernel, insight into the reproducing property can also obtained by using the equivalence between the continuous inner product and the dot product. This immediately yields

$$\langle f(x), k_\lambda(x) \rangle = f_i^n \cdot \phi_i^n(\lambda) \tag{2.27a}$$

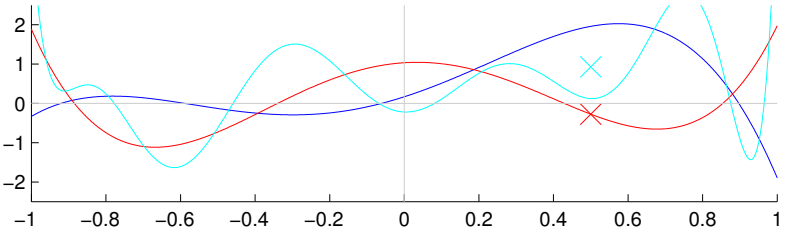
$$= \sum_{i=1}^{n} f_i \, \phi_i(\lambda) \tag{2.27b}$$

The last equation is just the reconstruction equation for $f(x)$ at λ. Hence

$$\langle f(x), k_\lambda(x) \rangle = f(\lambda) \tag{2.27c}$$

which again establishes the reproducing property.

Example 25. Verification of the reconstruction property for a function $f \in \mathcal{H}^4_{\text{Leg}}$ (red) for the reproducing kernel at $\lambda = 0.5$



The y-coordinate for the red cross is determined by Eq. 2.27b. The figure also shows that the reproducing property only hold for functions in the space and that it breaks down for a function $g \in \mathcal{H}^8_{\text{Leg}}$ (cyan).

The above example shows a central aspects of reproducing kernels: the reproducing property only holds for functions in the space $\mathcal{H}(X)$ that is associated with the reproducing kernel $k_{\bar{x}}(x)$. We will not have time to analyse the

error that results from incorrectly assuming that $f \in \mathcal{H}(X)$. However, it can be found in the literature.[16]

Exercise 13. *Show that* $\langle k_y(x), k_z(x) \rangle = k(z, y)$.

Solution: Follows from expanding both functions using Eq. 2.26 and then exploiting the orthonormality of $\{\phi_i\}_{i=1}^n$.

Exercise 14. *Show that the Gaussian*

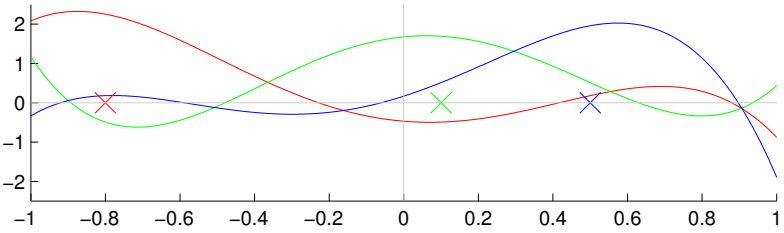
$$g_{\bar{x},\sigma}(x) = e^{-\frac{\|x-\bar{x}\|^2}{2\sigma^2}} \tag{2.28}$$

on $\mathbb{R}$ *is a reproducing kernel.*

Solution: General result is here: http://itb.biologie.hu-berlin.de/~minh/gaussianpaper-minh-2008-final.pdf. In a restricted case it is feasible to show the result.

2.4.2 Reproducing Kernel Bases

- Crucial observation: reproducing kernels at different points are different functions.

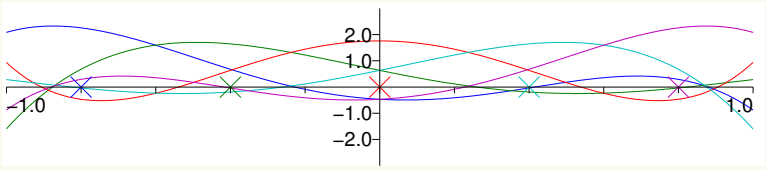


- Picking different points in X is equivalent to picking different functions in the space.
- As we have seen before, in many cases picking n functions will yield a spanning set, and hence a biorthogonal basis for the space.
- Introduce kernel and sampling matrices.
- Recovering basis representation from samples as a change of basis.

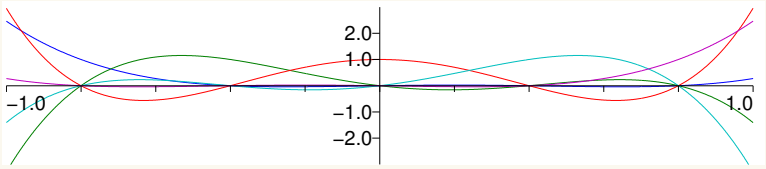
Exercise 15. *Construct a biorthogonal kernel basis for* H_4 *using random locations* λ_i *in* $[-1, 1]$. *Plot the primary and dual basis functions. Verify that you can reconstruct a signal from only knowing its value at the points* λ_i.

[16] Lessig, "Modern Foundations of Light Transport Simulation", Chapter 4.2.

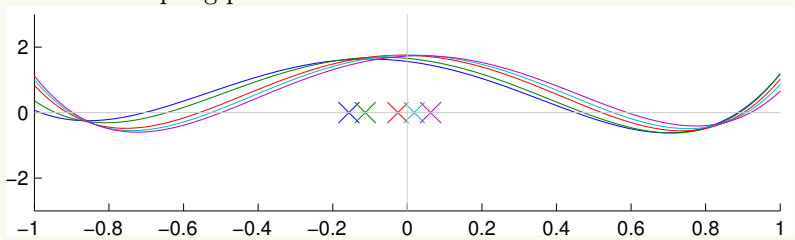
Solution: See `legPlotRKBasis.m` For well distributed sampling locations the reproducing kernel basis functions are given by:



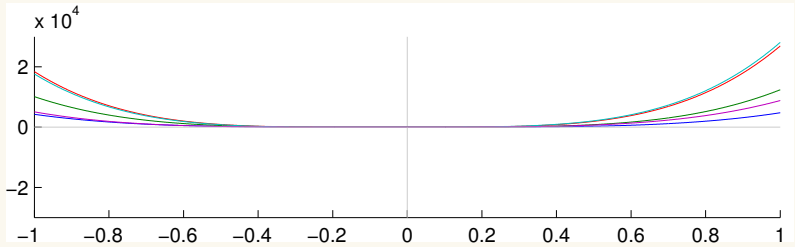
and the corresponding dual basis functions are



For clustered sampling points the kernel basis functions are



and the corresponding dual basis functions are



Note the scaling of the dual basis functions.

2.4.3 Choice of Sampling Points

- In general not any set of points will work.
- For spaces with globally supported bases, such as our example space $\mathcal{H}_4$ almost any set of distinct points will work; except we pick exactly the reproducing kernel points.
- Location can have great influence on quality.
- With ideal locations one would obtain orthogonal reproducing kernel basis. However, finding the locations that yield such a basis is highly nontrivial.

Additionally, on complex domains such points might not exist, which is already the case for the sphere.

- Oversampling as a strategy to make representation robust to non-bandlimited input.
- Fortunately, one can numerically optimize locations. This enables to numerically construct locations that are well suited for a setting. An alternative is to use so called *low discrepancy* or *quasi random points* which, when mapped appropriately to a domain, provide usually good points.
- For spaces with locally supported basis functions, for example wavelets, the situation is more complex. As a rule of thumb one then has to employ points that are distributed so that they match the local density of the basis functions. How to effectively combine this with nonlinear approximation (or compressed sensing) is currently an open question.

Exercise 16. *Consider the Paley-Wiener space $\Omega_\pi(\mathbb{R})$ of Fourier-bandlimited functions with bandlimit π, cf. Example 23. Show that*

$$\left\{ \operatorname{sinc}_i(x) = \frac{\sin\left(\pi(i-x)\right)}{\pi(i-x)} \right\}_{i\in\mathbb{I}} \tag{2.29}$$

is an orthogonal basis for $\Omega_1(\mathbb{R})$. Begin by plotting $\operatorname{sinc}(i-x)$ for $i = -5, \cdots, 5$.

Solution: The result follows immediately from the fact that the zero crossings of $\operatorname{sinc}_i(x)$ are all integers except i, which follows from the zero crossings of the sine functions.

2.4.4 Pointwise Numerical Techniques

- Sampling theorems.
- Integration.

$$\int_X f(x)\,dx = \cdots \tag{2.30}$$

- Interpolation.

Exercise 17. *Let (x_i, y_i) be k given data points. Interpret the classical Lagrange interpolation formula*

$$L(x) = \sum_{j=0}^{k} y_j\, \ell_j(x) \tag{2.31a}$$

with

$$\ell_j(x) = \prod_{i \neq j} \frac{x - x_i}{x_j - x_i} \tag{2.31b}$$

from the point of view of reproducing kernel basis expansions.

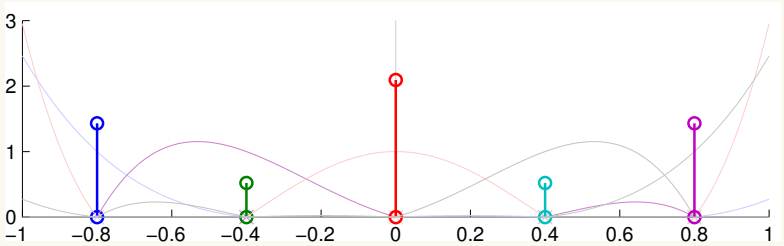
Solution: The $\ell_j(x)$ are interpolatory dual reproducing kernel basis functions for the space $P^{k-1}(\mathbb{R})$ spanned by all polynomials up to degree k. Since the "sampling points" x_i are arbitrary and depend on the

Homework 7. *Derive a quadrature rule for H_4 and compute the weights. Experiment how the distribution of sampling locations affects the quadrature weights. When the zeros of P_5 are used as quadrature nodes then this is known as Gauss-Legendre quadrature. Why is this a good choice?*

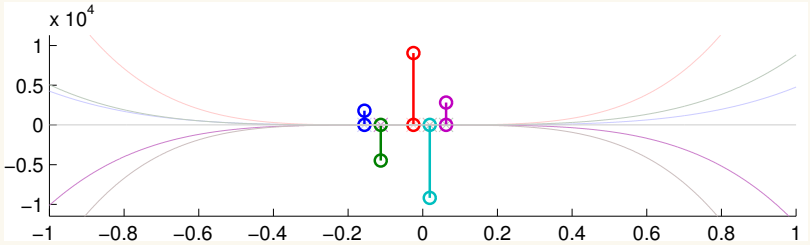
Solution: As for many classical polynomial bases, the Fourier basis, and many wavelets, the integral of all but the first Legendre polynomial vanishes. Hence, we obtain for Eq. 2.30,

$$w_i = \sum_{j=1}^{n} s_{ji} \int_{-1}^{1} P_i(x)\, dx = s_{0j}\, c_0 \tag{2.32}$$

where c_0 is the integral of $P_0(x)$ over $[-1, 1]$. Computing the weights for the well distributed points considered before, $\Lambda = \{-0.8, -0.4, 0.0, 0.4, 0.8\}$, we obtain (quadrature weights scaled by three for better visibility; the dual kernel functions $\tilde{k}_i(x)$ are shown in the background):



Quadrature weights for the clustered sampling points $\Lambda = \{-0.1562, -0.1125, -0.0250, 0.0187, 0.0625\}$ we used before are:



Note that the scaling of the weights differs by four order of magnitude. Additionally, for clustered sampling points one has positive and negative weights which is numerically undesirable.[17] As an example, we can integrate the Legendre polynomials $P_i(x)$, $1 \leq x \leq 4$, where we know that the integral vanishes. With the well distributed sampling points we obtain in double precision an average error of -6.94×10^{-18} while the clustered points yield -1.14×10^{-13}. The four digits that are lost compared to machine precision for the clustered points are roughly described by the magnitude of the quadrature weights.

The quadrature nodes that are classically used for integration in the spaces $\mathcal{H}^n_{\text{Leg}}$ are the n zero crossings of the Legendre polynomial P_{n+1} of degree $n+1$. The quadrature rule is then known as Gauss-Legendre quadrature. Using the zero crossings of P_{n+1} increases the accuracy of the quadrature rule, as can be seen using an error analysis.[18]

Exercise 18. *Devise an interpolation technique that employs the values $f(\lambda_i)$ of a function and its derivatives $f'(\lambda_i)$.*

TODO: Work out details

- Derivative is a linear operator so it commutes with the basis representation.
- One can reconstruct a reproducing kernel for derivative information.
- Combining these ideas should yield the result.

Example 26. In this example we will show how reproducing kernels enable a functional analytic interpretation of Monte Carlo integration. For this, we will also require the concept of a characteristic basis and its connection to reproducing kernels.

Let X be a set. The *characteristic function* $\chi_Y : X \to \mathbb{R}$ for a subset $Y \subset X$ is defined as

$$\chi_Y(x) = \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{otherwise} \end{cases} . \tag{2.33}$$

A *partition of* X is the collection of disjoint subsets X_i, $i = 1 \cdots n$, of X such that their union forms again X, that is

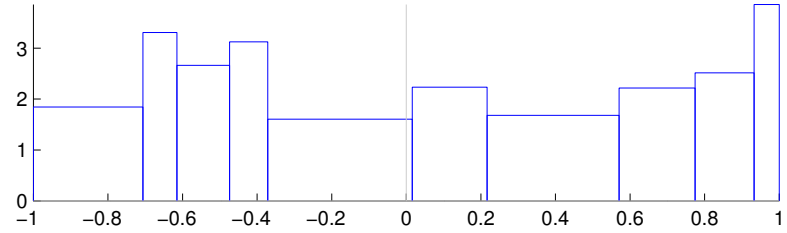
$$X = \bigcup_{i=1}^{n} X_i. \tag{2.34}$$

When we denote the characteristic function for each element X_i of a partition by $\chi_i(x)$ then we obtain a basis for the space $\text{span}(\chi_i(x)) = \mathcal{H}_\chi \equiv \mathcal{H}_\chi(X) \subset L_2(X)$. Since $\mathcal{H}_\chi(X)$ is a closed subspace of $L_2(X)$ it is a Hilbert space equipped with the L_2 inner product. We call $\{\chi_i(x)\}_{i=1}^n$ a *characteristic basis*.

Instead of unit height characteristic basis functions $\chi_i(x)$, we will in the following usually work with their normalized siblings

$$\bar{\chi}_i(x) = \frac{1}{\sqrt{|X_i|}}\chi_i(x) \tag{2.35}$$

where $|X_i|$ denotes the area of X_i. The basis $\{\bar{\chi}_i(x)\}_{i=1}^n$ is an orthonormal characteristic basis for $\mathcal{H}_\chi$. As an example, for $X = [-1, 1]$ an orthonormal basis is given by:



By Eq. 2.3 in the paper, the reproducing kernel for $\mathcal{H}_\chi$ at $\lambda \in X_j$ is given by

$$k_\lambda(x) = \sum_{i=1}^{n} \bar{\chi}_i(\lambda)\,\bar{\chi}_i(x). \tag{2.36a}$$

Since the product $\bar{\chi}_i(\lambda)\,\bar{\chi}_i(x)$ vanishes unless $\lambda \in X_j$ the sum collapses and we have

$$k_\lambda(x) = \bar{\chi}_j(\lambda)\,\bar{\chi}_j(x). \tag{2.36b}$$

By the definition of the normalized characteristic functions we thus obtain

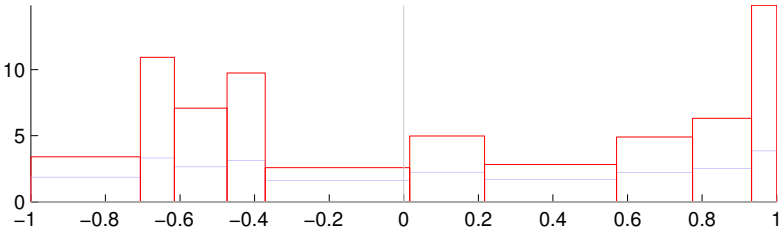
$$k_\lambda(x) = \frac{1}{|X_j|}\chi_j(\lambda) \tag{2.36c}$$

Hence, the reproducing kernel $k_\lambda(x)$ for $\mathcal{H}_\chi$ coincides with the orthonormal characteristic basis function $\bar{\chi}_i(x)$ up to a constant.

From Eq. 2.36 it follows that a reproducing kernel basis for $\mathcal{H}_\chi$ can be formed by choosing one location λ_i in each X_i. The reproducing kernel basis functions in Eq. 2.36c are then orthogonal, since their support is disjoint, but they are not orthonormal. Moreover, we cannot normalize the $k_i(x)$ since they would then lose the reproducing property; this is an instances where the general wisdom that every orthogonal basis can be carried over to an orthonormal basis by normalization of the basis functions is not true, or at least it would destroy the, for us crucial, reproducing property. From the biorthogonality condition it follows that the dual kernel functions for the reproducing kernel basis $\{k_i = 1/|X_i|\chi_x\}_{i=1}^n$ are given by $\tilde{k}_i(x) = \chi_i(x)$, that is by unnormalized characteristic functions. The basis pair for a characteristic reproducing kernel basis is hence

$$\left(\left\{k_i(x) = \frac{1}{|X_i|}\chi_i(x)\right\}\ ,\ \left\{\tilde{k}_i(x) = \chi_i(x)\right\}\right). \tag{2.37}$$

For the orthonormal characteristic basis that was shown above, the associated reproducing kernel basis is given by (reproducing kernels in red):



An important property of characteristic reproducing kernel bases is that these can be constructed and are practical for arbitrary domains X, including manifolds and high-dimensional spaces.

With the definition of characteristic reproducing kernel bases we are prepared to show how Monte Carlo integration can be obtained as a quadrature rule. Let $\{\chi_i\}_{i=1}^n$ be a characteristic basis over a partition for the set $X = [a, b] \subset \mathbb{R}$ with elements X_i. With locations $\Lambda = \{\lambda_i\}$ such that one λ_i lies in each X_i we obtain a characteristic reproducing kernel basis for the space $\mathcal{H}_\chi$ and the dual kernel functions $\tilde{k}_i(x)$ are then given by $\tilde{k}_i(x) = \chi_i(x)$. A quadrature rule associated with this reproducing kernel basis can be obtained using Eq. 2.9 in the paper. For the quadrature weights we then obtain

$$w_i = \int_X \tilde{k}_i(x)\, dx = \int_X \chi_i(x)\, dx = |\chi_i| = \frac{|X|}{n}. \tag{2.38}$$

The quadrature rule for the space $\mathcal{H}_\chi$ spanned by the χ_i is thus

$$\int f(x)\, dx = \sum_{i=1}^n w_i\, f(\lambda_i) = \frac{|X|}{n} \sum_{i=1}^n f(\lambda_i) \tag{2.39a}$$

$$= \frac{b-a}{n} \sum_{i=1}^n f(\lambda_i) \tag{2.39b}$$

Eq. 2.39b formally coincides with the standard Monte Carlo estimator for uniformly distributed sampling locations. From the definition of a probability (or measure) space, the result that Monte Carlo integration arises as a quadrature rule for the space spanned by characteristic functions is by no means surprising.[19] Also note that $\{\chi_i\}_{i=1}^n$ becomes dense in $L_2([a, b])$ as the number of partitions goes to infinity and hence asymptotically the quadrature rule is applicable for any $f \in L_2(X)$.

For samples drawn from an arbitrary probability distribution function $p(x)$ the requirement of one sample per unit height characteristic basis functions implies that the partition elements X_i can no longer equal size but have to have the form $X_i = [x_i, x_{i+1}]$ for suitable interval bounds $x_i \in [a, b]$. Choosing the x_i such that in the support of every $\chi_i(x)$ is on average one sample is then equivalent to

$$n\, P([x_i, x_{i+1}]) = n \int_{x_i}^{x_{i+1}} p(x)\, dx = 1 \tag{2.40}$$

[19] See (Rudin, *Real and Complex Analysis*).

which with $y \in [x_i, x_{i+1}]$ is to zero$^{\text{th}}$ order

$$n\ (p(y)\ (x_{i+1} - x_i)) = 1. \tag{2.41}$$

With λ_i being the samples in the support of χ_i one thus has $|\chi_i| = x_{i+1} - x_i = 1/(n\, p(\lambda_i))$. The quadrature rule with samples distributed according to $p(x)$ is therefore

$$\int f(\bar{x})\, d\bar{x} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(\lambda_i)}{p(\lambda_i)} \tag{2.42}$$

Eq. 2.42 formally coincides with the standard Monte Carlo estimator for importance sampling.

2.5 Linear Operators on Function Spaces

We have already abstractly considered linear maps in Chapter 1.5. In this section we will specialize to the case of linear spaces that are function spaces. One then typically speaks about *linear operators* or just operators. Let us begin by considering some examples.

Example 27. Linear functional is a special case of a linear operator that we already considered. Every linear functional $\varphi \in \mathcal{H}^*$ can be realized as

$$\varphi[f] = \int_X k_\varphi(x)\, f(x)\, dx \tag{2.43}$$

for a sufficiently general integral kernel $k_\varphi(x)$, which might no longer be a function in the classical sense. This is a generalization of the Riesz representation theorem in Theorem 1.1.

We have already seen simple examples for nonlinear operators in Example 5. We will see nonlinear differential and integral operators in the following sections.

2.5.1 Differential Operators

- We will only briefly discuss differential operators here and have a more detailed discussion in the next chapter.
  - It will be a good warm-up for the things that come later.
- This is the theory of partial differential equations that describe the evolution (typically time evolution) of "fields", that is continuous, not necessarily scalar functions describing quantities of physical interest.
  - All things in science are describe by ordinary or partial differential equations; at least in the hard sciences, and in particular in physics.
  - Obviously, our focus will be on linear partial differential equations. We will remark on nonlinear ones, which can be seen as a natural generalization, in Remark 12.

- Before we continue and discuss in more detail the theoretical aspects, let us consider an example

Example 28. Autonomous transport equation for function $\rho : \mathcal{M} \subset \mathbb{R}^3 \to \mathbb{R}$, for example fluid mass density in fluid mechanics, for a divergence free velocity vector field $\vec{v} : \mathcal{M} \subset \mathbb{R}^3 \to \mathbb{R}^3$[20] is given by

$$\frac{\partial \rho}{\partial t} + \vec{v} \cdot \frac{\partial \rho}{\partial x} = 0. \tag{2.44}$$

Interpretation:

- The mathematical formula is quite useless if one cannot interpret the equation and put some meaning into it.
- So called Eulerian representation.
- Right hand side is zero because it describes a conservation equation. Here the mass density ρ is globally conserved, that is mass is neither created nor destroyed.
- The equation is somewhat easier to interpret when it is written as

$$\frac{\partial \rho}{\partial t} = \vec{v} \cdot \frac{\partial \rho}{\partial x}. \tag{2.45}$$

- The change of $\rho(x)$ at x in time, this is what the derivative on the left hand side says, is given by the divergence on the right hand side. We will see what the divergence means later but the general form in the Eulerian representation, the change of the quantity in time is expressed by some differential operator is very important. To complete the picture one has to understand the differential operator. In the above case this is best done later when we introduced a differential operator known as Lie derivative.
- TODO: figure

General form of Linear Partial Differential Equations

- Derivative of functions on the real line is well known.
- We can look at it as a map

$$\frac{d}{dx} : \mathcal{F}(\mathbb{R}) \to \overline{\mathcal{F}}(\mathbb{R}) \tag{2.46}$$

  where $\mathcal{F}(R)$,$\overline{\mathcal{F}}(\mathbb{R})$ are suitable function spaces.
- Hence, d/dx is an operator.

[20] See Chapter 3.4 for the definition of divergence and its meaning.

- Is d/dx also a linear operator or map? By the standard rules of differentiation, this is indeed the case:

$$\frac{d}{dx}\Big(a\,f(x)+b\,g(x)\Big)=a\frac{df}{dx}+b\frac{dg}{dx}. \tag{2.47}$$

- Remark: d/dx vs. $\partial/\partial x$. We will typically use the latter in the following.

- We define $C^k(\mathbb{R})$ as the class of k-times continuously differentiable functions. Then

$$\frac{d}{dx}:C^{k+1}(\mathbb{R})\to C^k(\mathbb{R}) \tag{2.48}$$

  and $d/dx : C^\infty \to C^\infty$

- A general linear differential operator on $\mathbb{R}^n$ of order k acting on a function $f:\mathbb{R}^n\to\mathbb{R}$ has the form

$$A_x=\sum_{|\alpha|\le k} c_\alpha(x)\frac{\partial}{\partial x^\alpha} \tag{2.49}$$

  where $\alpha=(\alpha_1,\cdots\alpha_n)$ is a multi-index that determines the degree of the derivative in each dimension. In $2D$ a differential operator of degree 1 can for example have

$$\alpha=(0,0)\quad \frac{\partial}{\partial x^\alpha}f=\frac{\partial}{\partial x_1^0\,x_2^0}f=f \tag{2.50a}$$

$$\alpha=(1,0)\quad \frac{\partial}{\partial x^\alpha}f=\frac{\partial}{\partial x_1^1\,x_2^0}f=\frac{\partial f}{\partial x_1} \tag{2.50b}$$

$$\alpha=(0,1)\quad \frac{\partial}{\partial x^\alpha}f=\frac{\partial}{\partial x_1^0\,x_2^1}f=\frac{\partial f}{\partial x_2} \tag{2.50c}$$

$$\alpha=(1,1)\quad \frac{\partial}{\partial x^\alpha}f=\frac{\partial}{\partial x_1^0\,x_2^1}f=\frac{\partial f}{\partial x_1\,x_2} \tag{2.50d}$$

  The $c_\alpha(x)$ are coefficient functions that can on the position x; in the simplest case these are just constants.

- This is the coordinate form usually used in the theory of partial differential equations.

- Remark (appropriate function spaces): We can still defined C^k but its in general not a very useful notion. Most functions have a different differentiability class in each dimension. Anisotropic function space that allow for different differentiability, and more general regularity, at each point and each coordinate direction one needs much more complicated spaces. For questions like existence of well defined solutions one also needs Sobolev spaces that provide additional control over the magnitude of the derivatives.

Example 29. Concrete example: Laplacian in $\mathbb{R}^n$

$$\Delta f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2} \tag{2.51}$$

it is the sum of the pure or unmixed second partial derivatives; to be even more concrete, in $\mathbb{R}^3$ we have

$$\Delta f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i^2} = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \frac{\partial^2 f}{\partial x_3^2}. \tag{2.52}$$

Initial value problem and boundary conditions

- One knows the differential equation, the "law", for example for the time evolution, and one is interested how the system develops for finite times.
- The pde only gives evolution for infinitesimal times, and one has to extend this to finite times.
- Alternatively, one might be interested in the steady state, see Remark 11.

Boundary conditions:

- On a compact domain, as one typically considers in applications, the differential equations require boundary to be well defined, that is a description of what happens to the function at the boundary where the derivative is not well defined. For example, we could require that $f(x)$ and all derivatives of $f(x)$. Clearly the boundary conditions will affect the solution, possibly over the whole domain. The most common boundary conditions are:
  - Dirichlet boundary conditions: the function value at the boundary is specified.
  - Neumann boundary conditions: The derivative of $f(x)$ in the normal direction is specified (physically, this corresponds to in and out flux from a domain, which makes them physically very important).
  - TODO: figure
- Periodic boundary conditions with period z: $f(x+z) = f(x)$ which in particular means that one identifies points on "opposite boundaries" TODO: figure; this corresponds to a torus.

Solving:

- Solving a differential equation for given initial data can rarely be done analytically.

- We will talk about Green's functions as an analytic approach to the solution and about Galerkin projection for the numerical solution later. Additionally, the method of characteristics is often very useful, cf. Example 45.
- As a rule of thumb, and we talk about this later in more detail, one should understand as much about the application and mathematical structure of a partial differential equation before attempting to solve it numerically.
- There are recipes like finite element methods but they are far from foolproof.
- Mathematics: tries to prove existence of solutions. No general result.
  - Millenium prize for Navier-Stokes. Terence Tao's recent result for Navier-Stokes.
- Analytic techniques like separation of variables. Subject of a specialized course on the subject.
  - If possible, often a useful step before numerical treatment.

Remark 11. We see how the above partial differential equations describe time evolution. A stationary or steady state solution is also often of interest. It means that the quantity of interest, say $\phi(x,t)$ does not change. Hence, one has to have

$$\frac{\partial \phi}{\partial t} = 0. \tag{2.53}$$

For example, for the fluid density in Eq. 2.44 the steady state is given by

$$\frac{\partial \rho}{\partial t} = 0 = \vec{v} \cdot \frac{\partial \rho}{\partial x}. \tag{2.54}$$

Note that this by no means implies that the fluid velocity vanishes or that no mass is transported by the fluid. What is steady or time invariant is the value of the density at the location x.

We will leave it at this for the moment and return to the subject later. However, before we do this let us look at some examples:

Example 30. The heat equations for a function $u : \mathcal{M} \subset \mathbb{R}^3 \to \mathbb{R}$, for example the temperature at $x \in \mathcal{M}$, is given by

$$\frac{\partial u}{\partial t} + \Delta u = 0 \tag{2.55}$$

As the name suggests, it describes the diffusion of heat in an environment. As boundary conditions we can for example use Neumann boundary conditions with

$$\left.\frac{\partial u(x)}{\partial \vec{n}(x)}\right|_{\partial \mathcal{M}} = 0 \tag{2.56}$$

which corresponds to a perfect insulator, that is no heat flows through the boundary.

TODO: Figure with time evo. From Wikipedia? 1D or 2D?

Example 31. Scalar wave equation for $\phi : \mathcal{M} \subset \mathbb{R}^3 \to \mathbb{R}$

$$\frac{\partial^2 \phi}{\partial t^2} + \Delta \phi = 0 \tag{2.57}$$

The equation is for example central to acoustics.

The scalar wave equation in Eq. 2.57 is second order and hence more difficult to interpret than the autonomous transport equation in Example 28 where we could find an interpretation. However, by introducing

$$\vec{s} = \frac{\partial \phi}{\partial t} \;:\; \mathcal{M} \subset \mathbb{R}^3 \to \mathbb{R} \tag{2.58a}$$

$$\vec{r} = \frac{\partial \phi}{\partial x} \;:\; \mathcal{M} \subset \mathbb{R}^3 \to \mathbb{R}^3 \tag{2.58b}$$

as independent variables, Eq. 2.57 becomes

$$\frac{\partial s}{\partial t} = \sum_{i=1}^{3} \frac{\partial r}{\partial x_i} \tag{2.59a}$$

$$\frac{\partial r}{\partial t} = \frac{\partial s}{\partial x}. \tag{2.59b}$$

Hence, we have two first order transport equation that can be interpreted similarly to the autonomous transport equation in Example 28. The additional complexity of the second order equation is thereby "hidden" in the coupling between the two equations. [21]

Exercise 19. *Write the scalar wave equation in the form of a general linear differential operator in Eq. 2.49 by using space-time coordinates where $x_0 = t$.*

The wave equation can be seen as a space-time Laplace operator,

$$Wf = \sum_{i=0}^{n} \frac{\partial^2 f}{\partial x_i^2}. \tag{2.60}$$

However, as we will see later the time coordinate does have a different behaviour than the space coordinates so that the properties of W are quite different.

- We will return to the above examples in Chapter 3 where we will reconsider them from a more conceptual perspective.

[21] The reader might have seen something similar already in the context of ordinary differential equations where one can also reduce n^{th} order equations to a system of first order equations. The latter ones are typically used for numerical integration.

- Laplace operator arose as part of the heat equation and the Laplace equation. It also plays an important role, and has found a many, many applications in computer graphics. We will also consider it again from various perspective below.

Let us conclude this section with going beyond the linear regime and briefly discussing nonlinear differential operators.

Remark 12. As an extension of Eq. 2.49, the most important examples of nonlinear differential operators are of the form

$$N(f) = \left(\sum_{|\alpha|<k} c_i(f) \frac{\partial}{\partial x^\alpha} \right) f \tag{2.61}$$

where the $a_i(f)$ depend on the argument f and $\alpha = (\alpha_1, \cdots, \alpha_n)$ is again a multi-index. Such nonlinear operators have a multitude of applications in science. For example, the Euler equation, which describes the time evolution of the velocity vector field $\vec{v}(x,t)$ of an ideal fluid, is given by

$$\frac{\partial \vec{v}}{\partial t} + \nabla_{\vec{v}}\, \vec{v} = \nabla p \tag{2.62}$$

where $\vec{v}$ is the fluid velocity vector field and p the pressure, and we see that in the second term on the left hand side $\vec{v}$ is part of the differential operator and the vector-valued function the operator acts on. We will return to the Euler equation again in later chapters. Nonlinear generalization of the linear differential and integral operators that we consider below in Chapter 2.5.1 and Chapter 2.5.2.

2.5.2 Integral Operators

- The second class of linear operators we will consider are integral operators. We will see in the following when we consider Green's functions that in many instances an intimate connection between differential and integral operators exists.

- Definition: given by $K : \mathcal{F}(X) \to \overline{F}(Y)$:

$$g(y) = K(f) = \int_X k(x,y)\, f(x)\, dx \tag{2.63}$$

  and $k(x,y)$ is known as the *kernel* of the integral operator.

Example 32. Linear functionals are (somewhat degenerate) example where $\overline{F}(Y) = \mathbb{R}$

Example 33. *Hilbert-Schmidt integral operators.*

- The kernel is in L_2 as a function of x and y. This ensures that the operator is compact, and hence well behaved in a sense somewhat beyond our discussion.
- When the kernel is symmetric then the operator is self-adjoint.
- Compactness and self-adjointness are the essential condition for spectral theorem to apply, cf. Theorem 1.3. Hence, Hilbert-Schmidt operators have a well defined eigen decomposition, and, moreover, the necessary conditions are easy to verify.

Hilbert-Schmidt operators are the generalization of the above integral operators that have the same convenient properties.

Exercise 20. *Show that on $L_2(X)$ a Hilbert-Schmidt operator is self-adjoint when the kernel is symmetric.*

Solution: L_2 inner product and integral of operator action can change. L_2 integrability ensures that Fubini's theorem can be used and the interchange is actually valid. TODO: Verify the idea

Example 34. TODO: Rendering: scattering as Hilbert-Schmidt operator.

Remark 13. By a celebrated result by Laurent Schwartz, every "reasonable" operator on the space of compactly supported C^∞ test functions, cf. Ex. 21, can be realized as an integral operator. This in particular also includes differential operators. As an example, the integral representation of the identity operator is

$$\mathrm{Id}(f)|_{\bar{x}} = \int_X \delta_{\bar{x}}(x)\, f(x)\, dx \tag{2.64}$$

where $\delta_{\bar{x}}(x)$ is the Dirac delta distribution, cf. again Ex. 21.

Example 35. Markov chain. Integral operator $K : \mathcal{F}(X) \to \mathcal{F}(X)$ on a space with $\lambda = 1$ as largest eigenvalue, that is

$$f(y) = \int_X k(x, y)\, f(x)\, dx. \tag{2.65}$$

Integral kernel $k(x, y)$ then models the transition probability between the states $x \in X$ and $y \in X$. Hence, it also satisfies

$$1 = \int_X k(x, y) dy \tag{2.66}$$

to ensure that the transition probability is indeed a probability. The eigenfunction $f(x)$ associated with $\lambda = 1$ is the *steady state distribution*. The terminology comes from the fact that $f(x)$ is invariant under the integral operator.

Finite dimensional case, K is a matrix and the elements K_{ij} describe the transition probabilities between the finite state space, interpreted as $\mathbb{R}^n$, cf. Exercise 21. The steady state distribution is then the *left* eigenvector associated with K.

Markov chain: realization of a Markov process where the next state only depends on current state and not other states further in the past. This is crucial for matrix representation.

General insight: one can but does not have to look at things probabilistically.[22]

Exercise 21.[23] *We consider a Markov chain that is used weather. For simplicity we categorize weather as being in one of three states:* `rain`, `clouds`, `sun`. *The next day forecast is then given by*

	`rain`	`clouds`	`sun`
`rain`	0.5	0.25	0.25
`clouds`	0.5	0.0	0.5
`sun`	0.25	0.25	0.5

where the rows are tomorrow's weather and the column's today's, and the entries represent probabilities. For the Markov chain defined by the above transition matrix, compute the steady state distribution. Implement the power iteration method for this.

Solution: The steady state solution, given by the left eigenvector, is $f = [0.670.330.66]^T$. See `power_iteration.py`.

Remark 14. Markov chain's are the basis for Markov chain Monte Carlo methods that have various applications in science and engineering, and naturally also in computer graphics.[24]

Example 36. Example for nonlinear integral operator:

$$N(f) = \int_X k(x, y; f)\, f(x)\, dx \tag{2.67}$$

where the integral kernel $k(x, y; f)$ depends on the argument $f(x)$. In applications in physics one often has that an integral operator is in general nonlinear but that in a restricted regime the integral kernel is independent of f and one hence can with a linear integral operator, which is much easier to handle. An example is the scattering function in Ex. 34. For moderate light intensities $\rho(\omega, \bar{\omega})$ is independent of it but when the intensity becomes very high then material properties change and the scattering kernel has a nonlinear dependence on the intensity.

[22] For a reinterpretation of many ideas from machine learning from an analytic perspective see (Cucker and Zhou, *Learning Theory: An Approximation Theory Viewpoint*) and references therein.

[24] For example (Veach and Guibas, "Metropolis Light Transport"; Chenney and Forsyth, "Sampling Plausible Solutions to Multi-body Constraint Problems").

Green's Functions Let us begin by recalling the spectral theory that we introduced in Chapter 1.5.2. and that immediately applies to linear operators.

- Recall spectral theorem for operators.
- In our formulation is immediately applies to operators.
- Requires compactness. Generalizations exist but too technical for our purposes. For example, in general continuous spectrum possible, which is much more difficult to work with (although we will encounter it when we talk about the Fourier transform). Also, there might be no eigenfunctions at all, or they are or can only defined implicitly.

Definition 2.8. *Green's function*

Proposition 2.1. *Eigenvector expansion of Green's function*

Exercise 22. *Show the above proposition.*

Solution: TODO: Show it.

Example 37. Poisson equation

$$\Delta f = g \tag{2.68}$$

The Laplace equation is $g = 0$, that is the homogenous form of the equation. On $\mathbb{R}^3$ the Green's function of the Poisson equation is given by

$$G(x, \bar{x}) = -\frac{1}{4\pi}\frac{1}{|x - \bar{x}|}. \tag{2.69}$$

The Poisson and Laplace equations arise in a multitude of applications in CG.

- TODO: Provide a concrete example. For example surface reconstruction. Where was Green's function actually used?

Homework 8. *Eigenfunction's of heat operator H are Fourier and eigenvalues are ... Numerically construct the Green's function. For this, experiment with the number of eigenfunctions that are needed. Guess the analytic form*

Solution: TODO: Implement!

How to solve differential equations using Green's functions:

- Sometimes, like in the case of the heat equation, the Green's function has an analytic form and this form is sufficiently nice to enable an analytic solution.
  - Restricted to simple / symmetric domains.

  - "Canonical" differential equations. Although the number of "useful" differential equations is surprisingly small and one encounters the same equation typically over and over again.

- In principal one can use quadrature since one has an integral. However, this only yields the solution at single point. Galerkin projection, which will be discussed in the next section,

2.5.3 Galerkin Projection

Galerkin projection[25]

- Very general approach to obtain finite dimensional representations for operator equations.

  - Classical form and the once we will discuss are for linear operator equations but also extensions for nonlinear ones.

Let $A : F \to G$ be a linear operator between separable Hilbert spaces F and G, that is

$$A\, f = g. \tag{2.70}$$

Furthermore assume that $\{\varphi_i\}_{i=1}^{\infty}$ is an orthonormal basis for F and $\{\psi_j\}_{j=1}^{\infty}$ is an orthonormal basis for G. For Galerkin projection, we project Eq. 2.70 onto ψ_j,

$$\langle A\, f, \psi_j \rangle = \langle g, \psi_j \rangle \tag{2.71a}$$

and expand f in $\{\varphi_i\}_{i=1}^{\infty}$,

$$\left\langle A\Big(\textstyle\sum_i f_i\, \varphi_i\Big), k_j \right\rangle = \langle g, \psi_j \rangle\,. \tag{2.71b}$$

By linearity we then have

$$\sum_{i=1}^{\infty} f_i\, \langle A\, \varphi_i, \psi_j \rangle = \langle g, \psi_j \rangle\,. \tag{2.71c}$$

and by defining $A_{ij} = \langle A\, \varphi_i, \psi_j \rangle$ we obtain

$$\sum_{i=1}^{\infty} f_i\, A_{ij} = g_j. \tag{2.71d}$$

- The A_{ij} form an infinite matrix.

[25] Galerkin, "On Electrical Circuits for the Approximate Solution of the Laplace Equation"; Petrov, "Application of the Method of Galerkin to a Problem Involving the Stationary Flow of a Viscous Fluid".

- A numerically practical realization is obtained by truncating the bases $\{\varphi_i\}_{i=1}^{\infty}$ and $\{\psi_j\}_{j=1}^{\infty}$ at some finite n and m, respectively. Then

$$\bar{A}\,\bar{f} = \bar{g} \tag{2.72}$$

  where $\bar{A} \in \mathbb{R}^{n\times m}$ and $\bar{f}$ and $\bar{g}$ are n- and m-dimensional vectors, respectively.

- Brings us back to the question what are good approximation spaces, or
- With the discussion of Chapter 2.3 in mind, Galerkin projection is a linear ansatz in that one fixes a subspace a priori and not based on the data.
  - One can imagine choosing A based on the problem, for example by locally refining where necessary. An example of this idea are adaptive PDE methods.

Homework 9. *Perform a derivation analogous to those in Eq. 2.71 for biorthogonal bases for F and G.*

TODO: Do derivation.

Example 38. TODO: Finite Differences and finite volume methods Should we include it? What are applications where it is preferable over finite elements?

- Can one consider the finite volume method as a finite element method for constant test functions?
- Application of divergence theorem as a form of weak formulation?

Example 39. Finite elements Before Galerkin projection, write a linear equation

$$A\,f = g \tag{2.73}$$

with $A : W \to V$ in weak form

$$\phi(f, h) = g(h) \tag{2.74}$$

where $h \in \tilde{V}$ is an element in the dual space $\tilde{V}$ of V. Hence $g(h)$ refers to the natural pairing between V

- Analogous to how weak derivative is defined in Example XXX, and in fact equivalent in that the original PDE can then be applied to distributions, within a mathematically rigorous theory.
- Lax-Milgram theorem that asserts the existence of a unique solution.[26]

[26] Lax and Milgram, "Parabolic equations".

- Allows to work with derivative even for C^0 (basis) functions.
- Add figure with 2D elements.

The Poisson equation $\Delta f = g$ in one dimension is given by

$$\frac{\partial^2}{\partial^2 x} f = g. \tag{2.75}$$

As the domain $\mathcal{M}$ of the problem we will assume $\mathcal{M} = [0, 1] \subset \mathbb{R}$ with Dirichlet boundary conditions $f(0) = f(1) = 0$.

To obtain the weak formulation we require equality Eq. 2.75 to hold with respect to an arbitrary test function h from a suitable space. We will employ the Sobolev space $H_0^1([0, 1])$ of functions with bounded derivative which also satisfy $h(0) = h(1) = 0$ as test function space; since the boundary $\partial\mathcal{M} = 0, 1$ is a set of measure zero we can disregard it at the moment and enforce the correct values $f(0) = f(1) = 0$ of the solution at the end. Pairing with the test function and using integration by parts yields

$$\int_0^1 \left(\frac{\partial^2}{\partial^2 x} f(x) \right) h(x)\, dx = \int_0^1 g(x)\, h(x)\, dx \tag{2.76a}$$

$$\left[\frac{\partial f(x)}{\partial x}\, h(x) \right]_{x=0}^{x=1} - \int_0^1 \frac{\partial f(x)}{\partial x} \frac{\partial h(x)}{\partial x}\, dx = \int_0^1 g(x)\, h(x)\, dx \tag{2.76b}$$

and since $h(x)$ vanishes at the boundary the first term on the left hand side is zero. Hence, we have

$$\int_0^1 \frac{\partial f}{\partial x} \frac{\partial h}{\partial x}\, dx = \int_0^1 g(x)\, h(x)\, dx \tag{2.76c}$$

$$\phi(f, h) = \langle g, h \rangle\,. \tag{2.76d}$$

The last equation has the desired form of Eq. 2.74 with the left hand side being the bilinear pairing

$$\phi(f, h) = \int_0^1 \frac{\partial f}{\partial x} \frac{\partial h}{\partial x}\, dx \tag{2.77}$$

and the right hand side the usual L_2 inner product.

To obtain a finite approaximation of Eq. 2.76d we have to perform Galerkin projection and expand the left and right hand side in the piecewise-linear finite element basis $\{v_i\}_{i=1}^n$ which we assume, without loss of generality, to be defined over a regular grid, cf. Fig. XXX. This yields

$$\phi\left(\sum_{a=1}^n f_a v_a(x)\,, \sum_{b=1}^n h_b v_b(x) \right) = \left\langle \sum_{c=1}^n g_c v_c(x), \sum_{d=1}^n h_d v_d(x) \right\rangle. \tag{2.78a}$$

and by (bi-)linearity of ϕ and the inner product we obtain

$$\sum_{a=1}^{n}\sum_{b=1}^{n} f_a\, h_b\, \phi(v_a(x), v_b(x)) = \sum_{c=1}^{n}\sum_{d=1}^{n} g_c\, h_d\, \langle v_c(x), v_d(x)\rangle\,. \tag{2.78b}$$

Since the $v_i(x)$ are polynomials we can explicitly evaluate $\phi(v_a(x), v_b(x))$ and $\langle v_c(x), v_d(x)\rangle$. Without loss of generality, assume $b > a$. We then have

$$L_{ab} = \phi(v_a(x), v_b(x)) \tag{2.79a}$$

$$= \int_{ha-h}^{ha+h} \frac{\partial v_a}{\partial x}\,\frac{\partial v_b}{\partial x}\,dx. \tag{2.79b}$$

When the support of $v_a(x)$ and $v_b(x)$ is disjoint then the integral vanishes immediately. From Fig. XXX it folds that this is the case unless $b = a + 1$ in which case

$$L_{ab} = \int_{ha-h}^{ha} 1/h\, 0\, dx + \int_{ha}^{ha+1} 1/h\,(-1/h)\,dx + \int_{ha+1}^{ha+2} 0\,(-1/h)\,dx \tag{2.79c}$$

$$= -h\,(1/h^2) = -1/h. \tag{2.79d}$$

Hence, in general L_{ab} is nonzero if and only if $b = a + 1$ and or $b = a - 1$. The L_{ab} form the so called "stiffness matrix", a term that comes from early applications of the finite element methods to elasticity. Analogous to Eq. 2.79, for $\langle v_c(x), v_d(x)\rangle$ we have for $d > c$

$$M_{cd} = \langle v_c(x), v_d(x)\rangle \tag{2.80a}$$

$$= \int_{hc-h}^{hd+h} v_c(x)\, v_d(x)\, dx \tag{2.80b}$$

which is nonzero if and only if c and d are adjacent, in which case one has

$$M_{cd} = h/6 \tag{2.80c}$$

The M_{ij} form the "mass matrix". Note that both the stiffness matrix and the mass matrix are highly sparse due to the highly localized support of the $v_i(x)$.

Analogous for 2D. The only slight complication is integration by parts for which ...

Remark 15. Spectral methods: Galerkin with Fourier basis functions. The term is often used for any form of Galerkin projection with globally supported basis functions. These methods are highly efficient for smooth phenomena on regular domains but lose in accuracy when the solution has discontinuities or the domain is not a Cartesian subset of $\mathbb{R}^n$.

Example 40. We already introduced the Poisson equation in Example 37 and discussed its solution in one dimension using the finite element method in Example 39. On $[0, 2\pi,] \times [0, 2\pi]$, the equation is given by

$$\left(\frac{\partial^2}{\partial x_1} + \frac{\partial^2}{\partial x_2}\right) f(x) = g(x). \tag{2.81}$$

When we assume periodic boundary conditions the functions $f(x)$ and $g(x)$ can be expanded in the Fourier series as

$$f(x) = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \hat{f}_{a,b}\, e^{iax_1}\, e^{ibx_2} = \sum_{a,b=0}^{\infty} \hat{f}_{a,b}\, e^{i(ax_1+bx_2)} \tag{2.82a}$$

$$g(x) = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \hat{g}_{a,b}\, e^{iax_1}\, e^{ibx_2} = \sum_{a,b=0}^{\infty} \hat{g}_{a,b}\, e^{i(ax_1+bx_2)} \tag{2.82b}$$

for complex coefficients $\hat{f}_{a,b}$ and $\hat{g}_{a,b}$. Inserting the Fourier representations into the Poisson equation in Eq. 2.82 we have

$$\left(\frac{\partial^2}{\partial x_1} + \frac{\partial^2}{\partial x_2}\right) \sum_{a,b=0}^{\infty} \hat{f}_{a,b}\, e^{i(ax_1+bx_2)} = \sum_{a,b=0}^{\infty} \hat{g}_{a,b}\, e^{i(ax_1+bx_2)}. \tag{2.83a}$$

Using the linearity of differentiation we obtain

$$\sum_{a,b=0}^{\infty} \hat{f}_{a,b} \left(\frac{\partial^2}{\partial x_1} + \frac{\partial^2}{\partial x_2}\right) e^{i(ax_1+bx_2)} = \sum_{a,b=0}^{\infty} \hat{g}_{a,b}\, e^{i(ax_1+bx_2)} \tag{2.83b}$$

and since $(\partial/\partial x)e^{\alpha x} = \alpha e^{\alpha x}$ this can be written as

$$\sum_{a,b=0}^{\infty} \hat{f}_{a,b} \left(-a^2 - b^2\right) e^{i(ax_1+bx_2)} = \sum_{a,b=0}^{\infty} \hat{g}_{a,b}\, e^{i(ax_1+bx_2)} \tag{2.83c}$$

Since the Fourier coefficients of a function are unique we hence have to have

$$-\hat{f}_{a,b} \left(a^2 + b^2\right) = \hat{g}_{a,b}; \tag{2.84}$$

which, following the ansatz of Galerkin projection, could also be obtained by projecting both sides onto $e^{i(\bar{a}x_1+\bar{b}x_2)}$ and using the orthonormality of the Fourier basis functions. Eq. 2.84 enables to directly solve for the $\hat{f}_{a,b}$ given the coefficients $\hat{g}_{a,b}$ on the right hand side. To numerically implement this method one has to choose a finite bandlimit N and only consider the Fourier expansion up to this term.

For the given boundary condition, the Fourier basis hence enables a very efficient solution of the Poisson equation. In fact, the basis diagonalizes the operator, since $\hat{f}_{a,b}$ only depends on $\hat{g}_{a,b}$ and no other basis function coefficients of $g(x)$, cf. Chapter 2.5.4. Unfortunately, for more complex domains such a simple solution is no longer possible.[27]

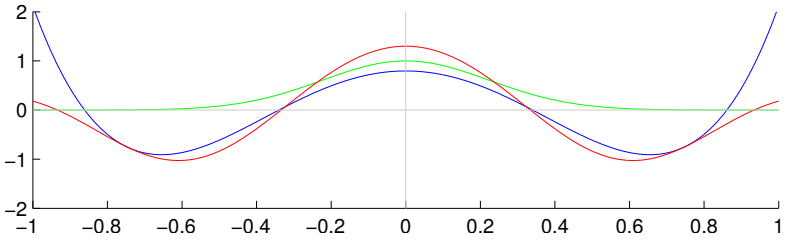
TODO: Should we write the above derivation to more closely follow the general Galerkin projection eq., which mainly corresponds to not explicitly expand the right hand side? TODO: Implement.

[27] For approaches to solve the Poisson equation relevant for graphics see for example (Botsch, Kobbelt, Pauly, Alliez, and Levy, *Polygon Mesh Processing*).

Example 41. The integral equation

$$\tilde{g}(y) = Kg|_y = \int_{-1}^{1} g(x) \left(15 \exp\left(-10 * (x-y)^2\right)\right) \, dx$$

with a Gaussian kernel (green, scaled by 1/15) for the Legendre polynomial $f(x) = P_4(x)$ (blue) as input:



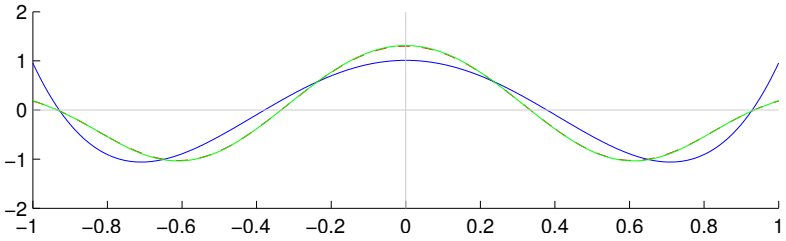
The result $g(x)$ is plotted in red. The approximation of the kernel in $\mathcal{H}^4_{\text{Leg}}$ is given by

$$K = \langle P_i(x)|k(x,y)|P_j(y)\rangle = \begin{pmatrix} 2.55 & 0.0 & -0.35 & 0.0 & -0.14 \\ 0.0 & 2.07 & 0.0 & -0.51 & 0.0 \\ -0.35 & 0.0 & 1.61 & 0.0 & -0.51 \\ 0.0 & -0.51 & 0.0 & 1.21 & 0.0 \\ -0.14 & 0.0 & -0.52 & 0.0 & 0.88 \end{pmatrix}$$

The sparsity results from the symmetry of the kernel. When we solve the above integral equation through its finite representation, that is

$$\tilde{g} = Kg = \begin{pmatrix} \tilde{g}_1^5 \\ \tilde{g}_2^5 \\ \tilde{g}_3^5 \\ \tilde{g}_4^5 \\ \tilde{g}_5^5 \end{pmatrix} = K \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.14 \\ 0.0 \\ -0.51 \\ 0.0 \\ 0.88 \end{pmatrix}$$

we obtain the approximation $\tilde{g}(y)$ (shown in blue)



An 11-dimensional approximation $\hat{g}(y) \in \mathcal{H}^{10}_{\text{Leg}}$ (green) is almost indistinguishable from the analytic solution (red, dashed). See `legPlotIntegralEquation.m` for the implementation.

Homework 10. *Let $\mathcal{H}(X)$ be a reproducing kernel Hilbert space with orthonormal reproducing kernel basis $\{k_i(x)\}_{i=1}^{\infty}$. Compute the Galerkin projection of $Af = g$ for a linear operator $A : \mathcal{H}(X) \to \mathcal{H}(X)$ by following the steps in Eq. 2.71.*

Solution: Following Eq. 2.71 we have

$$\langle A\, f, k_j \rangle = \langle g, k_j \rangle \tag{2.85a}$$

$$\left\langle A\left(\textstyle\sum_i f(\lambda_i)\, k_i\right), k_j \right\rangle = \langle g, k_j \rangle \tag{2.85b}$$

and by linearity

$$\textstyle\sum_i f(\lambda_i)\, \langle A\, k_i, k_j \rangle = \langle g, k_j \rangle\,. \tag{2.85c}$$

Exploiting now the reproducing kernel property of the k_i we obtain

$$\textstyle\sum_i f(\lambda_i)\, A(\lambda_i, \lambda_j) = g(\lambda_j). \tag{2.85d}$$

Hence, for reproducing kernel Galerkin projection the resulting matrix-vector equation only depends on pointwise values of the functions and the operator. Moreover, the "reconstruction kernels" for the right hand side are uniquely given by the reproducing kernel functions k_i.

TODO: Implement RK Galerkin projection for example? Use Ex. 41 again? H_{10} should be enough to avoid RK issues.

2.5.4 Fourier Theory Revisited

TODO: Work out.

Fourier transform, Fourier series, and linear time invariant systems.

- Translation operator T_a commutes with differentiation: first translate and then differentiate is the same as first differentiate and then translate.
  - Hence differentiation has some eigenfunctions as translation.
  - The argument generalizes to any operator that is translation invariant.
- $d/dx f = af \Rightarrow f = ce^{ax}$, for all $a \in \mathbb{R}$: this is just Fourier transform:

$$\hat{f}(\omega) = \int_{\mathbb{R}} f(x) e^{i\,\omega\,x} dx. \tag{2.86}$$

- The equation is just projection into the eigenbasis.
- Commutes with T_a. Hence T_a has the same eigenfunctions. And then every translation invariant operator has the Fourier basis as eigenfunctions.

- Consequently, translation invariant operators are diagonalized by Fourier basis.
- Application of operator is then multiplication by spectrum. Since $\mathbb{R}$ is non-compact the spectrum is continuous and it means applying a function to the continuous spectral representation.
- Spectrum is continuous for noncompact domain.
  - This is why the basis expansion is an integral and not a sum.
- But is there an obvious reason why it is discrete on compact domains?

Remark 16. When $X = \mathbb{R}^n$, then the Sobolev spaces $W^{k,p}(\mathbb{R}^n)$ can also be defined using the Fourier transform F as

$$W^{k,p}(\mathbb{R}^n) = \left\{ f \in L_p(\mathbb{R}^n) \mid F^{-1}\left((1+\xi)^{k/2}\hat{f}\right) \in L_p(\mathbb{R}^n) \right\}, \tag{2.87}$$

where $\hat{f}$ denotes the Fourier transform of the function f, that is $\hat{f} = F(f)$, and the behaviour of the derivatives is controlled by the weight $(1+\xi)^{k/2}$ which enforces a suitable decay of the Fourier coefficients as the frequency ξ goes to infinity. The above definition can be extended from $\mathbb{R}^n$ to other domains where a suitable generalization of the Fourier transform is defined, for example on the sphere,[28] and it also allows to introduce Sobolev spaces where k is not an integer.

Remark 17. Mention Fourier and derivative -> symbol calculus.

Remark 18. Beyond Fourier many other integral transforms that might be useful. For example, the Hilbert transform plays an essential role for the representation of wave signals.[29]

2.6 Further Reading

A comprehensive discussion of mathematical basis of signal processing can be found in the classic text by Mallat.[30] Introductions to the theory of frames can be found in the survey articles by Kovacevic and Chebira.[31] Approximation theory is very large subject in itself. A useful recent accounts on the classical theory is presented by Trefethen[32] and a good starting point for the modern

[28] For Sobolev spaces over the sphere see for example (Freeden, Gervens, and Schreiner, *Constructive Approximation on the Sphere (With Applications to Geomathematics)*; Hesse, "Complexity of numerical integration over spherical caps in a Sobolev space setting").

[29] Mandel and Wolf, *Optical Coherence and Quantum Optics*, Chapter 3.

[30] Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*.

[31] Kovacevic and Chebira, "Life Beyond Bases: The Advent of Frames (Part I)"; Kovacevic and Chebira, "Life Beyond Bases: The Advent of Frames (Part II)".

[32] (Trefethen, *Approximation Theory and Approximation Practice*), see also (Davis, *Interpolation and Approximation*; Cheney and Light, *A Course in Approximation Theory*).

theory for signals with varying regularity is again presented by Mallat.[33] For a more rigorous and mathematical treatment of function spaces see for example the classic text by Lax[34] or the books by Rudin.[35] For an alternative introduction to finite element methods see the article by Trenogin.[36]

[33](Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*), see for example also (R. A. DeVore and Lorentz, *Constructive approximation*; R. A. DeVore, "Nonlinear Approximation"; Wendland, *Scattered Data Approximation*; Freeden, Gervens, and Schreiner, *Constructive Approximation on the Sphere (With Applications to Geomathematics)*; Temlyakov, "Nonlinear Methods of Approximation"; Reimer, *Multivariate Polynomial Approximation*; Cucker and Zhou, *Learning Theory: An Approximation Theory Viewpoint*).

[34]Lax, *Functional Analysis*.

[35]Rudin, *Principles of Mathematical Analysis*; Rudin, *Real and Complex Analysis*; Rudin, *Functional Analysis*.

[36]Trenogin, "Galerkin Method".

Chapter 3

Manifolds and Tensors

Relevance for graphics: - Meshes, as discussed in more detail in the next section. - For simulation. - For manifold learning, i.e. the analysis of high dimensional data.[1] - Lie groups: Rotations are omnipresent in graphics, symmetry detection,[2] Lie group integrators.[3]

3.1 Preliminaries

In the following, it will be useful to have in mind some notions from algebra about the relationship between mathematical structures. We will therefore recall them at this point.

Definition 3.1. *A **homomorphism** is a structure preserving map between algebraic structures.*

Homomorphisms exists for different structures, such as algebras, Lie algebras, rings, and groups, and the precise meaning of the term depends on the structure of interest. The following examples demonstrate this.

Example 42. Let $(\mathcal{H}(X), \langle\,,\rangle)$ be a finite dimensional Hilbert space, and let $\{\varphi_i\}_{i=1}^n$ and orthonormal basis for $\mathcal{H}(X)$. Then $\{\varphi_i\}_{i=1}^n$ defines a Hilbert space homomorphism from $\mathcal{H}(X)$ to $\mathbb{R}^n$ with the usual Euclidean inner product.

[1] (Coifman, Lafon, Lee, Maggioni, Nadler, Warner, and Zucker, "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Diffusion Maps"; Coifman, Lafon, Lee, Maggioni, Nadler, Warner, and Zucker, "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Multiscale Methods") and references therein. A pivotal early paper is those by Belin and Niyogi ("Laplacian Eigenmaps for Dimensionality Reduction and Data Representation").

[2] (Lipman, Chen, Daubechies, and Funkhouser, "Symmetry Factored Embedding and Distance"); through the intimate relationship between rotations and the Laplace operator also methods that employ spherical harmonics can be considered to be based on Lie groups, see e.g. (Kazhdan, Funkhouser, and Rusinkiewicz, "Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors").

[3] Kobilarov, Crane, and Desbrun, "Lie Group Integrators for Animation and Control of Vehicles".

This is just a more formal version of the previous observation that, using an orthonormal basis such as $\{\varphi_i\}_{i=1}^n$, all operations in $\mathcal{H}(X)$ can equivalently be performed in $\mathbb{R}^n$.

Example 43. A Lie algebra homomorphism $\varphi : (\mathfrak{g}, [\,,]_\mathfrak{g}) \to (\mathfrak{h}, [\,,]_\mathfrak{h})$ between two Lie algebras $(\mathfrak{g}, [\,,]_\mathfrak{g})$ and $(\mathfrak{h}, [\,,]_\mathfrak{h})$ is a map such that $[X, Y]_\mathfrak{g} = [\varphi(X), \varphi(Y)]_\mathfrak{h}$ for all $X, Y \in \mathfrak{g}$.

Definition 3.2. *An **isomorphism** between two algebraic structures is a homomorphism whose inverse is also a homomorphism.*

An isomorphism thus defines an equivalence between two structures that enables to perform all operations on the first structure also with the second one, and then relate the result back to the first. When two structures A and B are isomorphic then this will be denoted as $A \cong B$.

Example 44. Extending Example 42, any finite dimensional Hilbert space $(\mathcal{H}, \langle\,,\rangle)$ of dimension n is isomorphic to Euclidean space $(\mathbb{R}^n, \cdot)$ with the inner product over $\mathcal{H}$ being the dot product over $\mathbb{R}^n$.

Definition 3.3. *An **endomorphism** is a homomorphism from an object onto itself.*

Definition 3.4. *An **automorphism** is an isomorphism from an object onto itself.*

Automorphisms are for example of importance in the context of (Lie) groups when it is acting on itself.

Definition 3.5. *A **homeomorphism** $\varphi : S \to T$ between topological spaces S and T is a bijective mapping that is continuous and has a continuous inverse.*

By the required continuity, a homeomorphism is an isomorphism in the category of topological spaces, preserving the topological structure of S and T. Some care is required to distinguish homomorphisms and homeomorphisms which are related—and orthographically almost coincide—but nonetheless distinct concepts.

3.2 Manifolds

3.2.1 Continuous Theory

3.2.2 Discrete Theory

3.3 Tensors

Distinguish from the tensors, that is multi-dimensional arrays, that sometimes appear in computer science.[4] - The connection is that the representation of the tensors we talk about can be multi-linear arrays.

[4] Vasilescu and Terzopoulos, "TensorTextures"; Vasilescu and Terzopoulos, "Multilinear Analysis of Image Ensembles: TensorFaces"; Tsai and Shih, "All-Frequency Precomputed Radi-

Exercises: Derivatives for matrices.

3.4 Differential Forms and Exterior Calculus

3.4.1 Continuous Theory

Use integration as example application. Also the crucial aspect for discretization. Remark on L_p spaces on manifolds.

3.4.2 Discrete Theory

3.5 The Lie Derivative

Develop continuous theory.

Make some remarks on discrete theory but no well developed theory so far.

3.6 Riemannian Manifolds

3.6.1 Continuous Theory

3.6.2 Discrete Theory

3.7 Lie Groups

Also talk about representations of matrices: this is how they occur almost always in applications.

For the examples build on the material in Mechanics and Symmetry.

3.7.1 The Euclidean Group

In Example 7 we already showed that a linear space has a group structure.

3.7.2 The Rotation Group

SO(3) and its different realizations (including quaternions) as example.

3.7.3 Diffeomorphism Groups

3.8 Further Reading

A comprehensive and rigorous discussion of manifolds and tensor calculus can be found in the book by Marsden, Ratiu, and Abraham.[5] A shorter introduction can be found in the "little Spivak".[6] A classical presentation from the point of

ance Transfer Using Spherical Radial Basis Functions and Clustered Tensor Approximation".

[5] Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*.

[6] Spivak, *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*.

physics that also tries to provide as much intuition as possible is those in the "Big Black Book: B^3" by Misner, Thorne, and Wheeler.[7]

[7] Misner, Thorne, and Wheeler, *Gravitation*.

Chapter 4

Dynamical Systems and Geometric Mechanics

- Simulation and animation; although so far almost exclusively used for simulation and control. - Fluid simulation.[1] - Control.[2] - Elastics.[3].

4.1 Motivation and Intuition

4.2 Ordinary and Partial Differential Equations

4.3 Hamiltonian Mechanics

4.3.1 Continuous Theory

Example 45. Method of characteristics, at least for $\dot{f} = -\{f, H\}$.

4.3.2 Symplectic Integrators

4.4 Lagrangian Mechanics

4.4.1 Continuous Theory

Remark 19 (Calculus of Variations)**.** Explain basic rules of classical classical calculus of variations.

[1] Elcott, Tong, Kanso, Schröder, and Desbrun, "Stable, Circulation-Preserving, Simplicial Fluids"; Pavlov, Mullen, Tong, Kanso, Marsden, and Desbrun, "Structure-preserving discretization of incompressible fluids"; Mullen, Crane, Pavlov, Tong, and Desbrun, "Energy-Preserving Integrators for Fluid Animation".

[2] Kobilarov, Crane, and Desbrun, "Lie Group Integrators for Animation and Control of Vehicles".

[3] Bergou, Wardetzky, Robinson, Audoly, and Grinspun, "Discrete Elastic Rods"; Bergou, Audoly, Vouga, Wardetzky, and Grinspun, "Discrete Viscous Threads"; Batty, Uribe, Audoly, and Grinspun, "Discrete Viscous Sheets".

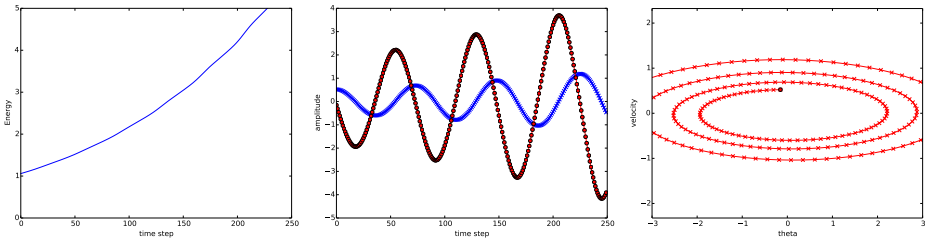


Figure 4.1: Explicit Euler integrator.

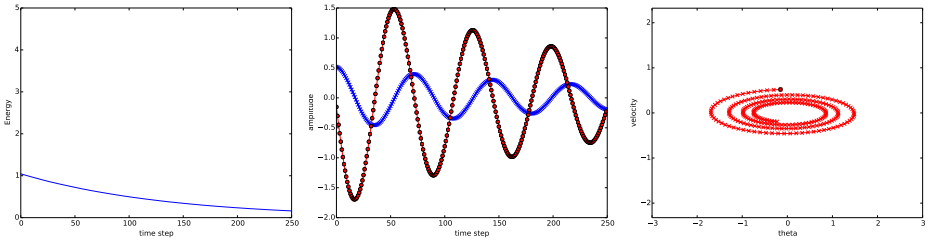


Figure 4.2: Implicit Euler integrator.

4.4.2 Variational Integrators

Begin with motivation:

- Explicit and implicit Euler for Pendulum. Always non-physical behaviour.
- Also recall mathematical form of explicit / implicit Euler.

In the continuous case, the equation motion is derived by inserting the Lagrangian for a system into the continuous Euler-Lagrange equation

$$\frac{\partial L}{\partial q} + \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = 0 \tag{4.1}$$

which we derived before in Sec. 4.4.1 from Hamilton's principle.

TODO: Add details on derivation of discrete Euler-Lagrange equations. Look at (Stern and Desbrun, "Discrete Geometric Mechanics for Variational Time Integrators"). For a variational integrator, the Lagrangian is discretized by

$$\hat{L}(q_k, q_{k+1}) = \int_{t_k}^{q_{k+1}} L(q, \dot{q})\, dt. \tag{4.2}$$

For a simple one point quadrature and equidistant nodes with $q_{k+1} - q_k = h$ one obtains

$$\hat{L}(q_k, q_{k+1}) = hL\left((1-\alpha)q_k + \alpha q_{k+1}\,,\ \frac{q_{k+1} - q_k}{h}\right) \tag{4.3}$$

with $\alpha \in [0, 1]$.

TODO: Choices for α.

The discrete Euler-Lagrange equation is

$$D_1 \hat{L}(q_k, q_{k+1}) + D_2 \hat{L}(q_{k-1}, q_k) = 0. \tag{4.4}$$

Example 46 (Variational Integrator for Pendulum)**.** The Lagrangian for the pendulum is

$$L(q, \dot{q}) = \frac{1}{2} l^2 \dot{q}^2 + gl \cos\theta \tag{4.5}$$

j where g is the gravitational constant and l the length of the pendulum. By Eq. 4.3, for $\alpha = 1$ the discrete Lagrangian is then

$$\hat{L}(\theta_k, \theta_{k+1}) = \frac{1}{2} l^2 v_{k+1}^2 + gl cos(\theta_k) \tag{4.6a}$$

where the discrete velocity is given by

$$v_{k+1} = v_{k+1}(q_k, q_{k+1}) = \frac{q_{k+1} - q_k}{h} \tag{4.6b}$$

and hence depends on both q_k and q_{k+1}. For the derivatives in the discrete Euler-Lagrange4.4 we thus have

$$D_1 \hat{L}(\theta_k, \theta_{k+1}) = h \left(\frac{1}{2} l^2 \left(2\, v_{k+1} \frac{1}{h} \right) - gl \sin(\theta_k) \right) \tag{4.7a}$$

$$D_2 \hat{L}(\theta_{k-1}, \theta_k) = h \left(\frac{1}{2} l^2 \left(2\, v_k \frac{1}{h} \right) + 0 \right) \tag{4.7b}$$

and inserting into the discrete Euler-Lagrange equation hence yields

$$\left(l^2\, v_{k+1} - h\, gl \sin(\theta_k) \right) + \left(l^2 v_k \right) = 0 \tag{4.8}$$

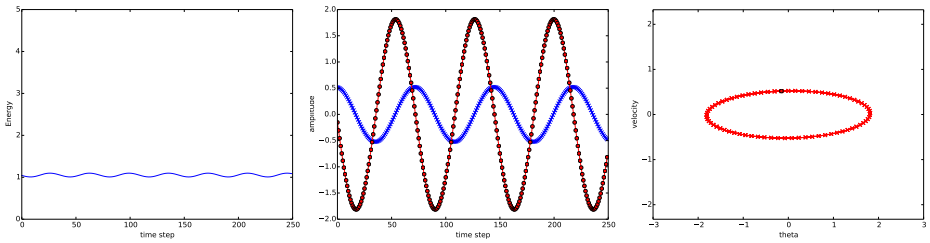
which only has one unknown, the velocity v_{k+1} at the next time step. Simplifying the equation we obtain

$$v_{k+1} = v_k + h \frac{g}{l} \sin(\theta_k) \tag{4.9a}$$

which equals an explicit Euler step for the velocity, see above. The angle at the next time step can be obtained by solving Eq. 4.6b for q_{k+1}. that is

$$\theta_{k+1} = \theta_k + h v_{k+1} \tag{4.9b}$$

which is an implicit Euler step. Eq. 4.9 provides a complete update rule for the pendulum. Numerically, the integrator yields (from left to right: energy, time series, and phase portrait)



and, in contrast to the explicit and implict Euler integrator in Fig. 4.1 and Fig. 4.2, we obtain a qualitatively correct harmonic motion. Note that Eq. 4.9 differs from the formulas for the explicit and implicit Euler schemes only slightly, but with a tremendous difference in the numerics. Also the computation is not more expensive than an explicit Euler-Integrator, the simplest possible scheme, while the nonlinear solve that is needed for the implicit Euler-scheme is avoided.

4.5 Symmetries

4.6 Further Reading

Books on geometric mechanics are for example the classic texts by Abraham and Marsden[4] and Arnold[5]. More current treatments can be found in books by Marsden and Ratiu,[6] which discusses a wealth of physical systems, and the books by Holm and co-workers[7] that provide a more pedagogical exposition.

[4] Abraham and Marsden, *Foundations of Mechanics*.

[5] Arnold, *Mathematical Methods of Classical Mechanics*.

[6] Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*.

[7] Holm, *Geometric Mechanics: Dynamics and Symmetry*; Holm, *Geometric mechanics: Rotating, translating and rolling*; Holm, Schmah, and Stoica, *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*.

Bibliography

Abraham, R. and J. E. Marsden. *Foundations of Mechanics*. second. Addison-Wesley Publishing Company, Inc., 1978. URL: http://caltechbook.library.caltech.edu/103/.

Arnold, V. I. *Mathematical Methods of Classical Mechanics*. second. Graduate Texts in Mathematics. Springer, Sept. 1989.

Batty, C., A. Uribe, B. Audoly, and E. Grinspun. "Discrete Viscous Sheets". In: *ACM Transactions on Graphics* 31.4 (July 2012), pp. 1–7. URL: http://dl.acm.org/citation.cfm?id=2185520.2185609.

Belkin, M. and P. Niyogi. "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation". en. In: *Neural Computation* 15.6 (June 2003), pp. 1373–1396. URL: http://www.mitpressjournals.org/doi/abs/10.1162/089976603321780317?journalCode=neco.

Bellman, R. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1966.

Bergou, M., B. Audoly, E. Vouga, M. Wardetzky, and E. Grinspun. "Discrete Viscous Threads". In: *ACM Transactions on Graphics* 29.4 (July 2010), p. 1. URL: http://dl.acm.org/citation.cfm?id=1778765.1778853.

Bergou, M., M. Wardetzky, S. Robinson, B. Audoly, and E. Grinspun. "Discrete Elastic Rods". In: *ACM SIGGRAPH 2008 papers on - SIGGRAPH '08*. Vol. 27. 3. New York, New York, USA: ACM Press, Aug. 2008, p. 1. URL: http://dl.acm.org/citation.cfm?id=1399504.1360662.

Botsch, M., L. Kobbelt, M. Pauly, P. Alliez, and B. Levy. *Polygon Mesh Processing*. A K Peters/CRC Press, 2010. URL: http://www.amazon.com/Polygon-Mesh-Processing-Mario-Botsch/dp/1568814267.

Bungartz, H.-J. and M. Griebel. "Sparse Grids". In: *Acta Numerica* 13 (May 2004), pp. 147–269. URL: http://journals.cambridge.org/abstract_S0962492904000182.

Candès, E. J. and B. Recht. "Exact Matrix Completion via Convex Optimization". In: *Foundations of Computational Mathematics* 9.6 (Apr. 2009), pp. 717–772. URL: http://link.springer.com/10.1007/s10208-009-9045-5.

Candès, E. J., J. Romberg, and T. Tao. "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information". In: *Information Theory, IEEE Transactions on* 52 (2006), pp. 489–509. URL: http://dx.doi.org/10.1109/TIT.2005.862083.

Candès, E. J. and M. B. Wakin. "An Introduction to Compressive Sampling". In: *Signal Processing Magazine, IEEE* 25 (2008), pp. 21–30. URL: http://dx.doi.org/10.1109/MSP.2007.914731.

Cheney, E. W. and W. A. Light. *A Course in Approximation Theory*. AMS Bookstore, 2009.

Chenney, S. and D. A. Forsyth. "Sampling Plausible Solutions to Multi-body Constraint Problems". In: *Proceedings of the 27th annual conference on Computer graphics and*

interactive techniques - SIGGRAPH '00. New York, New York, USA: ACM Press, July 2000, pp. 219–228. URL: http://dl.acm.org/citation.cfm?id=344779.344882.

Christensen, P. H. "Adjoints and Importance in Rendering: an Overview". In: *IEEE Transactions on Visualization and Computer Graphics* 9 (2003), pp. 329–340.

Coifman, R. R., S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Diffusion Maps". In: *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005), pp. 7426–7431. URL: http://www.pnas.org/content/102/21/7426.abstract.

— "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Multiscale Methods". In: *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005), pp. 7432–7437. URL: http://www.pnas.org/content/102/21/7432.abstract.

Cucker, F. and D.-X. Zhou. *Learning Theory: An Approximation Theory Viewpoint*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2007.

Davis, P. J. *Interpolation and Approximation*. Dover books on advanced mathematics. Dover Publications, 1975.

DeVore, R. *Capturing Functions in High Dimensions*. Tech. rep. Paris: Course Notes, 2009.

DeVore, R. A. "Nonlinear Approximation". In: *Acta Numerica* 7 (Jan. 1998), pp. 51–150. URL: http://journals.cambridge.org/abstract_S0962492900002816.

DeVore, R. A. and G. G. Lorentz. *Constructive approximation*. Springer, 1993. URL: http://books.google.com/books?id=cDqNW6k7_ZwC\&pgis=1.

Dick, J. and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, 2010.

Donoho, D. L. "Compressed Sensing". In: *IEEE Transactions on Information Theory* 52 (2006), pp. 1289–1306.

— "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality". In: *Mathematical Challenges of the 21st Century*. Los Angeles, CA, USA: American Mathematical Society, 2000. URL: http://www-stat.stanford.edu/~donoho/Lectures/CBMS/Curses.pdf.

Donoho, D. L., A. Maleki, I. U. Rahman, M. Shahram, and V. Stodden. "Reproducible Research in Computational Harmonic Analysis". In: *Computing in Science and Engineering* 11.1 (2009), pp. 8–18.

Elcott, S., Y. Tong, E. Kanso, P. Schröder, and M. Desbrun. "Stable, Circulation-Preserving, Simplicial Fluids". In: *ACM Transactions on Graphics* 26.1 (2007).

Finkelstein, A. and D. H. Salesin. "Multiresolution Curves". In: *SIGGRAPH '94: Proceedings of the 21st annual {C}onference on {C}omputer {G}raphics and {I}nteractive {T}echniques*. New York, NY, USA: ACM Press, 1994, pp. 261–268.

Freeden, W., T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere (With Applications to Geomathematics)*. Oxford Sciences Publication. Clarendon Press, Oxford University, 1998.

Galerkin, B. G. "On Electrical Circuits for the Approximate Solution of the Laplace Equation". Russian. In: *Vestnik Inzh.* 19 (1915), pp. 897–908.

Golub, G. H. and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in Mathematical Sciences. The Johns Hopkins University Press, 1996.

Gortler, S. J., P. Schröder, M. F. Cohen, and P. Hanrahan. "Wavelet Radiosity". In: *Computer Graphics Proceedings, Annual Conference Series, 1993 (ACM SIGGRAPH '93 Proceedings)*. 1993, pp. 221–230.

Grinstead, Charles M and J Laurie Snell. *Introduction to Probability*. American Mathematical Society, 2006.

Hesse, K. "Complexity of numerical integration over spherical caps in a Sobolev space setting". In: *Journal of Complexity* In press (Sept. 2010). URL: http://dx.doi.org/10.1016/j.jco.2010.09.002.

Holm, D. D. *Geometric Mechanics: Dynamics and Symmetry*. Vol. Part I. London: Imperial College Press, 2008.

— *Geometric mechanics: Rotating, translating and rolling*. Vol. II. Imperial College Press, 2008.

Holm, D. D., T. Schmah, and C. Stoica. *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*. Oxford texts in applied and engineering mathematics. Oxford University Press, 2009, 515 p.

Horn, R. A. and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

Kazhdan, M., T. Funkhouser, and S. Rusinkiewicz. "Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors". In: *ACM International Conference Proceeding Series; Vol. 43* (2003). URL: http://portal.acm.org/citation.cfm?id=882370.882392.

Kobilarov, M., K. Crane, and M. Desbrun. "Lie Group Integrators for Animation and Control of Vehicles". In: *ACM Transactions on Graphics (TOG)* 28.2 (2009). URL: http://portal.acm.org/citation.cfm?id=1516522.1516527.

Kovacevic, J. and A. Chebira. "Life Beyond Bases: The Advent of Frames (Part I)". In: *Signal Processing Magazine, IEEE* 24.4 (2007), pp. 86 –104. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4286567.

— "Life Beyond Bases: The Advent of Frames (Part II)". In: *IEEE Signal Processing Magazine* 24.5 (2007), pp. 115–125. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4350231.

Lax, P. D. *Functional Analysis*. New York: Wiley-Interscience, 2002.

— *Linear Algebra and Its Applications*. second ed. Pure and Applied Mathematics. Wiley Interscience, 2007.

Lax, P. D. and A. N. Milgram. "Parabolic equations". In: *Contributions to the theory of partial differential equations*. Annals of Mathematics Studies. Princeton, N. J.: Princeton University Press, 1954, pp. 167–190.

Lehtinen, J. "A Framework for Precomputed and Captured Light Transport". In: *ACM Trans. Graph.* 26 (2007), p. 13.

Lehtinen, J., M. Zwicker, E. Turquin, J. Kontkanen, F. Durand, F. X. Sillion, and T. Aila. "A Meshless Hierarchical Representation for Light Transport". In: *ACM Trans. Graph.* 27 (2008), pp. 1–9.

Lessig, C. "Modern Foundations of Light Transport Simulation". Ph.D. thesis. Toronto: University of Toronto, 2012. URL: http://users.cms.caltech.edu/~lessig/dissertation/lessig_dissertation.pdf.

Lipman, Y., X. Chen, I. Daubechies, and T. Funkhouser. "Symmetry Factored Embedding and Distance". In: *ACM SIGGRAPH 2010 papers on - SIGGRAPH '10*. Vol. 29. 4. New York, New York, USA: ACM Press, July 2010, p. 1. URL: http://dl.acm.org/citation.cfm?id=1833349.1778840.

Mallat, S. G. *A Wavelet Tour of Signal Processing: The Sparse Way*. third ed. Academic Press, 2009.

Mandel, L. and E. Wolf. *Optical Coherence and Quantum Optics*. Cambridge, UK: Cambridge University Press, Sept. 1995.

Marsden, J. E. and T. S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. third ed. Texts in Applied Mathematics. New York: Springer-Verlag.

Marsden, J. E., T. S. Ratiu, and R. Abraham. *Manifolds, Tensor Analysis, and Applications*. third. Applied Mathematical Sciences. New York: Springer-Verlag, 2004.

Misner, C. W., K. S. Thorne, and J. A. Wheeler. *Gravitation*. San Francisco, CA, USA: W. H. Freeman, 1973.

Mitchell, D. P. "Spectrally Optimal Sampling for Distribution Ray Tracing". In: *SIGGRAPH Comput. Graph.* 25 (1991), pp. 157–164.

Mullen, P., K. Crane, D. Pavlov, Y. Tong, and M. Desbrun. "Energy-Preserving Integrators for Fluid Animation". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2009)* 28.3 (2009), pp. 1–8.

Öztireli, A. C., M. Alexa, and M. Gross. "Spectral Sampling of Manifolds". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2010)* 29.6 (Dec. 2010), p. 1. URL: http://dl.acm.org/citation.cfm?id=1882262.1866190.

Pavlov, D., P. Mullen, Y. Tong, E. Kanso, J. E. Marsden, and M. Desbrun. "Structure-preserving discretization of incompressible fluids". In: *Physica D: Nonlinear Phenomena* 240.6 (Mar. 2011), pp. 443–458. URL: http://dx.doi.org/10.1016/j.physd.2010.10.012.

Petrov, G. I. "Application of the Method of Galerkin to a Problem Involving the Stationary Flow of a Viscous Fluid". In: *Prikl. Matem. Mekh.* 4.3 (1940).

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. New York, NY, USA: Cambridge University Press, 1992.

Ramamoorthi, R. "Precomputation-Based Rendering". In: *Foundations and Trends in Computer Graphics and Vision* 3.4 (2007), pp. 281–369.

Reimer, M. *Multivariate Polynomial Approximation*. International Series of Numerical Mathematics. Basel: Birkhäuser Verlag, 2003.

Rudin, W. *Functional Analysis*. International Series in Pure and Applied Mathematics. New York: McGraw-Hill Inc., 1991.

— *Principles of Mathematical Analysis*. International Series in Pure & Applied Mathematics. McGraw-Hill Publishing Co., Sept. 1976.

— *Real and Complex Analysis*. McGraw-Hill International Editions: Mathematics Series. McGraw-Hill Education - Europe, Mar. 1987.

Schröder, P., S. J. Gortler, M. F. Cohen, and P. Hanrahan. "Wavelet Projections for Radiosity". In: *Fourth Eurographics Workshop on Rendering*. Paris, France, June 1993, pp. 105–114.

Schröder, P. and W. Sweldens. "Spherical Wavelets: Efficiently Representing Functions on the Sphere". In: *Proceedings of ACM SIGGRAPH 1995*. New York, NY, USA: ACM Press, 1995, pp. 161–172.

— "Spherical Wavelets: Texture Processing". In: *Rendering Techniques '95*. Ed. by P Hanrahan and W Purgathofer. Wien, New York: Springer Verlag, Aug. 1995, pp. 252–263.

Sorkine, O. "Laplacian Mesh Processing". In: *Eurographics 2005, State of the Art Reports*. 2005, pp. 53–70.

Spivak, M. *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. HarperCollins Publishers, 1971.

Stakgold, I. and M. J. Holst. *Green's Functions and Boundary Value Problems*. third ed. Pure and Applied Mathematics. John Wiley \& Sons, Inc., 2011.

Stern, A. and M. Desbrun. "Discrete Geometric Mechanics for Variational Time Integrators". In: *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses*. ACM, 2006. URL: http://portal.acm.org/citation.cfm?id=1185657.1185669.

Taubin, G. "A Signal Processing Approach to Fair Surface Design". In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques - SIGGRAPH '95*. New York, New York, USA: ACM Press, Sept. 1995, pp. 351–358. URL: http://dl.acm.org/citation.cfm?id=218380.218473.

Temlyakov, V. N. "Nonlinear Methods of Approximation". In: *Foundations of Computational Mathematics* 3.1 (Jan. 2003), pp. 33–107. URL: http://link.springer.com/10.1007/s102080010029.

Trefethen, L. N. *Approximation Theory and Approximation Practice*. Society for Industrial and Applied Mathematics, 2012.

Trenogin, V. A. "Galerkin Method". In: *Encyclopedia of Mathematics*. Ed. by M. Hazewinkel. Kluwer Academic Publishers, 2002. URL: http://www.encyclopediaofmath.org/index.php?title=Galerkin_method\&oldid=16286.

Tsai, Yu-Ting and Zen-Chung Shih. "All-Frequency Precomputed Radiance Transfer Using Spherical Radial Basis Functions and Clustered Tensor Approximation". In: *Transactions of Graphics (Proceedings of Siggraph 2006)* 25.3 (Proceedings of ACM SIGGRAPH 2006) (2006). URL: http://portal.acm.org/citation.cfm?id=1179352.1141981.

Vasilescu, M. A. O. and D. Terzopoulos. "Multilinear Analysis of Image Ensembles: TensorFaces". In: *Computer Vision-ECCV 2002*. Ed. by Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen. Vol. 2350. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, Apr. 2006. URL: http://www.springerlink.com/index/10.1007/3-540-47969-4.

— "TensorTextures". In: *ACM Transactions on Graphics* 23.3 (Aug. 2004), p. 336. URL: http://dl.acm.org/citation.cfm?id=1015706.1015725.

Veach, E. and L. J. Guibas. "Metropolis Light Transport". In: *SIGGRAPH '97: Proceedings of the 24th annual conference on computer graphics and interactive techniques*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 65–76.

Wendland, H. *Scattered Data Approximation*. Cambridge monographs on applied and computational mathematics. Cambridge University Press, 2005.

Zatz, H. R. "Galerkin Radiosity: A Higher Order Solution Method for Global Illumination". In: *International Conference on Computer Graphics and Interactive Techniques* (1993). URL: http://portal.acm.org/citation.cfm?id=166145.

Index

Symbols

A

B

C

D

E

F

G

H

I

K

‘‘My work always tried to unite
the truth with the beautiful,
but when I had to choose
one or the other,
I usually chose the beautiful.’’

Hermann Weyl

© Christian Lessig 2014