Chapter 3

Linear transformations

Linear transformations, or linear maps, between vector spaces represent an important class of functions, in their own right, but also as approximations of more general nonlinear transformations.

A linear transformation acting on a Euclidian vector can be represented by a matrix. Many of the concepts we introduce in this chapter generalize to linear transformations acting on functions in infinite dimensional spaces, for example integral and differential operators, which are fundamental for the study of differential equations.

3.1 Matrix algebra

Linear transformation as a matrix

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ defines a *linear transformation*, if

(i) $f(x + z) = f(x) + f(z)$,

(ii) $f(\alpha x) = \alpha f(x)$,

for all $x, z \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. In the case that $m = n$, we refer to $f : \mathbb{R}^n \to \mathbb{R}^n$ as a *linear operator* in $\mathbb{R}^n$. In the standard basis $\{e_1, ..., e_n\}$ we can express the ith component of the vector $y = f(x) \in \mathbb{R}^m$ as

$$y_i = f_i(x) = f_i(\sum_{j=1}^{n} x_j e_j) = \sum_{j=1}^{n} x_j f_i(e_j),$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ for all $i = 1, ..., m$. In component form, we write this as

$$\begin{aligned} y_1 &= a_{11}x_1 + ... + a_{1n}x_n \\ &\vdots \\ y_m &= a_{m1}x_1 + ... + a_{mn}x_n \end{aligned} \tag{3.1}$$

with $a_{ij} = f_i(e_j)$. That is $y = Ax$, where A is an $m \times n$ *matrix*,

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}. \tag{3.2}$$

The set of real valued $m \times n$ matrices defines a vector space $\mathbb{R}^{m\times n}$, by the basic operations of (i) component-wise matrix addition, and (ii) component-wise scalar multiplication, that is

$$A + B = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}, \quad \alpha A = \begin{bmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & \ddots & \vdots \\ \alpha a_{m1} & \cdots & \alpha a_{mn} \end{bmatrix},$$

with $A, B \in \mathbb{R}^{m\times n}$ and $\alpha \in \mathbb{R}$.

Matrix-vector product

A matrix $A \in \mathbb{R}^{m\times n}$ defines a *linear map* $x \mapsto Ax$, by the operations of *matrix-vector product* and *component-wise scalar multiplication*,

$$\begin{aligned} A(x + y) &= Ax + Ay, & x, y \in \mathbb{R}^n, \\ A(\alpha x) &= \alpha Ax, & x \in \mathbb{R}^n, \alpha \in \mathbb{R}. \end{aligned}$$

In *index notation* we write a vector $b = (b_i)$, and a matrix $A = (a_{ij})$, with i the *row index* and j is the *column index*. For an $m \times n$ matrix A, and x an n-dimensional column vector, we define the matrix-vector product $b = Ax$ to be the m-dimensional column vector $b = (b_i)$, such that

$$b_i = \sum_{j=1}^{n} a_{ij}x_j, \quad i = 1, ..., m. \tag{3.3}$$

With a_j the jth column of A, an m-dimensional column vector, we can express the matrix-vector product as a linear combination of the set of column vectors $\{a_j\}_{j=1}^n$,

$$b = Ax = \sum_{j=1}^{n} x_j a_j, \tag{3.4}$$

or in matrix form

$$\begin{bmatrix} \\ b \\ \\ \end{bmatrix} = \left[\begin{array}{c|c|c|c} & & & \\ a_1 & a_2 & \cdots & a_n \\ & & & \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} \\ a_1 \\ \\ \end{bmatrix} + x_2 \begin{bmatrix} \\ a_2 \\ \\ \end{bmatrix} + ... + x_n \begin{bmatrix} \\ a_n \\ \\ \end{bmatrix}.$$

The vector space spanned by $\{a_j\}_{j=1}^n$ is the *column space*, or *range*, of the matrix A, so that $\text{range}(A) = \text{span}\{a_j\}_{j=1}^n$. The *null space*, or *kernel*, of an $m \times n$ matrix A is the set of vectors $x \in \mathbb{R}^n$ such that $Ax = 0$, with 0 the zero vector in $\mathbb{R}^m$, that is $\text{null}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$.

The dimension of the column space is the column *rank* of the matrix A, $\text{rank}(A)$. We note that the column rank is equal to the row rank, corresponding to the space spanned by the row vectors of A, and the maximal rank of an $m \times n$ matrix is $\min(m, n)$, which we refer to as *full rank*.

Matrix-matrix product

The *matrix-matrix product* $B = AC$ is a matrix in $\mathbb{R}^{l\times n}$, defined for two matrices $A \in \mathbb{R}^{l\times m}$ and $C \in \mathbb{R}^{m\times n}$, as

$$b_{ij} = \sum_{k=1}^{m} a_{ik} c_{kj}, \tag{3.5}$$

with $B = (b_{ij})$, $A = (a_{ik})$ and $C = (c_{kj})$.

We sometimes omit the summation sign and use the *Einstein convention*, where repeated indices imply summation over those same indices, so that we express the matrix-matrix product (3.5) simply as $b_{ij} = a_{ik}c_{kj}$.

Similarly as for the matrix-vector product, we may interpret the columns b_j of the matrix-matrix product B as a linear combination of the columns a_k with coefficients c_{kj}

$$b_j = Ac_j = \sum_{k=1}^{m} c_{kj} a_k, \tag{3.6}$$

or in matrix form

$$\left[\begin{array}{c|c|c|c} & & & \\ b_1 & b_2 & \cdots & b_n \\ & & & \end{array}\right] = \left[\begin{array}{c|c|c|c} & & & \\ a_1 & a_2 & \cdots & a_m \\ & & & \end{array}\right] \left[\begin{array}{c|c|c|c} c_1 & c_2 & \cdots & c_n \end{array}\right].$$

The composition $f \circ g(x) = f(g(x))$, of two linear transformations $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^l$, with associated matrices $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{l\times w}$, corresponds to the matrix-matrix product AC acting on $x \in \mathbb{R}^n$.

Matrix transpose and the inner and outer products

The *transpose* (or *adjoint*) of an $m \times n$ matrix $A = (a_{ij})$ is defined as the matrix $A^T = (a_{ji})$, with the column and row indices reversed.

Using the matrix transpose, the inner product of two vectors $v, w \in \mathbb{R}^n$ can be expressed in terms of a matrix-matrix product $v^T w$, as

$$(v, w) = v^T w = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} = v_1 w_1 + ... + v_n w_n. \tag{3.7}$$

Similarly, the *outer product*, or *tensor product*, of two vectors $v, w \in \mathbb{R}^n$, denoted by $v \otimes w$, is defined as the $m \times n$ matrix corresponding to the matrix-matrix product vw^T, that is

$$v \otimes w = vw^T = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} = \begin{bmatrix} v_1 w_1 & \cdots & v_1 w_n \\ \vdots & & \vdots \\ v_m v_1 & & v_m w_n \end{bmatrix}.$$

In tensor notation we can express the inner and the outer products as $(v, w) = v_i w_i$ and $v \otimes w = v_i w_j$, respectively.

The transpose has the property that $(AB)^T = B^T A^T$, and thus satisfies the equation $(Ax, y) = (x, A^T y)$, for any $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$, which follows from the definition of the inner product in Euclidian vector spaces, since

$$(Ax, y) = (Ax)^T y = x^T A^T y = (x, A^T y). \tag{3.8}$$

A square matrix $A \in \mathbb{R}^{n\times n}$ is said to be *symmetric* (or *self-adjoint*) if $A = A^T$, which gives that $(Ax, y) = (x, Ay)$. If in addition $(Ax, x) > 0$ for all non-zero $x \in \mathbb{R}^n$, we say that A is a *symmetric positive definite* matrix. A square matrix is said to be *normal* if $A^T A = AA^T$.

Matrix norms

To measure the size of a matrix, we first introduce the *Frobenius norm*, corresponding to the l_2-norm of the matrix A interpreted as an mn-vector, that is

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}. \tag{3.9}$$

The Frobenius norm is induced by the following inner product over the space $\mathbb{R}^{m\times n}$,

$$(A, B) = \operatorname{tr}(A^T B), \tag{3.10}$$

with the *trace* of a square $n \times n$ matrix $C = (c_{ij})$ defined by

$$\operatorname{tr}(C) = \sum_{i=1}^{n} c_{ii}. \tag{3.11}$$

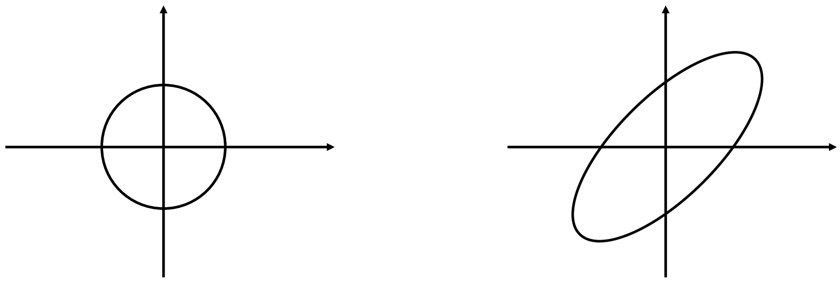

Figure 3.1: Illustration of the map $x \mapsto Ax$; of the unit circle $\|x\|_2 = 1$ (left) to the ellipse Ax (right), corresponding to the matrix A in (3.13).

Matrix norms for $A \in \mathbb{R}^{m\times n}$ are also induced by the respective l_p-norms on $\mathbb{R}^m$ and $\mathbb{R}^n$, in the form

$$\|A\|_p = \sup_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{\|Ax\|_p}{\|x\|_p} = \sup_{\substack{x\in\mathbb{R}^n \\ \|x\|_p=1}} \|Ax\|_p. \tag{3.12}$$

The last equality follows from the definition of a norm, and shows that the induced matrix norm can be defined in terms of its map of unit vectors, which we illustrate in Figure 3.1 for the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}. \tag{3.13}$$

We further have the following inequality,

$$\|Ax\|_p \leq \|A\|_p \|x\|_p, \tag{3.14}$$

which follows from (3.12).

Determinant

The *determinant* of a square matrix A is denoted $\det(A)$ or $|A|$. For a 2×2 matrix we have the explicit formula

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \tag{3.15}$$

For example, the determinant of the matrix A in (3.13) is computed as $\det(A) = 1 \cdot 2 - 2 \cdot 0 = 2$.

The formula for the determinant is extended to a 3×3 matrix by

$$\begin{aligned} \det(A) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= a(ei - fh) - b(di - fg) + c(dh - eg), \end{aligned} \tag{3.16}$$

and by recursion this formula can be generalized to any square matrix.

For a 2×2 matrix the absolute value of the determinant is equal to the area of the parallelogram that represents the image of the unit square under the map $x \mapsto Ax$, and similarly for a 3×3 matrix the volume of the parallelepiped representing the mapped unit cube. More generally, the absolute value of the determinant $\det(A)$ represents a scale factor of the linear transformation A.

Matrix inverse

If the column vectors $\{a_j\}_{j=1}^n$ of a square $n \times n$ matrix A form a basis for $\mathbb{R}^n$, then all vectors $b \in \mathbb{R}^n$ can be expressed as $b = Ax$, where the vector $x \in \mathbb{R}^n$ holds the coordinates of b in the basis $\{a_j\}_{j=1}^n$.

In particular, all $x \in \mathbb{R}^n$ can be expressed as $x = Ix$, where I is the square $n \times n$ *identity matrix* in $\mathbb{R}^n$, taking the standard basis as column vectors,

$$x = Ix = \left[\begin{array}{c|c|c|c} \\ e_1 & e_2 & \cdots & e_n \\ \\ \end{array} \right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

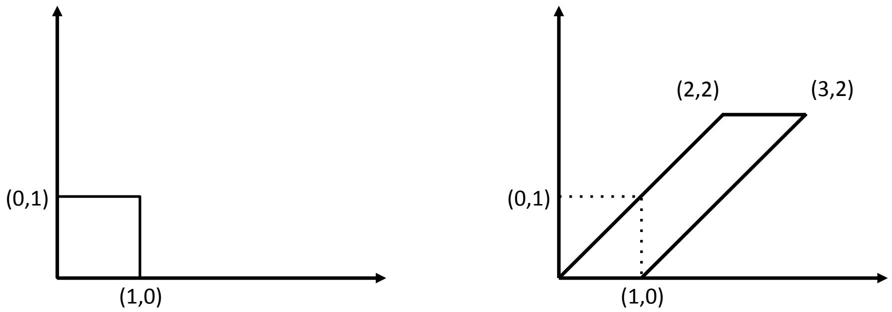

Figure 3.2: The map $x \mapsto Ax$ (right) of the unit square (left), for the matrix A in (3.13), with the corresponding area given as $|\det(A)| = 2$.

with the vector entries x_i corresponding to the Cartesian coordinates of the vector x.

A square matrix $A \in \mathbb{R}^{n\times n}$ is *invertible*, or *non-singular*, if there exists an *inverse matrix* $A^{-1} \in \mathbb{R}^{n\times n}$, such that

$$A^{-1}A = AA^{-1} = I, \tag{3.17}$$

which also means that $(A^{-1})^{-1} = A$. Further, for two $n \times n$ matrices A and B, we have the property that $(AB)^{-1} = B^{-1}A^{-1}$.

Theorem 4 (Inverse matrix)**.** *For a square matrix $A \in \mathbb{R}^{n\times n}$, the following statements are equivalent:*

(i) *A has an inverse A^{-1},*

(ii) $\det(A) \neq 0$*,*

(iii) $\text{rank}(A) = n$*,*

(iv) $\text{range}(A) = \mathbb{R}^n$

(v) $\text{null}(A) = \{0\}$*.*

The matrix inverse is unique. To see this, assume that there exist two matrices B_1 and B_2 such that $AB_1 = AB_2 = I$; which by linearity gives that $A(B_1 - B_2) = 0$, but since $\text{null}(A) = \{0\}$ we have that $B_1 = B_2$.

3.2 Orthogonal projectors

Orthogonal matrix

A square matrix $Q \in \mathbb{R}^{n\times n}$ is *ortogonal*, or *unitary*, if $Q^T = Q^{-1}$. With q_j the columns of Q we thus have that $Q^T Q = I$, or in matrix form,

$$\begin{bmatrix} \quad q_1 \quad \\ \hline q_2 \\ \hline \vdots \\ \hline q_n \end{bmatrix} \begin{bmatrix} \\ q_1 & \Big| & q_2 & \Big| & \cdots & \Big| & q_n \\ \\ \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix},$$

so that the columns q_j form an orthonormal basis for $\mathbb{R}^n$.

Multiplication by an orthogonal matrix preserves the angle between two vectors $x, y \in \mathbb{R}^n$, since

$$(Qx, Qy) = (Qx)^T Qy = x^T Q^T Qy = x^T y = (x, y), \tag{3.18}$$

and thus also the length of a vector,

$$\|Qx\| = (Qx, Qx)^{1/2} = (x, x)^{1/2} = \|x\|. \tag{3.19}$$

For example, counter-clockwise rotation by an angle θ in $\mathbb{R}^2$, takes the form of multiplication by an orthogonal matrix,

$$Q_{rot} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \tag{3.20}$$

whereas reflection in the line with a slope given by the angle θ, corresponds to multiplication by the orthogonal matrix,

$$Q_{ref} = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}, \tag{3.21}$$

where we note the general fact that for a rotation $\det(Q_{rot}) = 1$, and for a reflection $\det(Q_{ref}) = -1$.

Orthogonal projector

A *projection matrix*, or *projector*, is a square matrix P such that

$$P^2 = PP = P. \tag{3.22}$$

It follows that

$$Pv = v, \tag{3.23}$$

for all vectors $v \in \text{range}(P)$, since v is of the form $v = Px$ for some x, and thus $Pv = P^2x = Px = v$. For $v \notin \text{range}(P)$ we have that $P(Pv - v) = P^2v - Pv = 0$, so that the projection error $Pv - v \in \text{null}(P)$.

The matrix $I - P$ is also a projector, the *complementary projector* to P, since $(I - P)^2 = I - 2P + P^2 = I - P$. The range and null space of the two projectors are related as

$$\text{range}(I - P) = \text{null}(P), \tag{3.24}$$

and

$$\text{range}(P) = \text{null}(I - P), \tag{3.25}$$

so that P and $I - P$ separates $\mathbb{R}^n$ into the two subspaces $S_1 = \text{range}(P)$ and $S_2 = \text{range}(I - P)$, since the only $v \in \text{range}(P) \cap \text{range}(I - P)$ is the zero vector; $v = v - Pv = (I - P)v = \{0\}$.

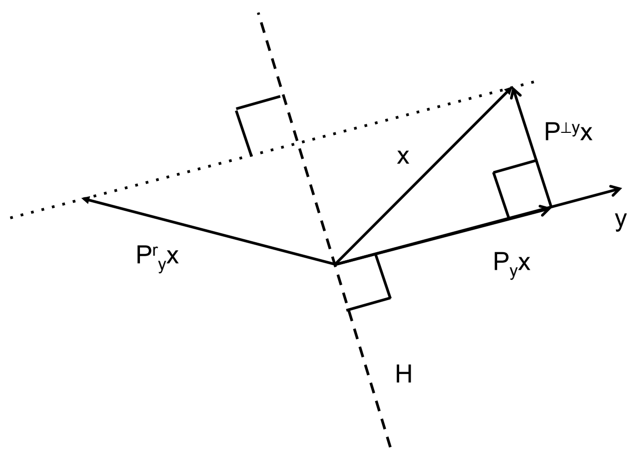

Figure 3.3: The projector P_yx of a vector x in the direction of another vector y, its orthogonal complement $P^{\perp y}x$, and $P_y^r x$, the reflector of x in the hyperplane H defined by y as a normal vector.

If the two subspaces S_1 and S_2 are orthogonal, we say that P is an *orthogonal projector*. This is equivalent to the condition $P = P^T$, since the inner product between two vectors in S_1 and S_2 then vanish,

$$(Px, (I - P)y) = (Px)^T(I - P)y = x^T P^T (I - P)y = x^T(P - P^2)y = 0,$$

and if P is an orthogonal projector, so is $I - P$.

For example, the orthogonal projection of one vector x in the direction of another vector y, expressed in (2.19), corresponds to an orthogonal projector P_y, by

$$\frac{(x,y)y}{\|y\|^2} = \frac{y(y,x)}{\|y\|^2} = \frac{y(y^T x)}{\|y\|^2} = \frac{yy^T}{\|y\|^2} x = P_y x. \tag{3.26}$$

Similarly we can define the orthogonal complement $P^{\perp y}x$, and $P_y^r x$, the reflection of x in the *hyperplane* H defined by y as a normal vector, so that

$$P_y = \frac{yy^T}{\|y\|^2}, \quad P^{\perp y} = I - \frac{yy^T}{\|y\|^2}, \quad P_y^r = I - 2\frac{yy^T}{\|y\|^2}, \tag{3.27}$$

defines orthogonal projectors, where we note that a hyperplane is a subspace in V of codimension 1.

3.3 Exercises

Problem 10. *Prove the equivalence of the definitions of the induced matrix norm, defined by*

$$\|A\|_p = \sup_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{\|Ax\|_p}{\|x\|_p} = \sup_{\substack{x\in\mathbb{R}^n \\ \|x\|_p=1}} \|Ax\|_p. \tag{3.28}$$

Problem 11. *For $A \in \mathbb{R}^{m\times l}, B \in \mathbb{R}^{l\times n}$, prove that $(AB)^T = B^T A^T$.*

Problem 12. *For $A, B \in \mathbb{R}^{n\times n}$, prove that $(AB)^{-1} = B^{-1}A^{-1}$.*

Problem 13. *Prove the inequality (3.14).*

Problem 14. *Prove that an orthogonal matrix is normal.*

Problem 15. *Show that the matrices A and B are orthogonal and compute their determinants. Which matrix represents a rotation and reflection, respectively?*

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad B = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix} \tag{3.29}$$

Problem 16. *For P a projector, prove that* $\text{range}(I - P) = \text{null}(P)$, *and that* $\text{range}(P) = \text{null}(I - P)$.

Problem 17. *For the vector $y = (1, 0)^T$, compute the action of the projectors $P_y, P^{\perp y}, P_y^r$ on a general vector $x = (x_1, x_2)^T$.*

Chapter 4

Linear operators in $\mathbb{R}^n$

In this chapter we give some examples of linear operators in the vector space $\mathbb{R}^n$, used extensively in various fields, including computer graphics, robotics, computer vision, image processing, and computer aided design.

We also meet differential equations for the first time, in the form of matrix operators acting on discrete approximations of functions, defined by their values at the nodes of a grid.

4.1 Approximation of differential equations

Difference and summation matrices

Subdivide the interval $[0, 1]$ into a *structured grid* $\mathcal{T}^h$ with n intervals and $n + 1$ *nodes* x_i, such that $0 = x_0 < x_1 < x_2 < ... < x_n = 1$, with a constant interval length, or *grid size*, $h = x_i - x_{i-1}$ for all i, so that $x_i = x_0 + ih$.

For each $x = x_i$, we may approximate the primitive function $F(x)$ of a function $f(x)$, expressed here as a definite integral with $f(0) = 0$, by

$$F(x_i) = \int_0^{x_i} f(s)ds \approx \sum_{k=1}^{i} f(x_k)h \equiv F_h(x_i), \quad (4.1)$$

which defines a function $F_h(x_i) \approx F(x_i)$ for all nodes x_i in the subdivision, based on *Riemann sums*.

The function F_h defines a linear transformation L_h of the vector of sampled function values at the nodes $y = (f(x_1), ..., f(x_n))^T$, which can be expressed by the following matrix equation,

$$L_h y = \begin{bmatrix} h & 0 & \cdots & 0 \\ h & h & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ h & h & \cdots & h \end{bmatrix} \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} f(x_1)h \\ f(x_1)h + f(x_2)h \\ \vdots \\ \sum_{k=1}^n f(x_k)h \end{bmatrix}, \tag{4.2}$$

where L_h is a summation matrix, with an associated inverse L_h^{-1},

$$L_h = h \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad \Rightarrow \quad L_h^{-1} = h^{-1} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & -1 & 1 \end{bmatrix}. \tag{4.3}$$

The inverse matrix L_h^{-1} corresponds to a difference matrix over the same subdivision $\mathcal{T}^h$, approximating the slope (derivative) of the function $f(x)$. To see this, multiply the matrix L_h^{-1} to $y = (f(x_i))$,

$$L_h^{-1} y = h^{-1} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} (f(x_1) - f(x_0))/h \\ (f(x_2) - f(x_1))/h \\ \vdots \\ (f(x_n) - f(x_{n-1}))/h \end{bmatrix},$$

where we recall that $f(x_0) = f(0) = 0$.

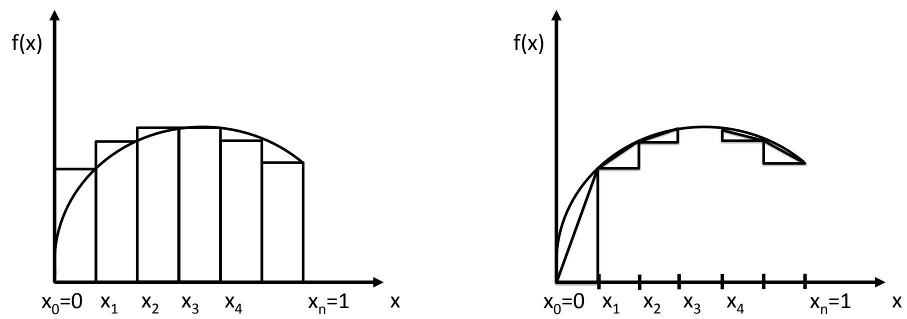

Figure 4.1: Approximation of the integral of a function $f(x)$ in the form of Riemann sums (left), and approximation of the derivative of $f(x)$ by slopes computed from function values in the nodes x_i (right), on a subdivision of $[0, 1]$ with interval length h.

As the interval length $h \to 0$, the summation and difference matrices converge to integral and differential operators, such that for each $x \in (0,1)$,

$$L_h y \to \int_0^x f(s)ds, \quad L_h^{-1} y \to f'(x). \tag{4.4}$$

Further, we have for the product of the two matrices that

$$y = L_h L_h^{-1} y \to f(x) = \int_a^x f'(s)ds, \tag{4.5}$$

as $h \to 0$, which corresponds to the *Fundamental theorem of Calculus.*

Difference operators

The matrix L_h^{-1} in (4.3) corresponds to a backward difference operator D_h^-, and similarly we can define a forward difference operator D_h^+, by

$$D_h^- = h^{-1} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \quad D_h^+ = h^{-1} \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \\ 0 & \cdots & 0 & 0 & -1 \end{bmatrix}.$$

The matrix-matrix product $D_h^+ D_h^-$ takes the form,

$$D_h^+ D_h^- = h^{-2} \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & 1 & -2 & 1 \\ 0 & \cdots & 0 & 1 & -2 \end{bmatrix}, \tag{4.6}$$

which corresponds to an approximation of a second order differential operator. The matrix $A = -D_h^+ D_h^-$ is *diagonally dominant*, that is

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|, \tag{4.7}$$

and symmetric positive definite, since

$$\begin{aligned} x^T A x &= ... + x_i(-x_{i-1} + 2x_i - x_{i+1}) + ... + x_n(-x_{n-1} + 2x_n) \\ &= ... - x_i x_{i-1} + 2x_i^2 - x_i x_{i+1} - x_{i+1} x_i + ... - x_{n-1} x_n + 2x_n^2 \\ &= ... + (x_i - x_{i-1})^2 + (x_{i+1} - x_i)^2 + ... + x_n^2 > 0, \end{aligned}$$

for any non-zero vector x.

The finite difference method

For a vector $y = (u(x_i))$, the ith row of the matrix $D_h^+ D_h^-$ corresponds to a *finite difference stencil*, with $u(x_i)$ function values sampled at the nodes x_i of the structured grid representing the subdivision of the interval $I = (0,1)$,

$$[(D_h^+ D_h^-)y]_i = \frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1})}{h^2} = \frac{\frac{u(x_{i+1}) - u(x_i)}{h} - \frac{u(x_i) - u(x_{i-1})}{h}}{h}.$$

Similarly, the difference operators D_h^- and D_h^+ correspond to finite difference stencils over the grid, and we have that for $x \in I$,

$$(D_h^+ D_h^-)y \to u''(x), \quad (D_h^-)y \to u'(x), \quad (D_h^+)y \to u'(x), \tag{4.8}$$

as the grid size $h \to 0$.

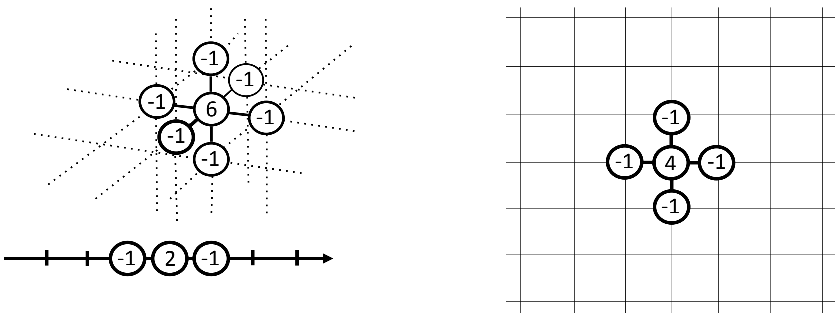

Figure 4.2: Example of finite difference stencils corresponding to the difference operator $-(D_h^+ D_h^-)$ over structured grids in $\mathbb{R}$ (lower left), $\mathbb{R}^2$ (right) and $\mathbb{R}^3$ (upper left).

The *finite difference method* for solving differential equations is based on approximation of differential operators by such difference stencils over a grid. We can thus, for example, approximate the differential equation

$$-u''(x) + u(x) = f(x), \tag{4.9}$$

by the matrix equation

$$-(D_h^+ D_h^-)y + (D_h^-)y = b, \tag{4.10}$$

with $b_i = (f(x_i))$. The finite difference method extends to multiple dimensions, where the difference stencils are defined over structured (Cartesian) grids in $\mathbb{R}^2$ or $\mathbb{R}^3$, see Figure 4.2.

Solution of differential equations

Since the second order difference matrix $A = -(D_h^+ D_h^-)$ is symmetric positive definite, there exists a unique invers A^{-1}. For example, in the case of $n = 5$ and the difference matrix A below, we have that

$$A = 1/h^2 \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix} \quad \Rightarrow \quad A^{-1} = h^2/6 \begin{bmatrix} 5 & 4 & 3 & 2 & 1 \\ 4 & 8 & 6 & 4 & 2 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 8 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}.$$

The matrix A^{-1} corresponds to a symmetric integral (summation) operator, where the matrix elements decay with the distance from the diagonal. The integral operator has the property that when multiplied to a vector $y = (y_i)$, each element y_i is transformed into an average of all the vector elements of y, with most weight given to the elements close to y_i.

Further, for $y = (u(x_i))$ and $b = (f(x_i))$, the solution to the differential equation

$$-u''(x) = f(x) \tag{4.11}$$

can be approximated by

$$y = A^{-1}b. \tag{4.12}$$

We can thus compute approximate solutions for any function $f(x)$ on the right hand side of the equation (4.11). Although, we note that while the $n \times n$ matrix A is *sparse*, with only few non-zero elements near the diagonal, the inverse A^{-1} is a *full matrix* without zero elements.

In general the full matrix A^{-1} has a much larger memory footprint than the sparse matrix A. Therefore, for large matrices, it may be impossible to hold the matrix A^{-1} in memory, so that instead iterative solution methods must be used without the explicit construction of the matrix A^{-1}.

4.2 Projective geometry

Affine transformations

An *affine transformation*, or *affine map*, is a linear transformation composed with a translation, corresponding to a multiplication by a matrix A, followed by addition of a position vector b, that is

$$x \mapsto Ax + b. \tag{4.13}$$

For example, an object defined by a set of vectors in $\mathbb{R}^2$ can be scaled by a diagonal matrix, or rotated by a *Givens rotation* matrix,

$$A_{rot} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \tag{4.14}$$

with θ a counter-clockwise rotation angle.

Any triangle in the Euclidian plane $\mathbb{R}^2$ is related to each other through an invertible affine map. There is also an affine map from $\mathbb{R}^2$ to a surface (manifold) in $\mathbb{R}^3$, although this map is not invertible, see Figure 4.4.

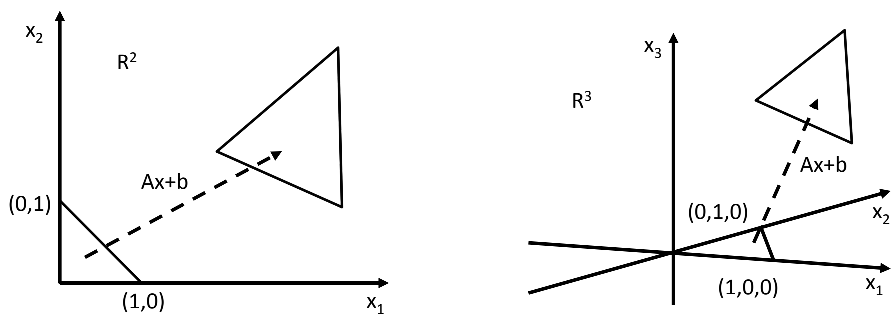

Figure 4.3: Affine maps $x \mapsto Ax + b$ of the *reference triangle*, with corners in $(0,0)$, $(1,0)$, $(0,1)$; in $\mathbb{R}^2$ (left); to a surface (manifold) in $\mathbb{R}^3$ (right).

Homogeneous coordinates

By using *homogeneous coordinates*, or *projective coordinates*, we can express any affine transformation as one single matrix multiplication, including translation. The underlying definition is that the representation of a geometric object x is homogeneous if $\lambda x = x$, for all real numbers $\lambda \neq 0$.

An $\mathbb{R}^2$ vector $x = (x_1, x_2)^T$ in standard Cartesian coordinates is represented as $x = (x_1, x_2, 1)^T$ in homogeneous coordinates, from which follows that any object $u = (u_1, u_2, u_3)$ in homogeneous coordinates can be expressed in Cartesian coordinates, by

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} u_1/u_3 \\ u_2/u_3 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} u_1/u_3 \\ u_2/u_3 \end{bmatrix}. \tag{4.15}$$

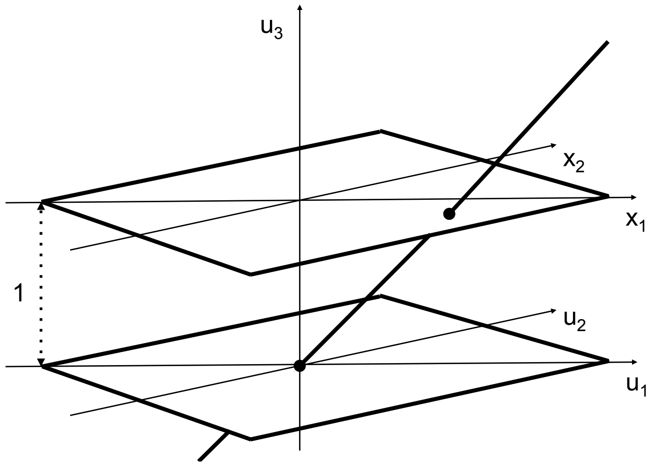

Figure 4.4: The relation of homogeneous (projective) coordinates and Cartesian coordinates.

It follows that in homogeneous coordinates, rotation by an angle θ and translation by a vector (t_1, t_2), both can be expressed as matrices,

$$A_{rot} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_{trans} = \begin{bmatrix} 1 & 0 & t_1 \\ 0 & 1 & t_2 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.16}$$

An advantage of homogenous coordinates is also the ability to apply combinations of transformations by multiplying the respective matrices, which is used extensively e.g. in robotics, computer vision, and computer graphics. For example, an affine transformation can be expressed by the matrix-matrix product $A_{trans}A_{rot}$.

4.3 Computer graphics

Vector graphics

Vector graphics is based the representation of primitive objects defined by a set of parameters, such as a circle in $\mathbb{R}^2$ defined by its center and radius, or a cube in $\mathbb{R}^3$ defined by its corner points. Lines and polygones are other common objects, and for special purposes more advances objects are used, such as NURBS (Non-uniform rational B-splines) for computer aided design (CAD), and PostScript fonts for digital type setting.

These objects my be characterized by their parameter values in the form of vectors in $\mathbb{R}^n$, and operations on such objects can be defined by affine transformations acting on the vectors of parameters.

Raster graphics

Whereas vector graphics describes an image in terms of geometric objects such as lines and curves, *raster graphics* represent an image as an array of color values positioned in a grid pattern. In 2D each square cell in the grid is called a *pixel* (from picture element), and in 3D each cube cell is known as a *voxel* (volumetric pixel).

In 2D image processing, the operation of a *convolution*, or *filter*, is the multiplication of each pixel and its neighbours by a *convolution matrix*, or *kernel*, to produce a new image where each pixel is determined by the kernel, similar to the stencil operators in the finite difference method.

Common kernels include the *Sharpen* and *Gaussian blur* filters,

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \qquad \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \tag{4.17}$$

where we note the similarity to the finite difference stencil of the second order derivative (Laplacian) and its inverse.

Figure 4.5: Raster image (left), transformed by a Sharpen (middle) and a blur (right) filters.