universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

The ubiquitous role of Linear Algebra within Applied Mathematics

Verfasser

Bayram ÜLGEN

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, im August 2011

Studienkennzahl lt. Studienbuchblatt:	A 190 406 313
Matrikel-Nummer:	0602262
Studienrichtung lt. Studienbuchblatt:	UF Mathematik. UF Geschichte, Sozialkunde, Politische Bildung
Betreuer:	Univ.-Prof. Dr. Hans Georg Feichtinger

Contents

Preface

Let me start by saying a few words about the title of this thesis and how I made up my mind to write about linear algebra and applied mathematics.

In fact, linear algebra has become a powerful tool for plenty of mathematical applications. Today, we are confronted with linear algebra almost everywhere, e.g., in natural sciences, like physics, biology and chemistry, in various areas of engineering (electrical, mechanical, etc.) and economy (e.g., statistics). In particular, modern signal and image processing have linear algebra at their disposal. Data compression, digitizing images and "repairing" defective signals are representative examples where linear algebra comes into action. This is the explicit reason for choosing the title:"The ubiquitous role of Linear Algebra within Applied Mathematics".

For me personally, it was the course *Applied Mathematics* for SSTAP[1]-students lectured by Prof. Feichtinger, where applied numerical linear algebra and linear algebra exposed me first to the viewpoint. This original inspiration made me attend several consolidation seminars about applied (linear) algebra until I decided to write my thesis about linear algebra and applied mathematics in English. Nowadays, English has become an international language and the standard language of science. Thus, scientists prefer writing their papers, books, etc. in English in order to reach a broad audience. I share this intention. This is why the language of this thesis is English, too.

In this thesis, I present mathematical applications from physics and electrical engineering, statistics and demography, signal and image processing. I work with both more elementary literature, such as [Str03], [OS06], [Bro91] and [Usm87], and more technical-theoretical mathematical contributions, such as [Dem97], [Goc10] and [Sad08]. Hence, I combine both and introduce mathematical applications by means of linear algebra neither on a pure mathematical nor an elementary level but on a level, where both coexist and are not neglected. In addition, I worked with MATLAB and the TikZ

[1]Secondary School Teacher Accreditation Program

environment in order to generate both the figures and tables. The whole typesetting was done in LaTeX.

The first chapter combines physics and electrical engineering and presents three fundamental laws of Kirchhoff and Ohm, respectively, which combine into the *fundamental network equation*. Therefore, we consider the four subspaces, graphs and incidence matrices.

The second chapter is all about applications to eigenvalues and eigenvectors. We consider statistics and demographic processes by introducing *Markov chains*, image compression by introducing the *singular value decomposition*, and (minimal norm) least squares applications.

The third chapter deals with basics of signal and image processing. We provide an introduction to Fourier analysis and discuss some important tools for signal processing, e.g., *filters*.

The last chapter constitutes an attempt to review our work with respect to mathematics education, teaching and, in general, didactics.

At this point, I would like to express my sincere gratitude to my supervisor Prof. Hans Georg Feichtinger. First and foremost, I thank him for inspiring me to write this thesis with his courses and seminars. I learned from him that the interplay of pure mathematics and applied sciences is possible and (more than) important.

Next, I would like to thank Dr. Kayhan Ince for several discussions about the first chapter, in particular about Kirchhoff, Ohm and electrical circuits, and how linear algebra contributes to electrical engineering, in general. Further, I thank my friends and colleagues from software engineering, technical mathematics, pedagogy and educational studies for motivating me with new ideas, comments and questions.

Finally, I wish to thank my parents, who have supported me in many ways, for their patience and understanding.

1 The backbone of Linear Algebra: The Four Subspaces

1.1 Motivation

We start with an m by n matrix A with $m = 5$ rows and $n = 4$ columns.

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \tag{1.1}$$

It is a special matrix with entries $a_{ij} = -1, +1$ and 0 where i describes the i-th row and j the j-th column of the matrix A.

First of all, mathematicians want to learn about some properties of this matrix A before letting it operate on given inputs x, for example.

So the questions we ask ourselves are:

- What can we say about the **column space** of A, $C(A)$, and the **row space** $C(A^\top)$, respectively, where $A^\top$ denotes the transpose of a matrix A?
- Are the column vectors and the row vectors belonging to $C(A)$ and $C(A^\top)$ linearly dependent or independent?
- How is the interplay between $C(A)$ and $C(A^\top)$?
- Is the matrix A invertible? How is its inverse linked to the **nullspace** of A, $N(A)$? How about the **"left" nullspace** $N(A^\top)$?

- What about the $rank(A) = r(A)$, $dim\ C(A)$, $dim\ C(A^\top)$?
- ...

The series of questions can be continued. In principle they shall help revealing all the properties of the matrix A. Now we are ready for mathematical operations.

1.2 The Four Subspaces and Kirchoff's Laws

The reader might wonder where there is a connection between the subspaces and Kirchoff. Well, it exactly is the matrix A we introduced above. Just wait and see.

Let us start with Gaussian elimination. After elementary operations on the rows of the matrix A we get the echelon form U which is an upper triangular matrix.

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \longrightarrow U = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

What do we infer from the new matrix U? What has happened? Elimination made us possible to look behind the scene, i.e., if the matrix A was invertible it had to had linearly independent pivot columns or rows. Hence, there must (at least) be one solution to $Ax = \mathbf{0}$, x being other than the zero vector. Subsequently $N(A)$, the nullspace of A, is not "empty"[1]. The rank of A is equal to the number of pivot columns or rows; here $r = 3$. There are three fixed variables, the pivots, and one free variable which can be read from U. By seetting x_4 to 1 we get one solution to

$$Ux = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

with $x_1 = x_2 = x_3 = x_4 = 1$. All other solutions are multiples of x, i.e., $N(A)$ contains

[1] Actually, $N(A)$ is never empty as it consists of at least the zero vector.

all the vectors of the form $c \cdot (1,1,1,1)^\top$.

We have chosen the echelon form of A because U and A have the same nullspace. The reasoning behind it is clear [Str03, 48]: A goes to U by elimination means that A is multiplied by some *elementary* matrices E_{ij} and *permutation* matrices P_{km} from its left which yields U. The matrices E_{ij} subtract a multiple k of row j from row i. P_{km} cause row exchanges of row k with row m. Thus we have

$$E_{43}E_{32}P_{45}P_{34}E_{32}E_{21}A = \begin{pmatrix} 1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&-1&1&0\\0&0&0&0&1 \end{pmatrix}\begin{pmatrix} 1&0&0&0&0\\0&1&0&0&0\\0&0&-1&1&0\\0&0&0&1&0\\0&0&0&0&1 \end{pmatrix}\begin{pmatrix} 1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&0&1\\0&0&0&1&0 \end{pmatrix}$$
$$\begin{pmatrix} 1&0&0&0&0\\0&1&0&0&0\\0&0&0&1&0\\0&0&1&0&0\\0&0&0&0&1 \end{pmatrix}\begin{pmatrix} 1&0&0&0&0\\0&1&0&0&0\\0&-1&1&0&0\\0&0&0&1&0\\0&0&0&0&1 \end{pmatrix}\begin{pmatrix} 1&0&0&0&0\\-1&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1 \end{pmatrix}$$
$$\begin{pmatrix} -1&1&0&0\\-1&0&1&0\\0&-1&1&0\\0&-1&0&1\\0&0&-1&1 \end{pmatrix} = \begin{pmatrix} -1&1&0&0\\0&-1&1&0\\0&0&-1&1\\0&0&0&0\\0&0&0&0 \end{pmatrix} = U$$

The elimination and permutation matrices E_{ij} and P_{km} do not change anything; they make the matrix A more "transparent". Indeed, those multiplications are fairly linked to the LPU-factorization of invertible matrices where L is the inverse of E (cf. [Str03, 83,101f]).[2]

After this short trip, we would like to continue our analysis on the matrix A. The vector $x = (1,1,1,1)^\top$ found above is a basis for $N(A)$. As the dimension of a space is given by the number of basis vectors, $dim\ N(A) = 1$ holds. Furthermore we know that A has three linearly independent rows spanning the whole row space $C(A^\top) \in \mathbf{R}^n$; the $dim\ C(A^\top) = 3$. Is there any link between $N(A)$ and $C(A^\top)$? Initially, both are subspaces of $\mathbf{R}^n$. If we keep a beady eye on their vectors, we become aware of an important feature of any m by n matrix. MATLAB helps us achieving our goal faster:

```
>> i=1:3;
>> x=[1,1,1,1]';
>> U(i,:)
ans =
    -1     1     0     0
     0    -1     1     0
     0     0    -1     1
>> U(i,:)*x
```

[2] As U is **u**pper triangular, L is a **l**ower triangular matrix.

```
ans =
     0
     0
     0
```

We just multiplied each row i by $x = (1,1,1,1)^\top$ which yielded 0 every single time. We conclude that any $x \in N(A)$ is perpendicular or *orthogonal* to any vector $w \in C(A^\top)$. Hence the subspaces $N(A)$ and $C(A^\top)$ are orthogonal.

Now, let us have a look at the column space by considering Ax:

$$Ax = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_2 - x_1 \\ x_3 - x_1 \\ x_3 - x_2 \\ x_4 - x_2 \\ x_4 - x_3 \end{pmatrix}$$

$Ax = (x_2 - x_1, x_3 - x_1, x_3 - x_2, x_4 - x_2, x_4 - x_3)^\top =: b$ is true if and only if the right hand side b is in $C(A)$. The column space consists of all linear combinations of the columns of A. Hence $x = (c,c,c,c)^\top$ with constants c yields the correct solution to $Ax = b$. We already know that $r(A) = 3$; there are three linearly independent columns. Thus the dimension of $C(A)$ is 3.

Finally, we have to deal with the fourth subspace, the "left" nullspace or the nullspace of the row space or the abbreviation $N(A^\top)$ (cf. [Str03, 175]). It contains the solutions to

$$A^\top y = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Given the *dim* $C(A)$ we deduce that the *dim* $N(A^\top)$ has to be 2. As an analogue to $C(A^\top)$ and $N(A)$, $C(A)$ and $N(A^\top)$ are subspaces of $\mathbf{R}^m$. Thus their dimensions have to add to $m = 5$. We could have argued that the echelon form of $A^\top$ has to have three fixed and two free variables. The number of free variables is an indicator of the basis and dimension of $N(A^\top)$. The nullspace of $A^\top$ consists of all linear combinations

satisfying $A^\top y = (0)$, i.e.,

$$N(A^\top) = \left\{ c \underbrace{\begin{pmatrix} -1 \\ 1 \\ 0 \\ -1 \\ 1 \end{pmatrix}}_{y_\alpha} + d \underbrace{\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{y_\beta} \quad where\ c, d \in \mathbf{R} \right\}$$

Special about y_α and y_β is that they are orthogonal to $b \in C(A)$. Any multiples of y_α and y_β are perpendicular to b. Hence, the subspaces $N(A^\top)$ and $C(A)$ are orthogonal.

We have entirely revealed the properties of the matrix A. We know everything about the subspaces $C(A)$, $N(A^\top)$, $C(A^\top)$ and $N(A)$. And we are ready to sum up all the information we have gathered in one figure (cf. [Str03, 185]):

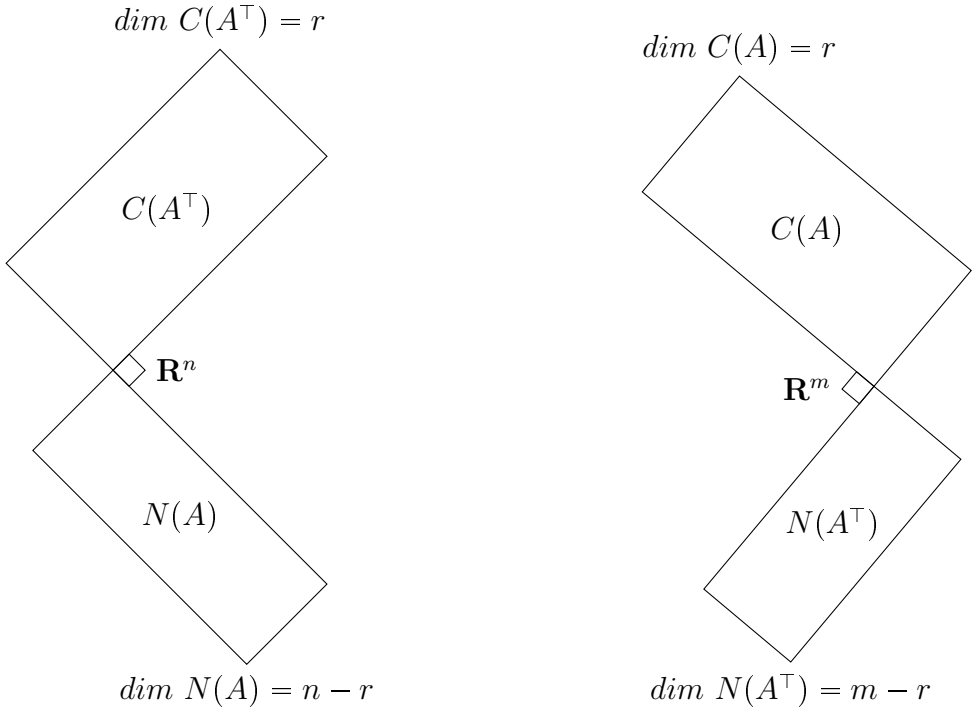
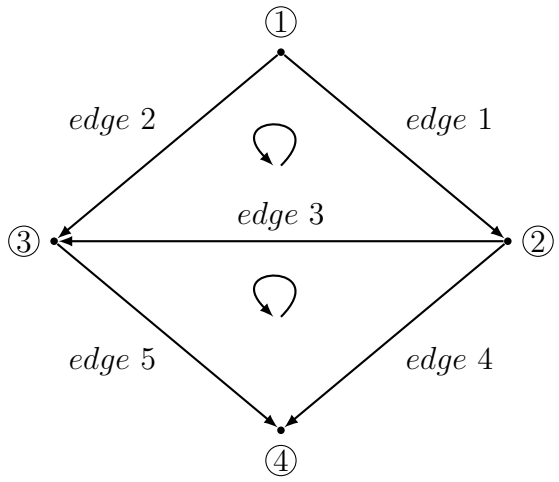


Figure 1.1: The four subspaces $C(A)$, $N(A^\top)$, $C(A^\top)$ and $N(A)$ at a glance.

Of course, we did not forget Kirchoff. We had to figure out all important features of (1.1) before we were able to link A to Kirchoff. The number of columns and rows of the matrix A correspond to the number of *nodes* and *edges* of a *directed graph*.

Definition 1. [Dem97, 288] A *directed graph* is a finite collection of *nodes* connected by a finite collection of directed *edges*, i.e., arrows from one node to another. A *path* in a directed graph is a sequence of nodes $n_1, \ldots, n_m$ with an edge from each n_i to n_{i+1}. A *self* edge is an edge from a node to itself.

Graphs are crucial in *discrete mathematics* (cf. [Goc10, p.168]). They serve as an abstract representation of interconnections of objects belonging to a certain set. For example, we could think of a set of websites and how some of them are *linked* to each other. Another example is a genealogical tree depicting a set of people and their relationships. Actually, this is a graph without any *loops* and is called - not by chance - a *tree*. What we want to do is to study such types abstractly. Therefore we assign every graph to an *incidence matrix* and vice versa. The figure below illustrates the link between A and the directed graph:



$$\begin{array}{c} \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{array} \begin{array}{c} \begin{array}{cccc} n_1 & n_2 & n_3 & n_4 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \end{array} = A$$

Figure 1.2: The *graph* associated with its *incidence matrix*.

From the very first, we knew that we would work with A in the course of this chapter; thus the matrix A was not chosen arbitrarily. Row 1 corresponds to *edge* 1 which starts at *node* 1 and ends at *node* 2. We want the entry in row 1 column 1 to be -1 as the edge leaves node 1. Hence, entering node 2 requires $+1$ in row 1 column 2.[3] Rows 2 to 5 are constructed analogous. The sign $\pm$ 1 is determined by the orientation of the arrows. Nevertheless, this does not imply that a flow has to go towards the direction of the edge; it can travel in the opposite direction, too.

When applying the incidence matrix A to the input vector x, as we did above, we get

[3] cf. [OS06, 123f] for different notation.

a vector of differences:

$$Ax = \begin{pmatrix} x_2 - x_1 \\ x_3 - x_1 \\ x_3 - x_2 \\ x_4 - x_2 \\ x_4 - x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix} = b$$

Most commonly x_1, x_2, x_3, x_4 represent **potentials** at the nodes and $Ax = b$ is the vector containing **potential differences** across the edges. These differences cause flows. x could also represent the heat and generally the temperature at the nodes, or the voltages at the nodes, or the heights of the nodes (cf. [Str07, 144]). Then b contains the heat and respectively the temperature differences, the voltage differences or the height differences.

We have already shown that the nullspace of A is a line $x = (1, 1, 1, 1)^\top \in \mathbf{R}^n$. As a conclusion, all potentials $(x_i)_{i=1,\dots,4}$ are equal. Furthermore, the rows give a record of the edges. As $r(A) = 3$, any 3 rows are linearly independent if their corresponding edges form a tree, e.g. edges 2, 3 and 4 or 1, 2 and 5, etc. We also want to mention the fact that any vector w is in the row space if and only if $w \perp x,\ x \in N(A)$.

As an analogue to the rows of $C(A^\top)$, any 3 columns of $C(A)$ are independent. But the sum of each four columns yields the zero vector. A unique solution to $Ax = b$ necessitates b to be perpendicular to y_α and y_β. At this point, at last, we introduce the first of two fundamental laws of *circuit theory*, namely **Kirchhoff's Voltage Law** *KVL*. Therefore we consider the three loops depicted in the graph; two small loops each in similar triangles and a big loop which will turn out to be a combination of the small loops. In general, KVL says that "the sum of voltage drops b_i around every closed loop is zero" [Str07, 147]. This is how a physicist or an electrical engineer would interpret the language of Linear Algebra. Adding potential differences around a closed loop yields zero. In particular, we have for example

$$(x_3 - x_2) + (x_4 - x_3) - (x_4 - x_2) \quad = \quad 0 \quad \text{(loop with edges } 3, 5, -4)$$

Provided that $b \in C(A)$, it holds $b_3 + b_5 - b_4 = 0$.

The left nullspace $N(A^\top)$ is two-dimensional containing y_α and y_β as a basis. They satisfy $A^\top y_\alpha = 0$ and $A^\top y_\beta = 0$. It is **Kirchhoff's Current Law** *KCL* which exactly

deals with $A^\top y = 0$ (cf. [SB97, 293f]), i.e.,

$$A^\top y = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} -y_1 - y_2 \\ y_1 - y_3 - y_4 \\ y_2 + y_3 - y_5 \\ y_4 + y_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The first equation tells the currents along the incident edges 1 and 2 to leave node 1. In other words, $-y_1 = y_2$ means that the flow into node 1 is the same as the flow out of it. As a result the total flow or net flow is zero. These equations are called *balance equations*, they are in *equilibrium* (cf. [Str07, 146] and [Str03, 417]). Thankfully, we did our job above by finding the solutions to $A^\top y = 0$. The currents y_α and y_β balance themselves. y_α represents the loop current around the rhombus; it goes backwards on edge 1, forward on edge 2, forward on edge 5 and finally backwards on edge 4. There are two more loops depicted in Figure 1.2. One small loop is given by y_β. Due to $dim\ N(A^\top) = 2$, the third loop is only a combination of y_α and y_β, i.e.,

$$y_\alpha - y_\beta = \underbrace{\begin{pmatrix} -1 \\ 1 \\ 0 \\ -1 \\ 1 \end{pmatrix}}_{\text{big loop}} - \underbrace{\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{\text{small loop}} = \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{pmatrix}}_{\text{small loop}} = y_\gamma \tag{1.2}$$

Remark. The graph in Figure 1.2 lies in a plane. We have *Euler's Formula*: $\#\ nodes - \#\ edges + \#\ loops = 1$. Referring to the incidence matrix and the graph, there were $n = 4$ nodes, $m = 5$ edges and $m-(n-1) = 2$ (independent) loops. Thus, $4-5+2 = 1$. In addition, $m - (n - 1)$ is the dimension of the nullspace of $A^\top$. Hence, the number of independent loops coincides with the dimension of $N(A^\top)$. ◇

Having introduced *KVL* and *KCL*, we are interested in their interplay when applied to networks. In networks, **Ohm's Law** *OL* holds for each particular current y_m. Actually, there is a direct proportionality between voltages and resistances or currents, respectively, i.e., $b = Ry$. R is called the *resistance matrix*. It is at least symmetric and positive definite (cf. [OS06, 303]). We prefer $R^{-1} = C$, where C is the *conductance matrix* with "material constants" $c_m = \frac{1}{R_m}$ which are assigned to each edge, and which

measure the physical properties of the edges (cf. [Str86, 88]). Rewriting the equation above yields $y = -Cb$. We had to change the sign of Ax. In circuit theory a flow goes from higher potential to lower potential. The potential difference across edge 1 is given by $x_1 - x_2 > 0$ providing positive current (cf. [Str03, 419]). With OL, we have figured out three fundamental laws of electricity. Combining $-Ax = b$, $y = Cb$ and $A^\top y = 0$, we get $A^\top CAx = 0$. We have put those equations together but, furthermore, we want some external powers such as current sources f and voltage sources p, e.g., batteries, to put something into action. Thus, we change $-Ax = b$ to $p - Ax = b$ and KCL from $A^\top y = 0$ to $A^\top y = f$. There are batteries p_m along the edges and current sources f_n at the nodes. The modified formulas combine into following *fundamental equation of equilibrium* (cf. [Str07, 152]):

$$\begin{array}{rcrcl} C^{-1}y & + & Ax & = & p \\ A^\top y & & & = & f \end{array} \quad \Longleftrightarrow \quad \begin{pmatrix} C^{-1} & A \\ A^\top & 0 \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} p \\ f \end{pmatrix} \tag{1.3}$$

which becomes

$$A^\top C(p - Ax) = f \quad \Longleftrightarrow \quad A^\top CAx = A^\top Cp - f. \tag{1.4}$$

$K = A^\top CA$ is called the *resistivity matrix*. Its entries depend on the conductances of the edges in the network (cf. [OS06, 304]). $A^\top CA$ is central to mathematical applications. Gilbert Strang's approach to applied Linear Algebra and applied mathematics is generally built on K (cf. [Str03, 419] and [Str86]). After much input, we would like to give an example with both a current source and a battery.

Example. (cf. [OS06, 308], [Str03, 8.2.(13)] and [Str86, 110–113]) Consider an electrical network along the edges e_m of the graph in Figure 1.3 (for convenience, we have chosen the graph introduced above). Edges e_1 and e_2 contain conductances c_1 and c_2 being 2 *Siemens* and edges e_3 to e_5 have conductances c_3, c_4 and c_5 with 3 *Siemens*. Further, there is a 9-volt battery connected along e_3 and a 2-ampere current source applied at node 1 (coming out at node 2). We want to figure out how much current flows through each edge.

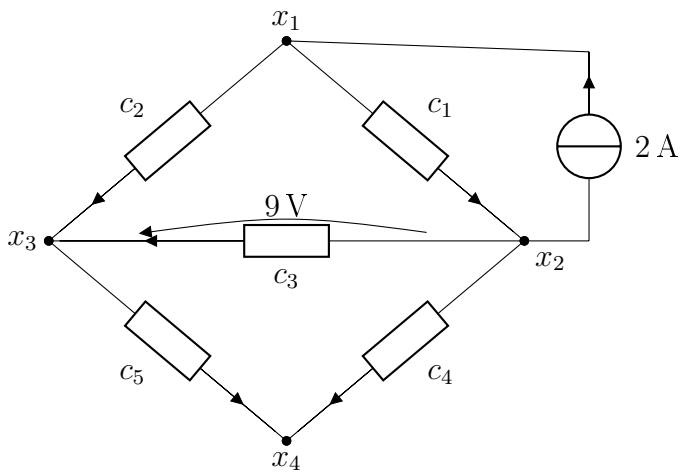


Figure 1.3: Electrical network with voltage source and current source.

According to the fundamental network equation, we start with computing the resistivity matrix K:

$$K = A^\top CA = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$
$$\begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & -2 & -2 & 0 \\ -2 & 8 & -3 & -3 \\ -2 & -3 & 8 & -3 \\ 0 & -3 & -3 & 6 \end{pmatrix}$$

We know that the columns of A and the rows of $A^\top$ are linearly dependent. Their nullspace are not empty. Hence, we cannot expect K to be non-singular and invertible. In order to solve for all potentials one node needs to be *grounded*. This is a technical term used by (electrical) engineers in terrestrial electricity, where "the earth is assumed to have zero potential. Specifying a particular node to have zero potential is physically equivalent to grounding that node." Applied to our example, we decide to ground node 4, i.e., we set $x_4 = 0$. As a consequence, we may drop the fourth column of A and thus the fourth row of $A^\top$. Now K has become a positive definite, invertible 3 by 3 matrix.

We are ready to solve the right hand side of the equation $A^\top CAx = A^\top Cp - f$:

$$A^\top Cp - f = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 9 \\ 0 \\ 0 \end{pmatrix}$$
$$- \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -27 \\ 27 \end{pmatrix} - \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -29 \\ 27 \end{pmatrix}$$

The rest is done via MATLAB:

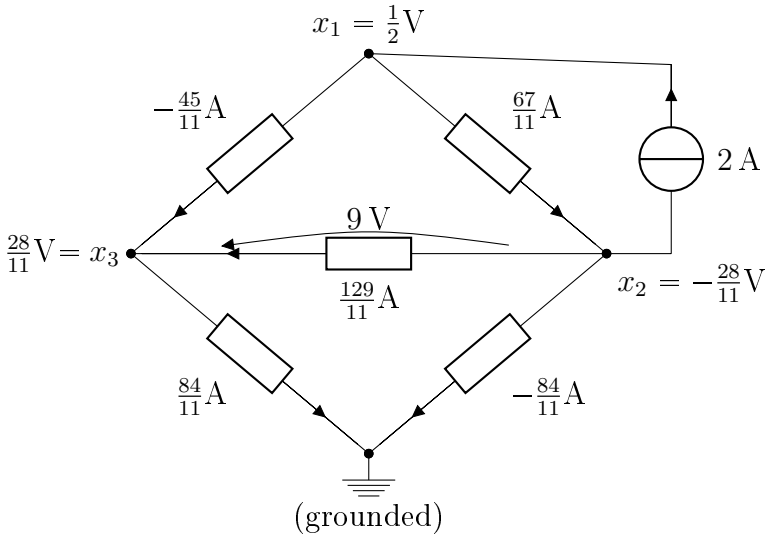
```
>> rhs=A'*C*p-f        %rhs abbr. for right hand side;
>> x=inv(K)*rhs
x =
        1/2
     -28/11
      28/11
```

Having figured out the potentials or node voltages x_1, x_2 and x_3, OL $y = Cb = C(p - Ax)$ yields the currents:

```
>> y=C*(p-A*x)
y =
      67/11
     -45/11
     129/11
     -84/11
      84/11
```

We conclude that the largest current flows through the battery edge y_3, while y_2 transmits the least. The minus sign only indicates y_2 flowing into the opposite direction.

Following network records our results for potentials x_n and currents y_m:



⬠

Summary. (cf. [Str03, 417]) Basic features of Linear Algebra made us possible to look into the world of physics and engineering. Here is a record of central ideas of the first chapter:

1. The key to applied Linear Algebra: (understanding) the four subspaces $C(A)$, $C(A^\top)$, $N(A)$ and $N(A^\top)$.

2. Each graph has its corresponding incidence matrix and vice versa.

3. $dim\ C(A) = dim\ C(A^\top) = n-1$, $dim\ N(A) = 1$ with $N(A)$ containing constant vectors $(c, c, \ldots, c)^\top$ (cf. [OS06, Prop. 2.51]). and $dim\ N(A^\top) = m-(n-1)$.

4. **Kirchhoff's Voltage Law**: Potential differences Ax add to zero around each closed loop.

5. **Ohm's Law**: $y = Cb$, i.e., *Current = Conductance* $\times$ *voltage drop.*

6. **Kirchhoff's Current Law**: Consider $A^\top y = 0$. Flows balance themselves at each node, i.e., flow in equals flow out.

7. Combining KVL, OL and KCL generates the *fundamental network equation*: $A^\top CAx = A^\top Cp - f$ with voltage sources p and current sources f.

2 Eigenvalues

This section is devoted to sample applications of eigenvalues and eigenvectors. We present Markov chains and stochastic processes, elementary image compression with the SVD, the pseudo-inverse and applications to (minimal norm) least squares.

2.1 Markov chains

Definition 2. (cf. [Sen81, 113], [FW79, 1] and [Beh00, Chapt. 1]) Given a finite set $S = s_0, s_1, \ldots, s_{n-1}$, the *state space* with n *states*, with the conditional probability $p(s_{k+1}|s_k, s_{k-1}, \ldots, s_1, s_0) = p(s_{k+1}|s_k)$. This probability property is called *Markov property*. Further, let ρ denote a probability vector with entries $(p_i)_{i \in S}$ where $p_i \geqslant 0 \ \forall j$ and $\sum p_i = 1$, and let $T = (p_{ij})_{i,j \in S}$ be a *transition matrix* where $p_{ij} \geqslant 0$ and $\sum_i p_{ij} = 1 \ \forall j$. The tuple (S, T) endowed with the Markov property constitutes a *finite Markov chain*.

Informally (cf. [Bro91, 5,181]), we consider a system with a countable number of states and assume that the state of the system tomorrow only depends on the current state of the system (and not on the past, i.e., it has no memory) (cf. [Sad08, 126]). The system moves from state i to one of the n states, say j, with the probability p_{ij}. In addition, each column of the *stochastic matrix* T is represented by a probability vector p_i.

Actually, we do not want to know what happens with the system after a single step or two steps. More generally, we are interested in the case after k steps, i.e., what is the conditional probability $p(s_k|s_0)$? Before introducing an example, we would like to remind the reader of some crucial facts about *eigenvectors* and *eigenvalues* which will facilitate a lot.

Initially, if we know the eigenvalues and eigenvectors of a square matrix A, we can easily deal with its powers, A^k, by *diagonalizing* A. Although not every n by n matrix is diagonalizable, we can find some matrix M such that $M^{-1}AM = J$. Then A is "as nearly diagonal as possible" (cf. [Str03, 346]). J denotes the *Jordan canonical form*. Nevertheless, we will work with diagonalizable matrices throughout this section.

Definition 3. (cf. [OS06, 410] and [Str03, 288]) A square matrix A is called diagonalizable if there exists a matrix S, where each of its columns is represented by exactly one of n linearly independent eigenvectors of A, such that

$$S^{-1}AS = \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \quad \text{or, equivalently we have} \quad A = S\Lambda S^{-1}.$$

Λ is a diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$, the eigenvalues of A.

Proof. See [OS06, 410f] or [Str03, 288f]. □

The definition above holds for square matrices. We want the matrix A to be symmetric, too, i.e., we consider $A = A^\top$.

Theorem 1. (cf. [OS06, 413] and [Str03, 318f]) Let A be real and symmetric. Then following holds:

1. All the eigenvalues are real.
2. Eigenvectors corresponding to different eigenvalues are orthogonal.

Proof. See [OS06, 415f] or [Str03, 320f]. □

Furthermore, $A = S\Lambda S^{-1}$ becomes $A = Q\Lambda Q^{-1}$ for symmetric matrices:

Theorem 2. Given a real, symmetric matrix A. Then there exists an orthogonal matrix Q and a real diagonal matrix Λ such that

$$A = Q\Lambda Q^{-1} \quad \text{with} \quad Q^{-1} = Q^\top.$$

Proof. See [Str03, 319–322]. □

Remark. (cf. [OS06, 418f] and [Str03, 319]) We are talking about the *spectral factorization* of A and, more generally, the *Spectral Theorem* in mathematics and the *Principal Axis Theorem* in physics. The columns of the orthogonal matrix Q correspond to eigenvectors of A which we can choose orthonormal. Hence, the columns of Q form an orthonormal basis of $\mathbf{R}^n$. ◇

Being aware of how diagonalization of a matrix is possible, we approach A^k. Let us start with $k = 2$. Therefore, we have

$$A^2 = S\Lambda S^{-1} S\Lambda S^{-1} = S\Lambda^2 S^{-1}$$

and by iteration

$$A^k = \underbrace{S\Lambda S^{-1} \cdots S\Lambda S^{-1}}_{k-times} = S\Lambda^k S^{-1}.$$

Obviously, the eigenvectors of A do not change and the eigenvalues are squared and taken to the k-th power, respectively. Analogous, if we have a real, symmetric matrix A, the above equations hold true for $S = Q$.

Now, we are given the conditions which make us possible to deal with a *first order linear iterative system*, i.e., $u_{k+1} = Tu_k$, representing a *Markov chain*.

Example. [OS06, Problem 10.4.2] A study has determined that, on average, the occupation of a boy depends on that of his father. If the father is a farmer, there is a 30% chance that the son will be a blue color laborer, a 30% chance he will be a white color professional, and a 40% chance he will also be a farmer. If the father is a laborer, there is a 30% chance that the son will also be one, a 60% chance he will be a professional, and a 10% chance he will be a farmer. If the father is a professional, there is a 70% chance that the son will also be one, a 25% chance he will be a laborer, and a 5% chance he will be a farmer.

1. What is the probability that the grandson of a farmer will also be a farmer?
2. In the long run, what proportion of the male population will be farmers?

The information from above is recorded in following figure:

SON \ *FATHER*	**farmer**	**laborer**	**professional**
farmer	0.4	0.1	0.05
laborer	0.3	0.3	0.25
professional	0.3	0.6	0.7

Table 2.1: Father-son professional relationship.

Of course, we could solve this problem via conditional probabilites and probability trees. The solution to the first part would easily be $p(f_2|f) = 0.4^2+0.1\cdot0.3+0.05\cdot0.3 = 0.205$.[1] This was not challenging as we only set $k = 2$. We do not want to utilize stochastics in order to calculate further steps.

Given the Markov property, we can read the transition matrix or *Markov matrix* from the figure above. Actually, it is a copy of it, i.e.,

$$T = \begin{pmatrix} 0.4 & 0.1 & 0.05 \\ 0.3 & 0.3 & 0.25 \\ 0.3 & 0.6 & 0.7 \end{pmatrix}.$$

The entries of T satisfy $0 \leqslant p_{ij} \leqslant 1$ and $\sum_i p_{ij} = 1\ \forall j$. Hence, each column of T represents a probability vector and T is a stochastic matrix. First of all, we are interested in the eigenvalues and eigenvectors of T. Therefore, we consider $det(T - \lambda I) = 0$, the characteristic equation:

$$det(T-\lambda I) = \begin{vmatrix} 0.4-\lambda & 0.1 & 0.05 \\ 0.3 & 0.3-\lambda & 0.25 \\ 0.3 & 0.6 & 0.7-\lambda \end{vmatrix} = \lambda^3 - \frac{7}{5}\lambda^2 + \frac{83}{200}\lambda - \frac{3}{200} = 0$$

yields $\lambda_1 = 1$ and $\lambda_{2,3} = \frac{4\pm\sqrt{10}}{20}$ with corresponding eigenvectors x_1, x_2 and x_3 which we have converted into probability vectors just by dividing by the sum of the elements of each vector:

$$\lambda_1 :\ x_1 = \begin{pmatrix} 0.0976 \\ 0.2683 \\ 0.6341 \end{pmatrix},\ \lambda_2 :\ x_2 = \begin{pmatrix} -5.84\cdot10^{15} \\ -.95\cdot10^{15} \\ 6.79\cdot10^{15} \end{pmatrix},\ \lambda_3 :\ x_3 = \begin{pmatrix} -.56\cdot10^{15} \\ 3.43\cdot10^{15} \\ -.29\cdot10^{15} \end{pmatrix}$$

satisfying $Tx_i = \lambda_i x_i, i = 1, 2, 3$. The elements of x_2 and x_3 are very big; we had to

[1] f_2 denotes the 2^{nd} generation farmer.

shorten them. Actually, we knew that one root of the characteristic equation would be 1. T is a regular stochastic matrix with entries $p_{ij} \neq 0$ and following theorem tells us more about it:

Theorem 3. (cf. [Sad08, 128] and [OS06, 542]) If T is a regular stochastic matrix, then

1. $\lambda_1 = 1$ is an eigenvalue of T. The entries of the corresponding eigenvector x_1 are all non-negative. We can normalize x_1 such that it becomes a probability vector.

2. T has no eigenvalue of magnitude greater than 1.

3. Any Markov chain with coefficient matrix T converges to the probability eigenvector: $u_k \to x_1$ $(k \to \infty)$.

Proof. See [Sad08, 128–130] and [OS06, 543]. □

Provided that the father is a farmer, the initial state vector is given by $u_0 = (1, 0, 0)^\top$. Additionally, we can diagonalize T and thus have a look at T^k. The initial vector is just a combination of the eigenvectors x_1, x_2 and x_3 with coefficients $c_1 = 1$, c_2 and c_3:

$$u_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.0976 \\ 0.2683 \\ 0.6341 \end{pmatrix} + c_2 \begin{pmatrix} -5.84 \cdot 10^{15} \\ -.95 \cdot 10^{15} \\ 6.79 \cdot 10^{15} \end{pmatrix} + c_3 \begin{pmatrix} -.56 \cdot 10^{15} \\ 3.43 \cdot 10^{15} \\ -.29 \cdot 10^{15} \end{pmatrix}$$

c_2 and c_3 are very small negative numbers. The columns of S contain the eigenvectors of T. MATLAB finds it quite exhausting to calculate $u_0 = Sc \Leftrightarrow c = S^{-1}u_0$, where $c = (c_1, c_2, c_3)^\top$, due to the large condition number of T which indicates that $det(T)$ is approximately zero and T is nearly singular.

Formally, multiplying $c_1x_1 + \cdots + c_nx_n$ by the k^{th} powers of the eigenvalues yields

$$\begin{aligned} u_k = c_1\lambda_1^k x_1 + \cdots + c_n\lambda_n^k x_n &= \begin{pmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \\ &= S\Lambda^k c. \end{aligned}$$

Obviously, we have $S\Lambda^k c = S\Lambda^k S^{-1}u_0 = T^k u_0$. Hence, the solution to the difference

system $u_{k+1} = Tu_k$ is $u_k = T^k u_0$.

We return to our problem and consider the state vectors for, e.g., $k = 2, 3, 5$ and 10:

$$u_2 = \begin{pmatrix} 0.2050 \\ 0.2850 \\ 0.5100 \end{pmatrix},\ u_3 = \begin{pmatrix} 0.1360 \\ 0.2745 \\ 0.5895 \end{pmatrix},\ u_5 = \begin{pmatrix} 0.1025 \\ 0.2691 \\ 0.6284 \end{pmatrix},\ u_{10} = \begin{pmatrix} 0.0976 \\ 0.2683 \\ 0.6341 \end{pmatrix}$$

The first entry of u_2 solves the first part of the problem. It says that, on average, a fifth of all grandsons will become a farmer. With each step we converge to a "steady-state limiting probability distribution". Actually, u_{10} is already practically equal to x_1. Since $|\lambda_{2,3}| < 1$, u_k approaches the limit quite quickly as $k \to \infty$, at least as fast as $(1 - 3 \cdot \lambda_2)^k$ (cf. [Sad08, 130]) , i.e.,

$$u_k = \underbrace{c_1 1^k x_1}_{\substack{\to x_1 \\ (k\to\infty)}} + \underbrace{c_2 \underbrace{\left(\frac{4+\sqrt{10}}{20}\right)}_{<1}^k x_2 + c_3 \underbrace{\left(\frac{4-\sqrt{10}}{20}\right)}_{<1}^k x_3}_{\substack{=0 \\ (k\to\infty)}} = x_1\ (k \to \infty) \left[= \begin{pmatrix} 0.0976 \\ 0.2683 \\ 0.6341 \end{pmatrix} \right].$$

Furthermore, we infer that the steady state x_1 does not depend on initial conditions. We could have chosen any initial state vector u_0 but, finally, the steady state would always turn out to be x_1. In order to answer the second part of the problem we just read x_1; accordingly, 10% of the whole population will become farmers, 27% will become laborers and 63% will become professionals, in the long run. ⬠

There are plenty of other applications where one can encounter Markov chains. *Matrix Models of Base Substitution* (cf. [AR04, chapt. 4.4]) in biology dealing with problems like "What is the probability that a base A [A is a nucleotide and one of the four molecular subunits of the DNA; it is an abbreviation for *adenine*, author's note] in the ancestral sequence will have mutated to become a base T [*thymine*, author's note] in the descendent sequence 100 time steps later?" ([AR04, 145]), or *population models* in demographic studies (cf. [BCC01])or *supply and demand models* in economy (cf. [Str03, 426–428]) are just three representative examples. For further details about stochastic processes and Markov chains, we refer the reader to [Beh00], [AR04] and [BCC01].

2.2 Image compression

We consider an m by n image as an m by n matrix. When digitizing an image we want to read its details by dividing it into pieces, i.e., a *lattice* with small rectangulars (=*pixels*). Each rectangular is assigned a number indicating the brightness of a pixel. Hence, the more pixels exist, i.e., the "smoothier" the lattice is, the more numbers we need. These numbers are stored in a matrix. A 320 by 200 lattice equivalently necessitates a matrix with the same size. Thus it stores $320 \cdot 200 = 64000$ numbers which must not be underestimated. Data processing by a computer is not appropriate if there is a vast amount of inputs to deal with. The hard disk does not only assign big data to a bigger storage, it is not sparse with regard to processing time at all. So, we want to *compress* the original image by replacing its corresponding matrix such that the initial image can be reconstructed from many few numbers without much loss of quality of the image. Therefore, we consider the *Singular Value Decomposition* (SVD) (cf. [Gra04, 106], [Dem97, 114] and [Str03, 352]):

Theorem 4. ([Dem97, 109]) Let A be an arbitrary m-by-n matrix with $m \geqslant n$. Then we can write $A = U\Sigma V^\top$, where U is m-by-n and satisfies $U^\top U = I$, V is n-by-n and satisfies $V^\top V = I$, and $\Sigma = diag(\sigma_1, \ldots, \sigma_n)$, where $\sigma_1 \geqslant \cdots \geqslant \sigma_n \geqslant 0$. The columns $u_1, \ldots, u_n$ of U are called *left singular vectors*. The columns $v_1, \ldots, v_n$ of V are called *right singular vectors*. The σ_i are called *singular values*. (If $m < n$, the SVD is defined by considering $A^\top$.)

Proof. See [Dem97, 110] or [Gra04, 109]. □

Gilbert Strang finds the SVD to be the "climax of Linear Algebra" (cf. [Str03, 357]) and the third and final part of the *fundamental theorems of Linear Algebra*[2]. It has some crucial algebraic and geometric properties which are enlisted below:

Theorem 5. ([Dem97, 111]) Let $A = U\Sigma V^\top$ be the SVD of the m-by-n matrix A, where $m \geqslant n$.

1. The eigenvalues of the symmetric matrix $A^\top A$ are σ_i^2. The right singular vectors v_i are corresponding orthonormal eigenvectors.

[2]The first part is about the dimension of the four subspaces $C(A)$, $N(A^\top)$ and $C(A^\top)$, $N(A)$ (cf. [Str03, 177]). The 2^{nd} fundamental theorem tells us how they are linked to each other (*orthogonality* of the four subspaces) (cf. [Str03, 187]).

2. The eigenvalues of the symmetric matrix $AA^\top$ are σ_i^2. The left singular vectors u_i are corrseponding orthonormal eigenvectors for the eigenvalues σ_i^2.

3. Suppose $\sigma_1 \geqslant \cdots \geqslant \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$. Then the rank of A is r. The null space of A is the space spanned by columns $r+1$ through n of V, i.e., $span(v_{r+1}, \ldots, v_n)$. The column space of A is the space spanned by columns 1 through r of U, i.e., $span(u_1, \ldots, u_r)$.

4. Let S^{n-1} be the unit sphere in $\mathbf{R}^n$: $S^{n-1} = \{x \in \mathbf{R}^n : \|x\|_2 = 1\}$. Let $A \cdot S^{n-1} = \{Ax, x \in \mathbf{R}^n \text{and} \|x\|_2 = 1\}$. Then $A \cdot S^{n-1}$ is an ellipsoid centered at the origin of $\mathbf{R}^m$, with principal axes $\sigma_i u_i$.

5. Write $V = \{v_1, \ldots, v_n\}$ and $U = \{u_1, \ldots, u_n\}$, so $A = U\Sigma V^\top = \sum_{i=1}^n \sigma_i u_i v_i^\top$ (a sum of rank-1 matrices). Then a matrix of rank $k < n$ closest to A (measured with $\|\cdot\|_2$) is $A_k = \sum_{i=1}^k \sigma_i u_i v_i^\top$, and $\|A - A_k\|_2 = \sigma_{k+1}$. We may also write $A_k = U\Sigma_k V^\top$, where $\Sigma_k = diag(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$.

Proof. (cf. [Dem97, 111f])

1. $A^\top A = (U\Sigma V^\top)^\top U\Sigma V^\top = V\Sigma U^\top U\Sigma V^\top = V\Sigma^2 V^\top$. We have diagonalized $A^\top A$ where V is an orthogonal matrix with eigenvectors of $A^\top A$ and Σ^2 contains the eigenvalues σ_i^2 on its diagonal.

2. Analogous to 1.

3. We split U up into $\hat{U} = [U_1, U_2]$ where U_1 is m-by-n and U_2 is m-by-$(m-n)$. Since $\hat{U}$ and V are regular, A and $\hat{U}^\top AV = [\Sigma_1, \mathbf{0}]^\top = \hat{\Sigma}$, where Σ_1 is n-by-n and $\mathbf{0}$ is $(m-n)$-by-n, have the same rank. Further, we infer from $A = U\Sigma V^\top \Leftrightarrow AV = U\Sigma$ that Av_i is a multiple of the left singular vectors u_i, i.e., $Av_i = \sigma_i u_i$. By assumption, $\sigma_{r+1} = \cdots = \sigma_n = 0$ and thus $Av_1 + \cdots + Av_r + Av_{r+1} + \cdots + Av_n = \sigma_1 u_1 + \cdots + \sigma_r u_r$. Hence, $Av_i = 0$ for $i = r+1, \ldots, n$ and $Av_i = \sigma_i u_i$ for $i = 1, \ldots, r$. So the nullspace of A is spanned by $v_{r+1}, \ldots, v_n$ and the column space of A is spanned by $u_1, \ldots, u_r$.

4. We start by multiplying S^{n-1} by $V^\top$. Since V is orthogonal, it maps unit vectors to other unit vectors and thus does not change the shape of S^{n-1}. Hence, $V^\top S^{n-1} = S^{n-1}$. Next, since $v \in S^{n-1}$ if and only if $\|v\|_2 = 1$, $x \in \Sigma S^{n-1}$ if and only if $\|\Sigma^{-1}x\|_2 = 1$ which is the same as $\sum_{i=1}^n (\frac{x_i}{\sigma_i})^2 = 1$. This is the well known equation for an ellipsoid with principal axes $\sigma_i e_i$, where e_i is the i^{th} column of I.

Finally, we multiply $x = \Sigma v$ by U such that each e_i becomes u_i, the i^{th} column of U.

5. By construction, the rank of A_k is k. Further,

$$\|A - A_k\|_2 = \| \sum_{i=k+1}^{n} \sigma_i u_i v_i^\top \|_2 = \sigma_{\max} = \sigma_{k+1}.$$

Next, we have to show that A_k is the matrix closest to A with respect to the rank. Therefore, we consider an arbitrary matrix B with $rank(B) = k$. Hence, the dimension of its nullspace is $n - k$. The dimension of the space spanned by $\{v_1, \ldots, v_{k+1}\}$ is $k + 1$. Since $(n - k) + (k + 1) > n$, we choose a unit vector h which is in their intersection. We have

$$\begin{aligned} \|A - B\|_2^2 &\geqslant \|(A - B)h\|_2^2 = \|Ah\|_2^2 \\ &= \|U\Sigma V^\top h\|_2^2 = \|\Sigma(V^\top h)\|_2^2 \\ &\geqslant \sigma_{k+1}^2 \|V^\top h\|_2^2 = \sigma_{k+1}^2. \end{aligned}$$

□

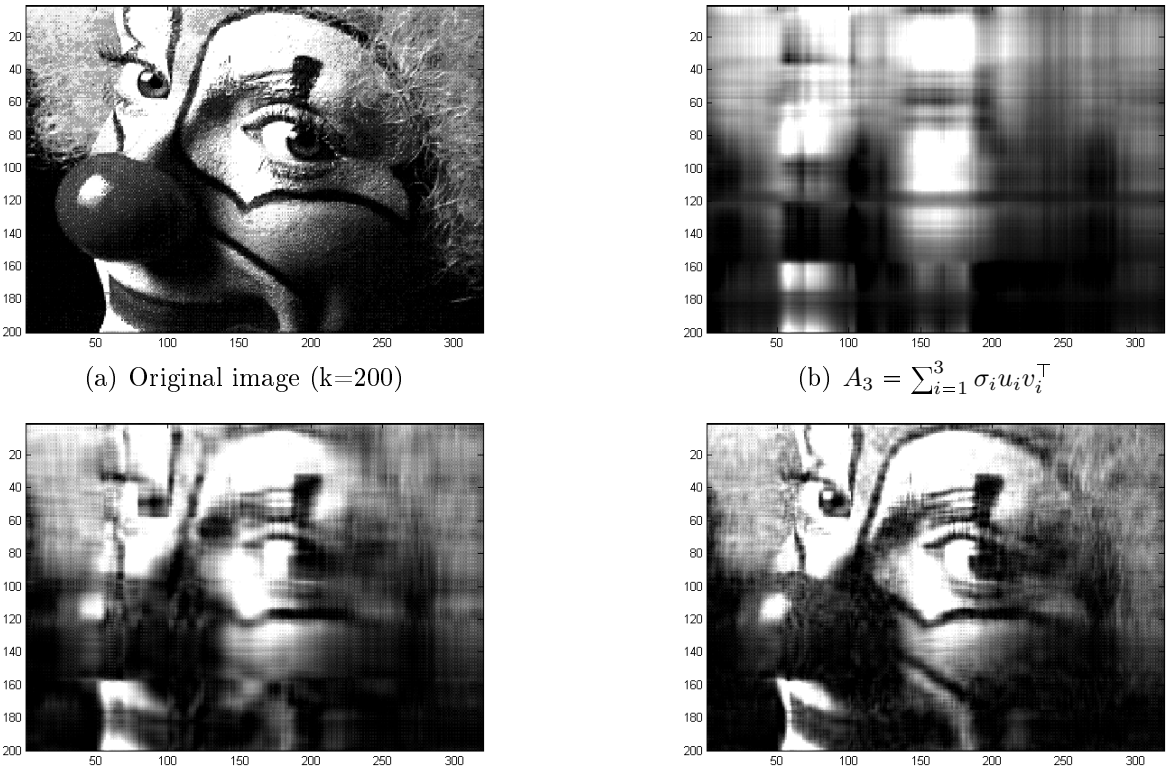
Remark. (cf. [Gra04, 103] and [OS06, 428]) Let A be symmetric. According to the theorem about orthogonal factorization, A decomposes to $A = Q\Lambda Q^\top$ with $Q = U = V$. We can write $A = \lambda_i q_i q_i^\top$, equivalently. This is a linear combination of *rank-1* matrices or projection matrices $q_i q_i^\top$, respectively. Thus, we interpret the SVD as a "generalization of the spectral factorization to non-symmetric matrices" ([OS06, 427]).

Next, we conclude from (3) that the orthogonal matrices U and V contain orthonormal bases for the four subspaces. In addition to (3), $\{v_1, \ldots, v_r\}$ is a basis for the row space and $\{u_{r+1}, \ldots, u_n\}$ is a basis for the left nullspace.

(4) gives a geometric interpretation of the SVD. Provided that A is a non-singular n-by-n matrix, U, V and Σ are square, too. U and V represent rotations and reflections, whereas Σ represents an orthogonal stretching transformation.

Finally, (5) tells us how to reconstruct A only by considering k singular values. In short, this is the *reduced singular value decomposition* (rSVD) giving an optimal low-rank approximation to A (cf. [Str03, 352]). In applications, we can ignore very small singular values which do not have any importance with regard to the reconstruction of A. ◇

Example. (cf. [Dem97, 114ff] and [Gra04, 107f])Consider the 320×200 image of a clown in Figure 2.1(a) below. It represents the original image with $rank(A) = 200$, i.e., the complete amount of singular values σ_i with $i = 1, \ldots, n$. We apply (5) of Theorem 5 in order to compress the image. Images (b) to (d) represent an approximation to A for $k = 3, 10$ and 20.



(a) Original image (k=200)

(b) $A_3 = \sum_{i=1}^{3} \sigma_i u_i v_i^\top$

(c) $A_{10} = \sum_{i=1}^{10} \sigma_i u_i v_i^\top$

(d) $A_{20} = \sum_{i=1}^{20} \sigma_i u_i v_i^\top$

Figure 2.1: Clowns compressed via SVD. (a) Original image, (b) Approximation to A for $k = 3$, (c) Approximation to A for $k = 10$ and (d) Approximation to A for $k = 20$

Note that we do not store $m \cdot n$ entries anymore. Since we take the first k vectors of U and V and the first k elements of Σ into consideration, this makes a total amount of $(m \cdot k) + (n \cdot k) + k = (m + n + 1) \cdot k$. We have recorded the compression ratios $(m + n + 1) \cdot k/(m \cdot n)$ and relative errors σ_{k+1}/σ_1 in following table:

k	$521k/64000$	σ_{k+1}/σ_1
3	.024	.155
10	.081	.077
20	.163	.040

Table 2.2: Compression ratio and relative error.

The images were produced in MATLAB by following commands:

```
load clown
[U,S,V]=svd(X) % MATLAB assigns the matrix X to the image.
colormap('gray')
image(U(:,1:k)*S(1:k,1:k)*V(:,1:k)') % set k=3,10,20
```

One can use the command `svds(X,k)` which just computes the first k entries of Σ where σ_i are in decreasing order. The command lines become

```
load clown
[U,S,V]=svds(X,k)
colormap('gray')
image(U*S*V').
```

⬠

Remark. The SVD has remarkable properties but it is not an ideal choice in applications like image processing and image compression respectively, statistics and finance. Computing the orthogonal matrices U and V are expensive. We need sparse matrices where computation is faster and cheaper. Therefore, we will introduce *filters* when dealing with elements of *Fourier-Analysis* in the course of this paper. For further details about *Limitations of the SVD* (cf. [Str07, 96f]). ◇

2.3 Minimal Norm Least Squares *MNLSQ*

A linear system of equations, i.e., $Ax = b$, has either no solution, a unique solution or infinitely many solutions. We try to read $C(A)$ properly and often want to know if $b \in C(A)$. In applications, we face *overdetermined* systems with more rows than columns where the right-hand side b does not constitute a linear combination of the columns of A. Thus, we look for the closest vector in the subspace, say p, with respect to b. Following figure illustrates the idea of what we are aiming at:

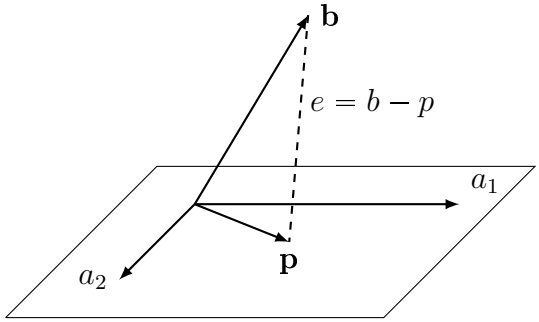


Figure 2.2: Projection of b onto the column space.

Actually, p is the *orthogonal projection* of b onto the column space of A spanned by a_1 and a_2, in particular. The *residual* $e = b - p$ is perpendicular to p and thus to the whole column space as $p = \sum_{i=1}^{n} x_i a_i$. In general, it holds

$$\langle e \,|\, a_i \rangle = 0 \Leftrightarrow A^\top (b - A\hat{x}) = 0 \Leftrightarrow A^\top A\hat{x} = A^\top b \tag{2.1}$$

for $i = 1, \ldots, n$. These are the so-called *normal equations*. Further, $\hat{x} = (A^\top A)^{-1} A^\top b$ is the unique solution to $Ax = b$ if and only if there exists $(A^\top A)^{-1}$ and if $\{a_1, \ldots, a_n\}$ is a basis for $C(A)$. Hence, $p = A\hat{x} = A(A^\top A)^{-1} A^\top b$ where $P = A(A^\top A)^{-1} A^\top$ is a projection matrix acting on b with the properties $P^2 P = P$ and $P^\top = P$. Briefly, this is what the *projection theorem* is all about (cf. [Goc10, 358–360]). Obviously, $A\hat{x} = p$ indicates $\|b - Ax\|_2$ to be minimal which is equivalent to minimizing $\|b - Ax\|_2^2$. We call $\hat{x}$ the *least squares solution* which minimizes the residuals or errors $\|e\|_2^2$ (cf. [Goc10, Theorem 291]).

Remark. (cf. [Sad08, 170] and [OS06, 196f]) Actually, we could choose any p-norm in order to minimize the residual e, e.g., the 1-norm $\|e\|_1 = \sum_{i=1}^{n} |e_i|$ or the maximum norm $\|e\|_\infty = \max\{|e_i|\}_{i=1}^{n}$. The reason why we have decided to work with the Euclidean norm $\|e\|_2^2$, endowed with the standard inner product $\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$, is that other norms would lead to nonlinear minimization problems which are much more complicated to handle. We prefer linear algebra when solving LSQ-minimization problems. ◇

Now, consider a *rank-deficient* matrix A with $m > n$ and $rank(A) < n < m$. Subsequently, $A^\top A$ is singular and A has linearly dependent columns. We cannot solve $Ax = b$ by direct means. We look for a vector x^+ which has minimal norm and is the unique least squares solution to $Ax = p$ and $Ax \approx b$, respectively. Since $x \in \mathbf{R}^n$, the general solution to a linear system is given by $x = x_r + x_s$ where $x_r \in C(A^\top)$ and

$x_s \in N(A)$. The system $Ax_r = p$ is uniquely solvable due to $Ax_s = 0$. Moreover, it follows from the pythagorean formula for right triangles that $\|x\|_2^2 = \|x_r\|_2^2 + \|x_s\|_2^2$. By setting $\|x_s\|_2^2 = 0 \Leftrightarrow \|x_s\|_2 = 0 \Leftrightarrow x_s = 0$, we obtain the *minimal norm solution* denoted by $x^+ = x_r$. It remains to find a way to compute x^+ satisfying $Ax^+ = p$.

We have already introduced how to deal with rectangular, ill-conditioned matrices. The SVD of such a matrix is the key to the right solution:

Definition 4. (cf. [Dem97, 126f]) Suppose that A is a rectangular matrix with $m > n$ and $rank(A) = r < n < m$, with $A = U\Sigma V^\top$ being the SVD of A, respectively. If $r < n$, the SVD reduces to $A = U\Sigma V^\top = U_r\Sigma_r V_r^\top$ where Σ_r is r-by-r and nonsingular and U_r and V_r have r columns. Then

$$A^+ = V_r\Sigma_r^{-1}U_r^\top \tag{2.2}$$

is called the *Moore-Penrose pseudoinverse* of a rank-deficient matrix A. We may also write $A^+ = V^\top\Sigma^+U$ where $\Sigma^+ = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}^+ = \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}$.

Remark. (cf. [Usm87, 84]) Let A be any m-by-n matrix. An n-by-m matrix A^+ is called the *generalized inverse* of A if and only if A^+ satisfies one or more of the following properties:

$$\begin{array}{lrcl} \textbf{1.} & AA^+A & = & A \\ \textbf{2.} & A^+AA^+ & = & A^+ \\ \textbf{3.} & (AA^+)^\top & = & AA^+ \\ \textbf{4.} & (A^+A)^\top & = & A^+A \end{array}$$

Furthermore, A^+ is said to be the pseudoinverse of A if and only if A^+ satisfies all the four enlisted properties. ◇

We have gathered enough information in order to solve the rank-deficient least squares problem. The "shortest" solution to $Ax \approx b$ is $x^+ = A^+b$. We call x^+ the *minimal norm least squares* solution (MNLSQ). Any other solution x_0 solving $A^\top Ax = A^\top b$ is longer than x^+. The projection of b onto the column space, i.e., the closest vector to b in the column space, is given by $p = Ax^+ = AA^+b$ where AA^+ represents the projection matrix acting on b. The reader might get a better view of the pseudoinverse by considering the following figure (cf. [Str03, 396]):

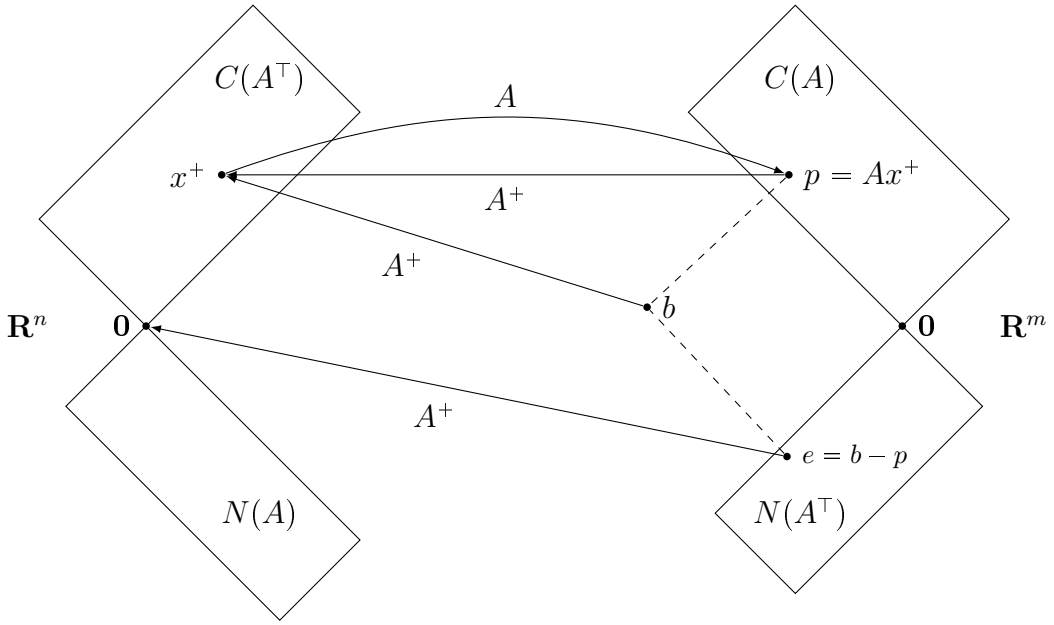


Figure 2.3: The actions of the pseudoinverse A^+.

There are plenty of applications of least squares problems such as *line/curve fitting*, *statistical* and *geodetic modeling* (cf. [Dem97, 101]). We want to focus on line/curve fitting which maybe is the most common application among (MN)LSQ-problems, e.g., forecasting and predicting future events are crucial to statistics, in particular. Given some data, say $(x_1, y_1), \ldots, (x_n, y_n)$, we initially want a line to go through each point in the plane. Typically, this is impossible as our data points are not colinear, i.e., there is no straight line matching them exactely. Thus we look for the *best* line $y = a_1 x + a_0$ fitting n outputs. This is equivalent to solving the equations

$$\begin{aligned} a_0 + a_1 x_1 &= y_1 \\ a_0 + a_1 x_2 &= y_2 \\ &\vdots \\ a_0 + a_1 x_n &= y_n \end{aligned} \tag{2.3}$$

or, briefly, $A\xi = b$ where

$$A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}; \quad \xi = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}; \quad b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

By the equations given in (2.1), we compute

$$A^\top A = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum (x_i)^2 \end{pmatrix} = n \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix},$$
$$A^\top b = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} = n \begin{pmatrix} \bar{y} \\ \overline{xy} \end{pmatrix} \tag{2.4}$$

with

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \ , \ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \ , \ \overline{x^2} = \frac{1}{n} \sum_{i=1}^{n} x_i^2 \ , \ \overline{xy} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i \tag{2.5}$$

denoting the *mean values* of the indicated variables (cf. [Sad08, 172f]). Finally, we have to solve a 2-by-2 system of equations with two unknowns a_0 and a_1. The solution (cf. [OS06, 198]) is

$$a_0 = \bar{y} - a_1 \bar{x} \quad \text{and} \quad a_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}. \tag{2.6}$$

We have figured out ξ which minimizes the least-squares error $\|e\|_2^2 = \|A\xi - b\|_2^2$. Hence, the best straight line fitting the given data in the least squares sense is

$$y = a_1(x - \bar{x}) + \bar{y}. \tag{2.7}$$

The slope of the line is given in (2.6) (cf. [OS06, 198]).

Remark. (cf. [Bos10, 161–164]) In statistics this phenomenon is called *linear regression*. Provided that (X, Y) is a two-dimensional stochastic variable, the regression line of y with respect to x is given by (2.7). We call a_1 the (empirical) regression coefficient of y with respect to x where $s_{xy} = \sum (x_i - \bar{x}) y_i$ denotes the *covariance* and $s_x^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ is the *variance* of a random sample of n observations $x_1, \dots, x_n$. The residual $\|e\|_2^2$ becomes $(n-1)(1-r^2)s_y^2$ (cf. [Bos10, 163]), where $s_y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$ and $r = \frac{s_{xy}}{s_x s_y}$ is called *correlation coefficient* of x and y. Obviously, r lies in the interval $[-1; 1]$. For further properties of r (cf. [Bos10, 134]). ◇

One might be dissatisfied with a simple linear model. We could replace the linear

function $y = a_1x + a_0$ by a parabola $y = a_2x^2 + a_1x + a_0$ or, generally, by a polynomial of n^{th} degree, i.e., $y = a_nx^n + \ldots + a_1x + a_0$. Now, the best curve fitting the data points is obtained by solving following system of equations:

$$\begin{aligned} a_0 + a_1x_1 + \ldots + a_nx &= y_1 \\ a_0 + a_1x_2 + \ldots + a_nx^2 &= y_2 \\ &\vdots \\ a_0 + a_1x_n + \ldots + a_nx^n &= y_n. \end{aligned} \tag{2.8}$$

In matrix notation, we have $V\xi = b$ with

$$V = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{pmatrix}; \quad \xi = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}; \quad b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}. \tag{2.9}$$

The $n \times (n+1)$ coefficient matrix V is called *Vandermonde matrix*. By setting $n := n+1$, V becomes square and invertible for distinct $x_1, \ldots, x_{n+1}$. In this case, $V\xi = b$ has a unique solution indicating $\|e\|_2^2$ to be zero. The solution is an *interpolating polynomial* which fits the data exactly (cf. [OS06, 201f]). Interpolation and LSQ-methods are useful means to approximate functions like $\sqrt{x}$, $\sin x$ or e^x by polynomials. Nevertheless, high-degree polynomials might be badly behaved which could have a negative impact on the columns of the Vandermonde matrix V. They consitute a poor basis and V often is ill-conditioned, albeit we divide each column vector by its norm and make them unit vectors (cf. [Str07, 441,587f]). Hence, high-degree polynomial interpolation is not used in practical applications (cf. [OS06, 205]).

Example. (cf. [Sad08, 176f]) The city of Austin in Texas was founded in 1839. The historical population is given in the following table:

Date	Population	Date	Population
1850	629	1940	87,930
1860	3,494	1950	132,459
1870	4,428	1960	186,545
1880	11,013	1970	251,808
1890	14,575	1980	345,890
1900	22,258	1990	465,622
1910	29,860	2000	656,562
1920	34,876	2010	790,390
1930	53,120	2020	best approx.

Table 2.3: Population of Austin in Texas since 1850.

Throughout this example, we will work with MATLAB by analyzing the measured data. We will use the data from 1850 through 1950 to find a model for Austin's population growth, and use the data from 1960 through 2010 to test our models. Initially, we want the best *linear* fit and are interested in estimating the population from 1960 to 2010. The typical approach to finding the best line is given by the normal equations in (2.4) and (2.1), respectively. Thus

$$A^\top A\xi = \begin{pmatrix} 11 & 20900 \\ 20900 & 39721000 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 394642 \\ 761879610 \end{pmatrix} = A^\top b$$

where A contains two linearly independent vectors with `ones(11,1)` generating an 11-by-1 column vector with ones on the one hand and the first 11 data entries (until 1950) from the table on the other hand. Hence, the LSQ-solution is

$$\xi = (A^\top A)^{-1} A^\top b = \begin{pmatrix} -2047181.5455 \\ 1096.3464 \end{pmatrix} \tag{2.10}$$

with the LSQ-line $y \approx 1.1 \cdot 10^3 x - 2.05 \cdot 10^6$ minimizing the total squared error $\|e\|_2$ with $\|e\|_2 = \|(A\xi - b)\|_2 \approx 6.1 \cdot 10^4$. For convenience, me might scale the date vector `(1850:10:1950)'` such that all values lie in the interval $[-1; 1]$. Following figure illustrates the best line versus the measured data:

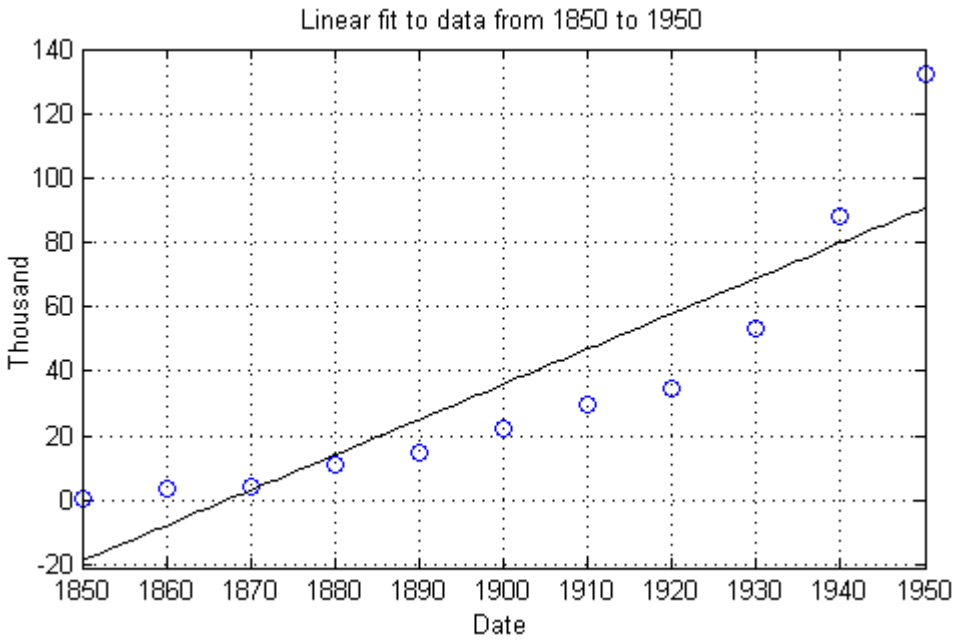


Figure 2.4: Linear LSQ-approximation.

There are some convenient commands in MATLAB generating the best line in a least-squares sense, anyway. The commands can be found in the appendix. Actually, the command `polyfit` is based on `vander(x)`, i.e., the Vandermonde matrix V in (2.9) with respect to the basis $B_V = \{1, x, \ldots, x^{n-1}, x^n\}$. If we want to fit a line to some data, we just take the last two columns of V, i.e., `V1=V(:,n-1:n)` where `n=length(x)`. The solution of the overdetermined system $V1\xi \approx b$ is obtained by the backslash operator \ or by the pseudoinverse of $V1$,respectively. The poylnomials represented by $\{p(x) = a_1 x + a_0$ with $p(x) \in \mathbf{P}(\mathbf{R})\}$ constitute a two-dimensional subspace of $\mathbf{R}^{11}$ [cf. lecture notes: applied mathematics for SSTAP]. Hence, we have to solve the system in a LSQ-sense.

Referring to Figure 2.4, we are not satisfied with this linear model with respect to predicting future data. For the period from 1960 to 2010 it yields the vector $(1.0166, 1.1262, 1.2358, 1.3455, 1.4551, 1.5647) \cdot 10^5$. By comparing the sample values with the measured values, we notice the huge gap between them. And we conclude that linear growth is not a good model for Austin's population. Thus, we seek for the best curve beginning with a parabola. Further, we would like to consider polynomials of higher degree ($n > 2$).

Let us start with a polynomial of degree $n = 2$. Therefore, we repeat the commands above and just substitute 2 for n. Indeed, the parabola turns out to be a better fit in contrast to the straight line:

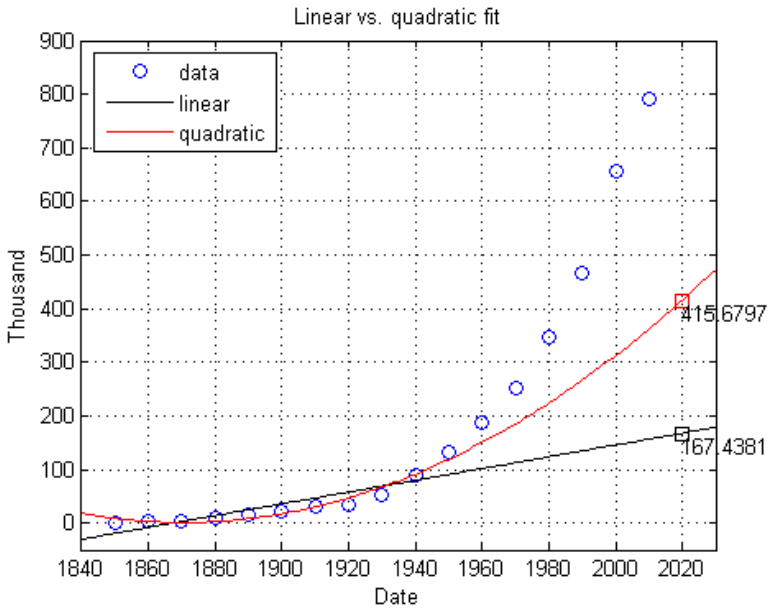


Figure 2.5: Two polynomials fitting the data in a LSQ-sense.

Well, it is better with respect to the estimated population for the year 2020 which we can read from the figure directly. We, however, find the quadratic model rather unreliable for future prediction as $415,680$ does not even reach the data from 1990. Thus, we wonder what happens if we allow more parameters, i.e., we consider polynomials of degree $n > 2$? A pictorial answer comes first:

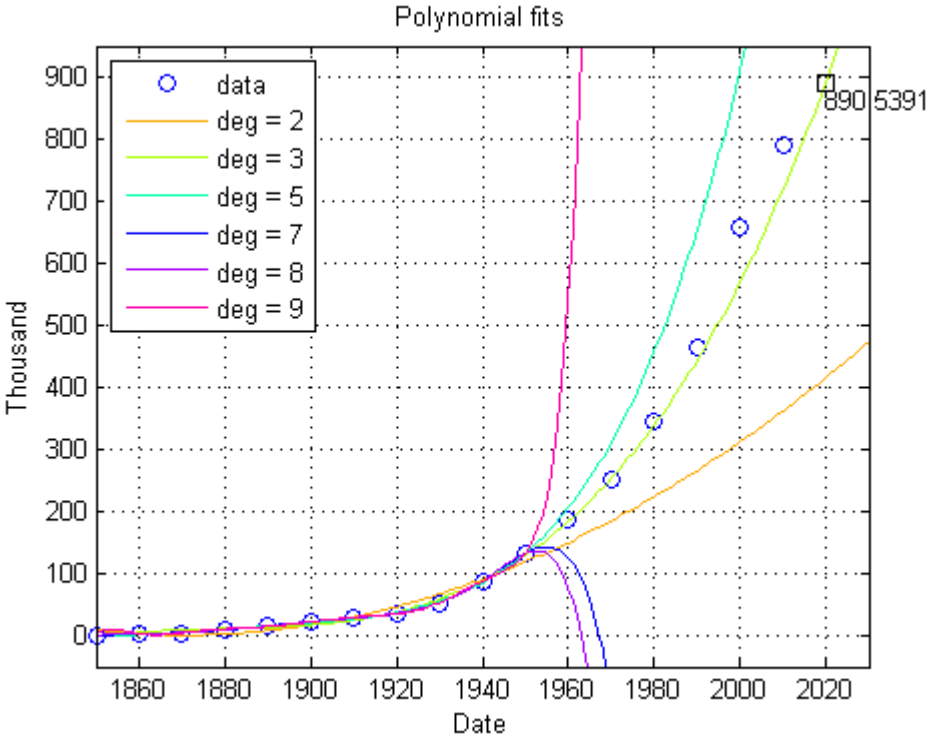


Figure 2.6: Polynomials of increasing degree fitting to original data.

As the degree of a polynomial increases, the more erratic it becomes. The cubic fit seems to be the best fit among the polynomials. It predicts a population of approximately $890,540$ for 2020, which we find quite possible. In contrast to the residual of the linear model (60.73 units), the total squared error decreased to 11.7 units. ⬠

3 An introduction to Signal (and Image) Processing

This section is devoted to an elementary introduction to signal (and image) processing. Initially, we provide some basics of (complex) vector and function spaces (ℓ_2, L^2), show their interconnection and say some words about finite and infinite dimensional spaces.

Next, we introduce three types of *Fourier* series and go into detail by considering some crucial properties of their (partial) sums and coefficients, respectively. This leads us to the *Discrete Fourier Transform*, DFT, and the fast DFT, i.e., the *Fast Fourier Transform*, FFT.

Finally, we work on function spaces when dealing with the Fourier integral $f(x)$ and the Fourier transform $\hat{f}(\omega)$. Further, we provide fundamental examples and applications and finish this section with a competitor to Fourier integral & Co., i.e., *wavelets*.

Throughout this section, our work is mainly based on [Sad08], [Str07], [Str03] and [OS06]. We recommend [Str07] to interested readers, who just aim at getting some idea about Fourier analysis, its applications, wavelets and how all of them contribute to signal and image processing (cf. [Str07, Chapt. 4]). [Str03] and [OS06] provide basics, whereas [OS06] confronts the reader with Fourier analysis from a view of point of signal processing from the very beginning. [Sad08] combines all three references but goes further by considering important properties of introduced subjects.

3.1 Basics

We have been working with euclidean norms and standard inner products on $\mathbf{R}^n$, which wholly fulfilled our purpose above. Nevertheless, we need to extend our "operation space" in order to introduce, e.g., (complex) Fourier series.

We obtain the complex norm of a complex vector z by multiplying by its conjugate transpose, i.e., $\|z\|^2 = \bar{z}^\top z = z^H z = |z_1|^2, |z_2|^2, \ldots, |z_n|^2$ equipped with the standard complex inner product $\langle u, v \rangle = u^H v = \overline{u_1} v_1 + \ldots + \overline{u_n} v_n$. The standard inner product on $\mathbf{C}^n$ is a sesquilinear form which is conjugate-symmetric and positive. It is called a positive Hermitian form (cf. [Sad08, 150]).

Analogous to (real) orthogonal matrices, there exist special complex matrices, i.e., Hermitian matrices, $A = \bar{A}^\top = A^H$ and unitary matrices U, respectively. A unitary matrix U has orthonormal columns.

When dealing with Fourier analysis, we will have to operate on infinite dimensional spaces. Therefore, we introduce the spaces ℓ_2 and L^2.

Definition 5. $\ell_2(\mathbf{R})$ is the space of infinte sequences of real scalars $x_1, x_2, x_3, \ldots$, such that the infinite sum $\sum_i |x_i|^2$ converges.

The infinite dimensional real vector space is denoted by $\ell_2(\mathbf{R})$. The inner product on $\ell_2(\mathbf{R})$ is given by $\langle x, y \rangle = \sum_{i=0}^{\infty} x_i y_i$ (cf. [Sad08, Theorem 6.8]). The space of infinite sequences of complex scalars x_i, $\ell_2(\mathbf{C})$, is defined similarly with inner product $\langle x, y \rangle = \sum_{i=0}^{\infty} x_i \bar{y}_i$. Since ℓ_2 is an infinite dimensional vector space where $\sqrt{\sum_i |x_i|^2} < \infty$, we can take infinite linear combinations (as long as they converge) from which we benefit a lot (see below).

Next, we consider a space of functions on real and complex subsets, denoted by $L^2(I, \mathbf{R})$ and $L^2(I, \mathbf{C})$, respectively, or even just L^2, "where an orthonormal basis makes the space look like ℓ_2" ([Sad08, 179]). The orthonormal basis will be sine and cosine functions and complex exponentials, and "the conversion from a function (in L^2) to an infinite set of coefficients (in ℓ_2) is called Fourier series" ([Sad08, 179]). This is what we discuss below.

Definition 6. Let I be a subset of $\mathbf{R}$. Then $L^2(I, \mathbf{R})$ is the space of real-valued functions $f(x)$ on I sucht that $\int_I |f(x)|^2 \, \mathrm{d}x < \infty$. $L^2(I, \mathbf{R})$ is a real inner product space with inner product $\langle f, g \rangle = \int_I f(x) g(x) \, \mathrm{d}x$. Analogous, we define $L^2(I, \mathbf{C})$, the space of complex valued functions $f(x)$ on I such that $\int_I |f(x)|^2 \, \mathrm{d}x < \infty$. Its inner product is $\langle f, g \rangle = \int_I \overline{f(x)} g(x) \, \mathrm{d}x$.

Spaces such as ℓ_2 and L^2, where the norm comes from an inner product, are called *Hilbert spaces* (cf. [Sad08, 178]).

3.2 Fourier series

First, we want a function $f \in L^2$ to decompose into an infinite sum of sines. Therefore, we consider the functions

$$B_k(x) = \sin kx \tag{3.1}$$

on the interval $[0, \pi]$. Their special property is that they form an orthogonal basis for L^2, i.e., they span the function space and are linearly independent. We just compute their inner product:

$$\begin{aligned} \langle B_k, B_l \rangle &= \int_0^\pi \sin kx \sin lx \, \mathrm{d}x \\ &= \frac{1}{2} \int_0^\pi \cos(k-l)x - \cos(k+l)x \, \mathrm{d}x \\ &= \begin{cases} \pi/2 & \text{if k=l} \\ 0 & \text{else.} \end{cases} \end{aligned} \tag{3.2}$$

Now, we expand the function $f(x)$ as an infinite linear combination of sines, i.e.,

$$f(x) = \sum_{i=1}^{\infty} b_k B_k = \sum_{i=1}^{\infty} b_k \sin kx. \tag{3.3}$$

Multiplying (3.3) by $\sin kx$ and integrating over $[0, \pi]$ yield the coefficients b_k:

$$b_k = \frac{\langle B_k, f \rangle}{\langle B_k, B_l \rangle} = \frac{2}{\pi} \int_0^\pi f(x) \sin kx \, \mathrm{d}x. \tag{3.4}$$

Taking inner products of functions f and $\tilde{f}$, both elements of L^2, shows how the "Fourier decomposition maps a function (an element of L^2) to an infinite set of coefficients (an element of ℓ_2), such that, aside from an overall factor of $[\pi/2]$, the inner prodcut is preserved" ([Sad08, 183]), i.e.,

$$\langle f, \tilde{f} \rangle = \left\langle \sum_k b_k \sin kx, \sum_k \tilde{b_k} \sin kx \right\rangle = \frac{\pi}{2} \sum_k \bar{b_k} \tilde{b_k}. \tag{3.5}$$

Analogous to $B_k(x)$, functions $A_k(x) = \cos kx$ satisfy (3.2) to (3.5). The expansion of $g \in L^2$ as an infinite linear combination of cosines is $g(x) = \sum_{i=0}^{\infty} a_k A_k = \sum_{i=0}^{\infty} a_k \cos kx$

with coefficients

$$a_0 = \frac{1}{\pi}\int_0^{\pi} g(x)\,\mathrm{d}x \quad \text{and} \quad a_k = \frac{2}{\pi}\int_0^{\pi} g(x)\cos kx\,\mathrm{d}x. \tag{3.6}$$

The constant term a_0 is the *average value* of the function $g(x)$ (cf. [Str07, 320]). The next step - this is the second type of Fourier series - is to combine the functions $f(x)$ and $g(x)$, i.e., we take linear combinations of both sines and cosines at once. Before doing that, we define our terms:

Definition 7 (Periodic functions)**.** A function $h(x)$ is said to be *periodic* with period T if $h(x+T) = h(x)\ \forall x$. It is said to satisfy *periodic boundary conditions* on $[0,T]$ if $h(0) = h(T)$ and $h'(0) = h'(T)$ (cf. [Sad08, 243]).

Subsequently, functions on the unit circle can be represented by periodic functions on the line with period T, in particular 2π and restricting the angle θ of a function on the unit circle, it becomes a function on $[0,T]$ with periodic boundary conditions (cf. [Sad08, 243]). The sines and cosines match this definition as they are periodic with period T. Thus, any linear combination is also periodic. In addition, every periodic function decomposes into a (infinite) sum of linear combinations of these trigonometric functions:

$$h(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{2\pi kx}{T}\right) + \sum_{k=1}^{\infty} b_k \sin\left(\frac{2\pi kx}{T}\right) \tag{3.7}$$

where $h \in L^2$. The coefficients a_k and b_k are given by

$$a_k = \frac{2}{T}\int_0^{T} h(x)\cos\left(\frac{2\pi kx}{T}\right)\mathrm{d}x \quad \text{and} \quad b_k = \frac{2}{T}\int_0^{T} h(x)\sin\left(\frac{2\pi kx}{T}\right)\mathrm{d}x. \tag{3.8}$$

(3.7) represents the most genereal case of a (infinite) Fourier series with period T and its coefficients (3.8). By setting $T = 2\pi$ we obtain (3.2) up to (3.6), the sine and cosine series on an particular interval (the factor multiplying a_k and b_k becomes $1/\pi$, clearly). Furthermore, we may write $h(x)$ as a sum of complex exponentials, i.e.,

$$h(x) = \sum_{k=-\infty}^{\infty} c_k \exp\left(\frac{2\pi ikx}{T}\right), \tag{3.9}$$

where

$$c_k = \frac{1}{T}\int_0^T h(x)\exp\left(\frac{-2\pi ikx}{T}\right)\mathrm{d}x. \tag{3.10}$$

It is not surprising that the functions $\left\{e^{2\pi ikx/T}\right\}$ form an orthogonal basis for L^2. We compute the inner product:

$$\begin{aligned}\left\langle \exp\left(\frac{2\pi ikx}{T}\right), \exp\left(\frac{2\pi ilx}{T}\right)\right\rangle &= \int_0^T \exp\left(\frac{2\pi i(k-l)x}{T}\right)\mathrm{d}x \\ &= \begin{cases} T & \text{if k=l} \\ 0 & \text{else.}\end{cases}\end{aligned} \tag{3.11}$$

With (3.3), (3.7) and (3.9), we have introduced three types of Fourier series. Indeed, the sets of functions $\{\sin kx\}$, $\{\sin(2\pi kx/T), \cos(2\pi kx/T)\}$ and $\{\exp(2\pi ikx/T)\}$ are orthogonal bases for the function space L^2. Actually, those functions are eigenfunctions which are obtained by solving the equation $A\xi = \lambda\xi$, where A denotes a Hermitian operator. We consider three different Hermitian operators (cf. [Sad08, 244f,254f]); first, d^2/dx^2 with Dirichlet boundary conditions, whose eigenvalues are $-k^2$, and whose eigenfunctions are $\{\sin kx\}$. Second, we use the same operator but with periodic boundary conditions. The eigenvalues are $-(2\pi k/T)^2$, and the eigenspace is given by $\{\sin(2\pi kx/T), \cos(2\pi kx/T)\}$. Finally, the third operator is $-id/dx$ with periodic boundary conditions. The eigenvalues are $2\pi k/T$ with eigenfunctions $\{\exp(2\pi ikx/T)\}$. For further details and a complete proof for the special functions to constitute a basis for L^2.

Before providing a specific and elementary example from signal processing, we have to deal with a special equation, namely $z^n = 1$. Therefore, we recall the complex numbers on the unit circle, i.e., $\zeta_n = \exp(2\pi i/n) = \cos(2\pi/n) + i\sin(2\pi/n)$, where $n = 1, 2, 3, \ldots$, being the so called *primitive n^{th} root of unity*. Hence, the n^{th} power of ζ_n is $\zeta_n^n = (\exp(2\pi i/n))^n = 1$. A complete set of roots of the polynomial $z^n - 1$ is given by the powers of ζ_n (cf. [OS06, 280f]):

$$\zeta_n^k = e^{2\pi ik/n} = \cos\left(\frac{2\pi k}{n}\right) + i\sin\left(\frac{2\pi k}{n}\right), \tag{3.12}$$

where $k = 0, 1, \ldots, n-1$. For example, we look for the 4^{th} roots of unity, i.e., we want solutions to $z^4 = 1$. Initially, $\zeta_4 = i$. It generates all the other solutions: 1 for $k = 0$, i for $k = 1$, -1 for $k = 2$ and $-i$ for $k = 3$.

It exactly is the *Discrete Fourier Transform (DFT)* which deals with the powers of ζ_n. Before going into detail, we need to introduce the *Fourier matrices* F_n and F_n^{-1} which produce the DFT and its inverse. We start with $n = 4$, in particular, in order to visualize the "shape" of a Fourier matrix:

$$F_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix} \tag{3.13}$$

In brief, we see each entry as a power of $\zeta_4 = i$. In genereal and more precisely, each entry in row j column k is $\omega^{jk} = (\zeta_n^k)^j = (e^{2\pi ik/n})^j$, where $j, k = 0, 1, \ldots, n-1$. Thus, the Fourier matrix with entries ω^{jk} is

$$F_n = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & \omega & \omega^2 & \ldots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \ldots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \ldots & \omega^{(n-1)^2} \end{pmatrix}. \tag{3.14}$$

The Fourier matrix F_n has orthogonal columns. It becomes unitary by dividing by its norm, e.g., $\frac{1}{2}F_4$ is unitary. It holds $(\frac{1}{\sqrt{n}}F_n^H)(\frac{1}{\sqrt{n}}F_n) = I$ and $\frac{1}{n}F_n^H = F_n^{-1}$ but F is not Hermitian by itself (cf. [Str07, 348] and [Str03, 491]).

With the DFT, we move from functions with infinite series to vectors $h_0, h_1, \ldots, h_{n-1}$ with finite sums. Subsequently, (3.9) becomes $h_j = \sum_{k=0}^{n-1} c_k \omega^{jk}$ which is tantamount to $h = F_n c$ (cf. [Str07, 349]). We reconstruct the vector h from its discrete Fourier coefficients. This procedure is known as the *Inverse DFT (IDFT)*. The DFT which transforms h to c is done by the inverse of the Fourier matrix with conjugate complex entries $\overline{\omega}^{jk}$. Here, the complex coefficients c_k in (3.10) become $c_k = \frac{1}{n}\sum_{j=0}^{n-1} h_j \overline{\omega}^{jk}$ which is $c = F_n^{-1}h$.

Example (Noise removal)**.** (cf. [OS06, Exercise 5.7.13]) Let $f(x) = x(2\pi - x)$ be sampled on $n = 128$ equally spaced points between 0 and 2π. We construct a corrupted signal by a random number generator in MATLAB, denoise it and afterwards compare the original signal with the new output. We have recorded the MATLAB commands of following figures in the appendix.

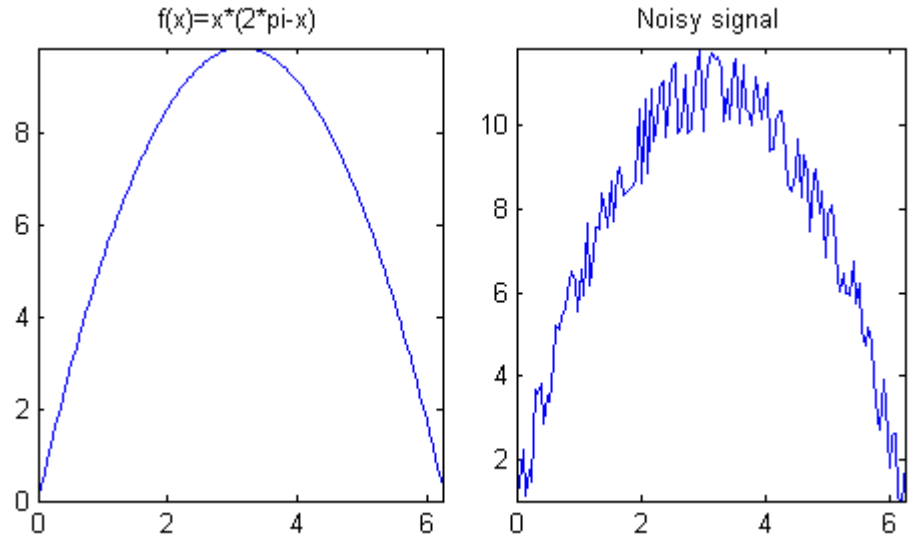


Figure 3.1: The original signal $f(x)$ and the noisy signal.

In order to denoise the corrupted signal we replace (3.9) by $\bar{h}(x) = \sum_{k=-l}^{l} c_k e^{ikx}$ and just consider the $2l+1 \ll n$ low frequency Fourier terms (l is much smaller than n as $l < \frac{1}{2}(n+1)$). They will eliminate high frequency, i.e., noise, and retain essential features of $f(x)$ (cf. [OS06, 285]). Hence, we do not use all of the 128 coefficients c_k but, for instance, $2l+1 = 3$, $2l+1 = 5$ or $2l+1 = 11$ coefficients. We have plotted the denoised signals below:

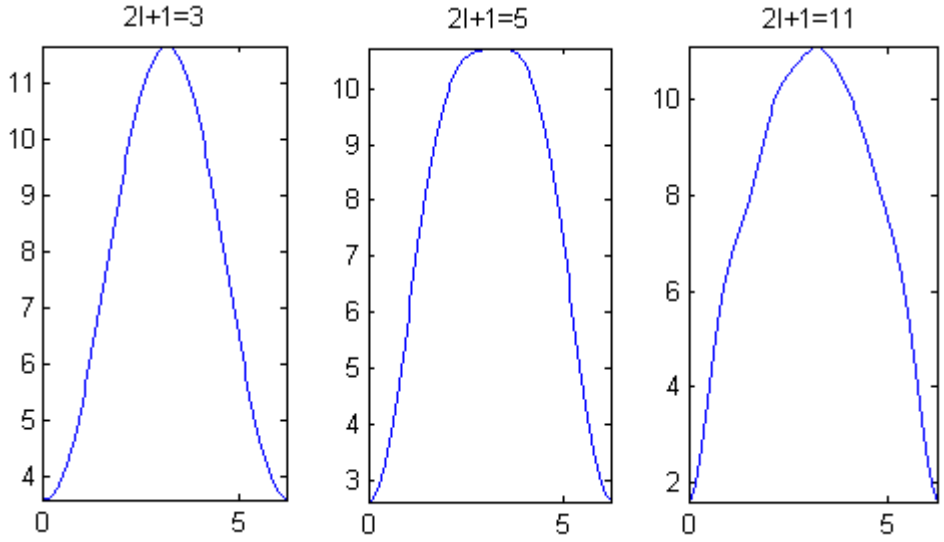


Figure 3.2: The denoised signals with $3, 5$ and 11 Fourier terms.

Of course, we used the (fast) DFT to obtain the coefficients, first. Those plots were generated by $\bar{h}(x) = \sum_{k=-l}^{l} c_k e^{ikx}$, where $l = 1, 3$ and 5. Finally, we compare the

denoised signals with the original signal:

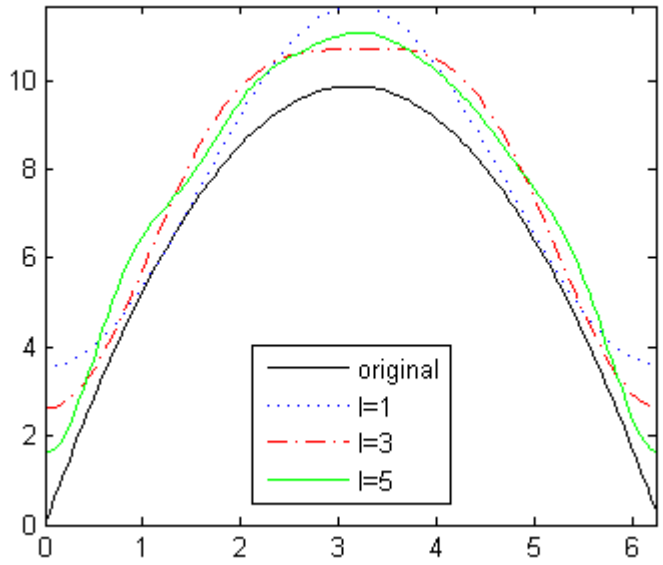


Figure 3.3: Comparison of the original signal with the denoised signals.

We admit that the denoised signals do not perfectly fit the original, but they constitute a good reconstruction of it. Notice that we only made use of $3, 5$ and 11 coefficients, respectively. The maximum error L^∞ of the difference of, e.g., the dotted representation and the original signal is $\|f(x) - \bar{h}_3\|_\infty = \max\{|f(x) - \bar{h}_3|,\ x \in [0, 2\pi]\} \approx 3.6$. The more terms we add, the worse we do. The first 3 and the first 5 coefficients give reasonable results. ⬠

We would like to accelerate the reconstruction of a signal from its coefficients. Normally, we need n^2 complex mulitiplications when multiplying c by F_n (this is the minimal amount of complex operations as we did not consider complex additions). Factorizing F_n would speed up computations. We are talking about a fast algorithm for the DFT, the *Fast Fourier Transform (FFT)*. The key idea is to connect F_n with $F_{n/2}$ (cf. [Str07, 350]). We again consider the Fourier matrix for $n = 4$ and its half-size matrix F_2 and connect them:

$$F_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{pmatrix} \quad \text{with} \quad \begin{pmatrix} F_2 & \\ & F_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & & \\ 1 & i^2 & & \\ & & 1 & 1 \\ & & 1 & i^2 \end{pmatrix}.$$

Obviously, those matrices are not the same. The full factorization of F_4 is

$$F_4 = \begin{pmatrix} 1 & & 1 & \\ & 1 & & i \\ 1 & & -1 & \\ & 1 & & -i \end{pmatrix} \begin{pmatrix} 1 & 1 & & \\ 1 & i^2 & & \\ & & 1 & 1 \\ & & 1 & i^2 \end{pmatrix} \begin{pmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{pmatrix},$$

where the right matrix is a permutation matrix which separates the coefficients into evens and odds, the half-size matrix in the middle acts on this permutation matrix and the matrix at the left "combines the two half-size outputs, in a way that produces the correct full-size output $[h = F_4 c]$" ([Str07, 350]).

The first step of the FFT for the general case follows the same idea (cf. [Str07, 351]). We have

$$F_n = \begin{pmatrix} I_{n/2} & D_{n/2} \\ I_{n/2} & -D_{n/2} \end{pmatrix} \begin{pmatrix} F_{n/2} & \\ & F_{n/2} \end{pmatrix} \begin{pmatrix} even - odd \\ permutation \end{pmatrix}. \tag{3.15}$$

$I_{n/2}$ is the identity matrix, $D_{n/2}$ denotes the diagonal matrix with entries $(1, \omega, \ldots, \omega^{\frac{n}{2}-1})$ using the n^{th} roots of unity. The half size matrices $F_{n/2}$ use following important relation: $\omega_{n/2} = \omega_n^2$. The permutation matrix is as introduced in the particular case which separates the vector c into $c' = (c_0, c_2, \ldots)$ and $c'' = (c_1, c_3, \ldots)$. Thus, $h = F_n c$ has constituents $h' = F_m c'$ and $h'' = F_m c''$, where $m = n/2$. To reconstruct h, we look at following equations:

$$\begin{aligned} h_j &= h'_j + \omega^j h''_j \\ h_{j+1} &= h'_j - \omega^j h''_j \end{aligned}, \tag{3.16}$$

where $j = 0, 1, \ldots, m-1$. To sum up, at each FFT step we first split the vector c into c' and c'', we compute h' and h'' and finally reconstruct h by (3.16). This is only one stage. Actually, there are $log_2 n$ steps (coming from $n = 2^r$) (cf. [OS06, 291]). Thus, we have "a total of $r \cdot n = n \cdot \log_2 n$ complex additions/subtractions and the same number of multiplications" ([OS06, 291]). Let us consider, e.g., $n = 2^{10}$. Then $1024 \cdot 10 = 10,240$ is much smaller than 1024^2. This is the reason why many books refer to the FFT as an "industrial" revolutionizer. For the complete FFT and some examples, we refer to [Str07, 351f], [OS06, 287–292], [Mut06, 280–289] and [GW99, Lesson 9], respectively. The latter is more accurate.

3.3 Fourier transform and signals

We want to decompose a function f into its basic constituents, i.e., sines and cosines, or complex exponentials, respectively, on the entire real line $\mathbf{R}$. It is not completely different from what we introduced above, however, it is more general (cf. [Sad08, 281]).

Definition 8. Let $f \in L^2(\mathbf{R})$. Then

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega t} \, \mathrm{d}\omega, \tag{3.17}$$

where

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} \, \mathrm{d}t. \tag{3.18}$$

The decomposition of a signal in the *time domain* (3.17) is called *Fourier integral*. The complex valued function $\hat{f}(\omega)$ is called the *Fourier transform* of the function $f(t)$. In signal processing $\hat{f}(\omega)$ describes the frequency behavior of the signal $f(t)$. Here, we speak of working in the *frequency domain* (cf. [Sad08, 287]).

There are some properties (without proofs) of Fourier transforms which need to be considered. We start with a proposition which says to treat the function $f(t)$ and its corresponding tranfsorm $\hat{f}(\omega)$ on an equal footing (cf. [Sad08, 287], [GW99, 193]).

Proposition (The Plancherel-Parseval equality)**.** Let $f, g \in L^2(\mathbf{R})$. Then following equations hold:

$$\int_{-\infty}^{\infty} \overline{f(t)} g(t) \, \mathrm{d}t = \int_{-\infty}^{\infty} \overline{\hat{f}(\omega)} \hat{g}(\omega) \, \mathrm{d}\omega \tag{3.19}$$

and

$$\int_{-\infty}^{\infty} |f(t)|^2 \, \mathrm{d}t = \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 \, \mathrm{d}\omega. \tag{3.20}$$

Theorem 6. (cf. [Str07, 370], [Sad08, 287] and [Pau07, 4f]) Let $f \in L^2(\mathbf{R})$, let $f_a(t) = f(t-a)$ be the *translated* function, let $g_b(t) = e^{ibt} f(t)$ be the *modulated*

function, and let $m(t) = \delta f/\delta t$. Then following holds:

$$\begin{aligned} \hat{f}_a(\omega) &= e^{-i\omega a}\hat{f}(\omega) \\ \hat{g}_b(\omega) &= \hat{f}(\omega - b) \\ \hat{m}(\omega) &= i\omega\hat{f}(\omega). \end{aligned} \tag{3.21}$$

In words, translating a function by a is tantamount to multiplying its transform by $e^{-i\omega a}$, or, to put it more concrete, " shift of f changes phase of $\hat{f}$" ([Str07, 371]). Next, a phase change of f causes $\hat{f}$ to shift. Third, the derivative of a function is the same as multiplying the transform by $i\omega$.

Example (Gaussian)**.** (cf. [Sad08, 282f], [Str07, 371]) Let $f(t) = \exp(-t^2/2)$. The transform is

$$\begin{aligned} \hat{f}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} e^{-i\omega t}\,\mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(t^2+2\omega i t)/2}\,\mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(t+i\omega)^2/2} e^{-\omega^2/2}\,\mathrm{d}t \\ &= \frac{e^{-\omega^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2}\,\mathrm{d}t \\ &= e^{-\omega^2/2}, \end{aligned} \tag{3.22}$$

where $u = t + \omega i$. We put this fascinating result into words: "the Fourier transform of a standard normal distribution is a standard normal distribution" ([Sad08, 283]). We could have computed the transform by using, e.g., the third property of Theorem 6 (cf. [Str07, 371]). ⬠

In the discrete case, one might be interested in the Fourier coefficients of the product of two functions f and g. It is not the product of c_k and d_k, however, it is the *convolution* of those coefficients, i.e., $\sum c_k d_{n-k}$ (cf. [Str07, 357]). Below, we introduce the convolution for functions f and g (from $\mathbf{R}$ to $\mathbf{C}$) and their transforms.

Definition 9. The *convolution* of two functions f and g is the function $f * g$ defined by

$$f * g(t) = \int_{-\infty}^{\infty} f(t-y)g(y)\,\mathrm{d}y. \tag{3.23}$$

The convolution of the transforms $\hat{f}$ and $\hat{g}$ is similarly defined:

$$\hat{f} * \hat{g}(\omega) = \int_{-\infty}^{\infty} \hat{f}(\omega - \omega')\hat{g}(\omega')\,\mathrm{d}\omega'. \tag{3.24}$$

There is following relevant interplay between convolution, multiplication and transform:

Theorem 7. (cf. [Sad08, Theorem 10.4]) Let $f, g, fg, f * g \in L^2$. Then

$$\widehat{f * g}(\omega) = \sqrt{2\pi}\hat{f}\hat{g} \quad \text{and} \quad \widehat{fg}(\omega) = \frac{1}{\sqrt{2\pi}}\hat{f} * \hat{g}. \tag{3.25}$$

Section (2.2) was devoted to compressing an image by means of the SVD. We mentioned that computing orthogonal matrices might be costly and that we need sparse matrices where computation is faster and cheaper. Now, we introduce *filters* which actually are convolutions. They constitute the key step in signal and image processing (cf. [Str07, 361]). We work with periodic signals and present two specific filters, namely the *lowpass* filter and the *highpass* filter. Indeed, these shall only be representative examples. We again refer interested readers as well as people with elementary background knowledge to [Str07] and, particularly, [SV04] and [GW99], respectively.

We start by defining a filter, informally. In brief, "a filter is a continuous, translation-invariant system" ([GW99, 14]), i.e., a shift by a vector does not change anything. Filters operate in the frequency domain and let some frequencies pass, while others are attenuated or even blocked (cf. [SV04, 295]).

The first example is a "second average" lowpass filter L (cf. [Str07, 361]. It has centered coefficients $c_{-1} = \frac{1}{4}$, $c_0 = \frac{1}{2}$ and $c_1 = \frac{1}{4}$. The output at time n is given by $y_n = \frac{1}{4}x_{n-1} + \frac{1}{2}x_n + \frac{1}{4}x_{n+1}$, which just is $l * x$, where $l = \frac{1}{4}(.,1,2,1,.)$ or, equivalently, Lx. L denotes a *Toeplitz* matrix with

$$L = \frac{1}{4}\begin{pmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{pmatrix}. \tag{3.26}$$

The lowpass response is denoted by $L(\omega)$, where $L(\omega) = \frac{1}{4}e^{-i\omega} + \frac{1}{2} + \frac{1}{4}e^{i\omega} = \frac{1}{2}(1 +$

$\cos\omega$). Hence, the lowest frequency $\omega = 0$ is passed unchanged through the filter which corresponds to the DC input $x = (1, 1, 1, 1)$. The highest frequency $\omega = \pi$ corresponding to the AC input $x = (1, -1, 1 - 1)$ is blocked. Here, the response is the zero vector. In between, the inputs $(1, i, -1, -i)$ and $(1, -i, -1, i)$ at $\omega = \pm\frac{\pi}{2}$ have outputs multiplied by $\frac{1}{2}$. Please notice that the numbers $\lambda = 0, 1, \pm\frac{1}{2}$ are eigenvalues of the filter matrix L. They satisfy $Lx = \lambda x$ with inputs x. One might use a lowpass filter like L in order to remove noise and thus high frequency from a signal (cf. [Str07, 362]).

The second example of a filter is about the "second difference" highpass filter H (cf. [Str07, 364]). Analogous to L, we use centered coefficients $c_{-1} = -\frac{1}{4}$, $c_0 = \frac{1}{2}$ and $c_1 = -\frac{1}{4}$. We have the output at time n with $y_n = -\frac{1}{4}x_{n-1} + \frac{1}{2}x_n - \frac{1}{4}x_{n+1}$, which is $h * x$, where $h = (., -1, 2, -1, .)$, or simply Hx. H again is a Toeplitz matrix with

$$H = \frac{1}{4}\begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix}. \tag{3.27}$$

The highpass response is given by $H(\omega) = -\frac{1}{4}e^{-i\omega} + \frac{1}{2} - \frac{1}{4}e^{i\omega} = \frac{1}{2}(1 - \cos\omega)$. $L(\omega)$ and $H(\omega)$ are fairly related, i.e., $H(\omega)$ is $L(\omega)$ shifted by π. Subsequently, the lowest frequency $\omega = 0$ is stopped by the filter, whereas the highest frequency $\omega = \pi$ is passed. At $\omega = \pm\frac{\pi}{2}$ we have the same outputs as above.

After having introduced a usefool tool (not only) for signal and image processing, we want to finish this section with a competitor to Fourier analysis, namely *wavelets*.

Given a function $\psi \in L^2(R)$ with $\|\psi\|_2 = 1$. It is called the *analyzing wavelet* or *mother wavelet* (cf. [GW99, Lesson 42]), if it satisfies

$$K = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|}{|\omega|}\,\mathrm{d}\omega < \infty, \tag{3.28}$$

where $\hat{\psi}$ denotes the transform of the wavelet. We construct the family of functions

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \tag{3.29}$$

where $a, b \in \mathbf{R}$ and $a \neq 0$. We infer that this family of functions is generated by

translations and dilations.

Definition 10. Fix ψ and let $f \in L^2(\mathbf{R})$. The *Wavelet transform* of f with respect to ψ denoted by $\mathcal{W}_\psi f$ is

$$\mathcal{W}_\psi f(a,b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{\infty} f(t)\overline{\psi_{a,b}(t)}\,\mathrm{d}t. \tag{3.30}$$

The function $\mathcal{W}_\psi f$ is continuous and bounded. The number $\mathcal{W}_\psi f(a,b)$ is called wavelet coefficient of f with respect to ψ in (a,b). The signal $f(t)$ is reconstructed by the *Inverse Wavelet transform*:

$$f(t) = \frac{1}{K} \int_{\mathbf{R}} \int_{\mathbf{R}} \frac{1}{a^2} \mathcal{W}_\psi f(a,b) \psi\left(\frac{t-b}{a}\right) \mathrm{d}a\,\mathrm{d}b. \tag{3.31}$$

We do not further go into detail by introducing other theorems, propositions and examples. What we want to do is to explain why wavelets have become a very usefull mathematical tool in applications and why it was, to our mind, necessary to give a faint idea of it.

The main idea underlying the Fourier transform is to "transform" a signal from its time domain to its frequency domain. The inverse is obtained by the Fourier integral. Thus, we consider a signal either in the *t-domain* or in the *ω-domain*. We cannot have it both ways (cf. *Heisenberg's Uncertainty Principle*, e.g., in [GW99, 197f]). In addition, the Fourier bases are global and only have a locality in the frequency domain. Moreover, they give a poor approximation at a jump (cf. *Gibbs phenomenon*, e.g., in [Str07, 318f]). To put it more concrete: if we want to represent signals and images "in a sparse and piecewise smooth way" ([Str07, 389]), we need to choose bases which are local, easily refined, fast to compute, and which yield a good approximation by a few terms (cf. [Str07, 389]). Wavelets achieve all of them. They represent signals in time and frequency. They capture high frequencies over short times and lower frequencies over longer times (cf. [Str07, 390]).

The most common application to wavelets is data compression. For instance, the FBI preferred wavelet methods for compressing digital fingerprints instead of a proposal based on JPEG (*Joint Photographic Experts Group*). For more information and details we again refer to some literature: [Str07, Chapt. 4.7], [Fra99], [MMP07], [SV04] and

[GW99]. [Fra99] introduces wavelets by (not necessarily) elementary linear algebra, [MMP07] and [SV04] provide many applications worked with MAPLE and MATLAB, especially [SV04], without going to deep into mathematical questions. [GW99] is all about fundamentals of signal processing.

4 "Applicable" Mathematics and Didactics

In this short section, we first introduce some general results of an empirical survey about "Teaching applications in mathematics education", which were presented by J. Humenberger in his habilitation thesis (in 1997). Afterwards, we have a look at syllabi of secondary academic schools (*Allgemeinbildende Höhere Schule*, AHS) and technichal high schools (*Höhere Technische Lehranstalt*, HTL) and want to know to which extent linear algebra and its applications are prevalent. Next, we focus on the math-textbook "Ingenieur-Mathematik 4", which is commonly used in technical high schools. We finish this section with some thoughts, ideas, comments and questions concerning the teaching and didactics of mathematical applications at (technical) high schools.

4.1 General part

J. Humenberger wrote his dissertation about "Fundamental ideas of Applied Mathematics", which was extended to a scientific contribution, i.e., a book with the same title (cf. [HR95]). According to the authors, this book constitutes a first approach to fundamental ideas of applied mathematics as a complete subject (cf. [HR95, V]) (at least restricted to the German area). This is the reason why we have explicitely chosen this author. In his habilitation thesis, Humenberger presents the results of an empirical investigation, where teachers, mathematics students and pupils were asked questions about "applications in mathematics teaching" (cf. [Hum97, 3]). We do not intend to present all of it, however, we are interested in following problems:

- Applicable and non-applicable mathematical topis (opinions of students and pupils).

- Computer use at school.
- Teachers' and pupils' opinion on teaching and resp. learning applied mathematics.

Both students as well as pupils have similar point of views about the first issue. They find percent calculation and stochastics (= probability theory + statistics) usefull, while integral and differential calculus, complex numbers and curve sketching and vector calculus, respectively, are among the least applicable mathematical topics.

Moreover, 74% of pupils never worked with computers in their math-lessons. Only a quarter of the teachers "sometimes/occasionally" ([Hum97, 45]) let their pupils work at a computer.

Third; almost 71% of invesitaged teachers find mathematical applications useful as these help pupils to better realize the purpose and essence of mathematics. But they complain about the complexity of "non-mathematical areas", i.e., physics, engineering, biology, etc., that even they themselves do not know much about non-mathematical contents.

What is interesting - and almost clear - is the pupils' reply to following question: "What do you like most at mathematics?". They answered: "Problems, where you can see how math is applicable" (55%).

4.2 Syllabi

We consider two different kinds of syllabi; first, the syllabus of secondary academic schools (AHS) and second, the syllabus of technical high schools (HTL). According to the homepage of the *Austrian Federal Ministry for Education, Arts and Culture*, there are 14 subject areas pupils might attend (HTL). We are interested in the syllabus of *electronics and electrical engineering* as this contains the most amount of mathematics hours (16 distributed over five years). We start with the first.

The AHS-syllabus does not explicitely distinguish between linear algebra and other mathematical topics (analysis, etc.). Following topics ($9^{th} - 12^{th}$ grades) can be assigned to linear algebra: (systems of) equations, trigonometry ($9^{th} grade$); vector calculus in $\mathbf{R}^2$ and $\mathbf{R}^3$ (9^{th} and 10^{th} grades); complex numbers and conic section (11^{th}

grade). The 12^{th} grade syllabus for math does not contain any topics of linear algebra and geometry.

The HTL-syllabus is more rich in content with respect to linear algebra. The 9^{th} and 10^{th} grades math contains vector calculus ($\mathbf{R}^2$ and $\mathbf{R}^3$), linear (in)equations and systems of equations, trigonometry and complex numbers. There is a logical reason behind introducing complex numbers that early. Since the subject area is electrical engineering, the complex number j in electronics (engineering, in genereal) which is i in mathematics has an overriding importance (especially for engineers).

The 11^{th} grade students are presented error propagation and estimation, numerical methods for solving equations, interpolation and difference equations. The next grade contains Fourier sums, matrices, conic section and algebraic structures. Analogous to the AHS-syllabus, there is no linear algebra content in the 13^{th} grade at all.

For more and detailed information visit the homepage of the Ministry for Education, Arts and Culture, which is provided in the *Bibliography* below.

4.3 Ingenieur Mathematik 4 - An analysis

"Ingenieur-Mathematik 4" (ger., "mathematics for engineers") is a mathematical textbook which is commonly used at technical high schools. We have analysed its approach to Fourier analysis.

It starts by introducing infinite sums (chapter 3) and thus Fourier sums (cf. [TK05, 101]). The authors provide coefficient properties and decay rates of Fourier coefficients and even say a word about the *Gibbs phenomenon* without explicitely mentioning it (cf. [TK05, 105f]).

Further, the book presents "Numerical Fourier Analysis" beginning with *Shannon's Sampling Theorem*. It provides formulas for the numerical computation of Fourier coefficients and introduces the FFT. This chapter ends here but chapter 4 is all about an important application of Fourier analysis, namely *differential equations*.

Chapter 5 is wholly devoted to "Transforms and Signals" and constitutes an introduction to signal processing (especially with examples from electrical engineering). We encounter basic signals like the step function $\sigma(t)$ and the *Delta impulse* $\delta(t)$. The

Fourier transform and some of its properties, and the Laplace transform follow next. Differential equations and electrical circuits (cf. [TK05, 223–232]) are introduced as applications to the Laplace transform.

We see that this approach to Fourier analysis is not completely elementary anymore. This book does not only contain formulas but provides essential explanations to introduced subjects. In fact, it intends at educating young engineers who become an asset for companies. Actually, each student of a technical high school was granted an "Ingenieur"-title along with the diploma until recently. Now, they need to work three years in a field which matches their subject area they attended at school (electrical engineering, mechanical engineering, mechatronics, etc.) in order to become an engineer.

4.4 Comments, ideas, remarks and questions

At this place, we want to review this work including the survey of J. Humenberger, the syllabi and the mathematical textbook analysed above with respect to mathematics education.

Almost 15 years have passed since the empirical investigation of Humenberger was carried out. Indeed, this is a long period of time and we wonder to which extent the inclusion of mathematical applications into math lessons has happened (in Austria). While the number in 1997 of computer using students was 74%, we are convinced that this number has decreased as *new-generation* teachers make use of computers in their classes and lessons (this is an assumption as we lack any scientific statistics). For instance, the mathematical software *GeoGebra* has become a popular computer algebra system. It particularly appeals to students of secondary academic schools. Furthermore, we know from some students of technical high schools of various subject areas that they use *MATLAB* and *Maple* in order to cope with complex computations and simulate physical, electrical, etc. processes.

When considering the mathematical contents of both syllabi, we realize a huge gap in between. One might take this for granted as AHS provides general and HTL specific (technical) education. While the overwhelming majority of AHS-students enrols at university, HTL-students prefer working in their subject area. However, the situation of latter students has started to change since the labor market has become more com-

petitive. Nowadays, companies do not wholly prefer HTL-students with a technical high school diploma anymore. More and more HTL-students continue their studies at university in their subject area. They often have a great background knowledge (referring to the syllabus and, e.g., "Ingenieur Mathematik 4") which facilitates a lot - at least in the first year after enrolment. The reader might wonder why we write about this. We think of AHS-students who want to study mathematics, electrical engineering, and other technical courses who actually are at disadvantage compared to their colleagues from a technical high scholl. We want to end this series of thoughts with some (provocative) questions: What is the ideal syllabus which does not neglect both the AHS-syllabus and the HTL-syllabus? Do HTL-students really need that much math, nowadays, where so much students with a high school diploma enrol at university? How about workshops, courses, introductory seminars, etc. in holidays for future mathematics students, especially about applied mathematics?.

Back to the survey: it is interesting that mathematics students and pupils found integral, differential and vector calculus and complex numbers not applicable. Indeed, this applies more to AHS-students than HTL-students. To our mind, this is a serious issue which needs to be considered properly. High school students often wonder where they need math and for which purpose "something mathematical" is introduced. This is due to the abstract world of mathematics which causes students trouble as they cannot "feel" mathematical abstractions.

From our vantage point, it is indispensable to incorporate technology, i.e., hand-held devices, CAS, etc., into math lessons. Further, we do not think that applied mathematics is a subject to be introduced only in special, time-consuming projects and other organisations. The *adequate* education for teachers, the *ideal* syllabus, the *technical environment* in classes (schools) and *motivated* students are the *only* prerequisites to make mathematics fun and interesting.

Appendix

Linear LSQfit:

```
xi=1850:10:1950;
b=[629 3494 4428 11013 14575 22258 29860 34876 53120 87930 132459];
p=polyfit(xi,b,1);
l=linspace(1850,1950);
v=polyval(p,l);
plot(l,v,'k-',xi,b,'bo')
grid on
```

or alternatively:

```
x=[1850;1860;1870;1880;1890;1900;1910;1920;1930;1940;1950];
y=[0.629;3.494;4.428;11.013;14.575;22.258;29.86;34.876; ...
53.12;87.93;132.459];
xsc=(x-1900)/50; % scaled x domain
V=vander(xsc);
n=length(V);
xil=pinv(V(:,n-1:n))*y; % coef. for lin. function
xin=1850:1950;
sxin=(xin-1900)/50;
vxin=polyval(xil,sxin);
grid on
hold on
plot(x,y,'bo',xin,vxin,'k-')
xlabel('Date')
ylabel('Thousand')
title('Linear fit to data from 1850 to 1950')
```

```
axis([1850 1950 -21 140])
```

Linear vs. quadratic fit:

```
x=[1850;1860;1870;1880;1890;1900;1910;1920;1930;1940;1950];
y=[0.629;3.494;4.428;11.013;14.575;22.258;29.86;34.876; ...
53.12;87.93;132.459];
xc=[1850;1860;1870;1880;1890;1900;1910;1920;1930;1940; ...
1950;1960;1970;1980;1990;2000;2010];
yc=[0.629;3.494;4.428;11.013;14.575;22.258;29.860;34.876; ...
53.12;87.93;132.459;186.545;251.808;345.89;465.622;656.56;790.39];
xsc=(x-1900)/50; % scaled x domain
V=vander(xsc);
n=length(V);
xcn=1840:2030; % instead linspace
axis([1840 2030 -50 900]);
hold on
plot(xc,yc,'bo'); % xc; abbrev. for x-complete (1850:2010)
and yc; abbrev. for y-complete (629 to 790.390);
xil=pinv(V(:,n-1:n))*y; % coef. for lin. function
xip=pinv(V(:,n-2:n))*y; % coef. for parabola
sxcn=(xcn-1900)/50; % scaled xcn
vxil=polyval(xil,sxcn); % evaluate line at sxcn
vxip=polyval(xip,sxcn); % evaluate parabola at sxcn
w=(2020-1900)/50;
wvl=polyval(xil,w); % estimate for 2020 (linear)
wvp=polyval(xip,w); % estimate for 2020 (quadratic)
plot(xcn,vxil,'k-'); % plot line
plot(xcn,vxip,'r-'); % plot parabola
plot(2020,wvl,'ks'); % plot square at x=2020 and y=wvl
plot(2020,wvp,'rs'); % plot square at x=2020 and y=wvp
text(2020,wvl-25,num2str(wvl)); % add text
text(2020,wvp-25,num2str(wvp)); % add text
xlabel('Date');
ylabel('Thousand');
title('Linear vs. quadratic fit');
```

```
legend('data','linear','quadratic',2);
grid on
```

Polynomial fits:

```
x=[1850;1860;1870;1880;1890;1900;1910;1920;1930; ...
1940;1950];
y=[0.629;3.494;4.428;11.013;14.575;22.258;29.86; ...
34.876;53.12;87.93;132.459];
xc=[1850;1860;1870;1880;1890;1900;1910;1920;1930; ....
1940;1950;1960;1970;1980;1990;2000;2010];
yc=[0.629;3.494;4.428;11.013;14.575;22.258;29.860; ...
34.876;53.12;87.93;132.459;186.545;251.808;345.89; ...
465.622;656.56;790.39];
xsc=(x-1900)/50; % scaled x domain
V=vander(xsc);
n=length(V);
xcn=1840:2030; % instead linspace
sxcn=(xcn-1900)/50; % scaled xcn
w=(2020-1900)/50;
plot(xc,yc,'bo');
hold on
axis([1850 2030 -50 950]);
col=hsv(9);
for d=[2 3 5 7 8 9]
xid=pinv(V(:,n-d:n))*y;
vxid=polyval(xid,sxcn);
plot(xcn,vxid,'color',col(d,:));
end
legend('data','deg = 2','deg = 3','deg = 5',...
'deg = 7','deg = 8','deg = 9',2)
xlabel('Date')
ylabel('Thousand')
title('Polynomial fits')
xi3=pinv(V(:,n-3:n))*y;
vxi3=polyval(xi3,w);
```

```
plot(2020,vxi3,'ks');
text(2020,vxi3-30,num2str(vxi3));
grid on
```

Original signal, corrupted signal, denoised signals, and comparison:

```
%original signal%

t=0:pi/64:127*pi/64;
f=2*pi*t-t.^2;

%corrupted signal%

fn=f+2*rand(size(t));

%plot of original signal and corrupted signal%

hold on
subplot(1,2,1)
plot(t,f,'b-')
axis tight
title('f(x)=x*(2*pi-x)')
subplot(1,2,2)
plot(t,fn,'b-')
axis tight
title('Noisy signal')

%denoised signals with coefficients%

c=fft(eye(128))*fn';

fd3=c(128)*exp(-i*t)+c(1)+c(2)*exp(i*t);
fd5=c(128)*exp(-i*t)+c(127)*exp(-2*i*t)+c(1)+c(2)*...
exp(i*t)+c(3)*exp(2*i*t);
fd11=c(128)*exp(-i*t)+c(127)*exp(-2*i*t)+c(126)*...
exp(-3*i*t)+c(125)*exp(-4*i*t)+...
```

```
c(124)*exp(-5*i*t)+c(1)+c(2)*exp(i*t)+c(3)*exp(2*i*t)+...
c(4)*exp(3*i*t)+c(5)*exp(4*i*t)+c(6)*exp(5*i*t);

%subplot of denoised signals%

hold on
subplot(1,3,1)
plot(t,fd3,'b-')
title('2l+1=3')
axis tight
subplot(1,3,2)
plot(t,fd5,'b-')
axis tight
title('2l+1=5')
subplot(1,3,3)
plot(t,fd11,'b-')
axis tight
title('2l+1=11')
hold off

%comparison of denoised signal with original signal%

hold on
plot(t,f,'k-',t,fd3,'b:',t,fd5,'r-.',t,fd11,'g-')
axis tight
legend('original','l=1','l=3','l=5')
```

Bibliography

[AR04] E.S. Allmann and J.A. Rhodes. *Mathematical Models in Biology. An Introduction.* Cambridge University Press, Cambridge, 2004.

[BCC01] F. Bauer and C. Castillo-Chávez. *Mathematical Models in Population Biology and Epidemiology.* Springer-Verlag, New York, 2001.

[Beh00] E. Behrends. *Introduction to Markov Chains with special Emphasis on Rapid Mixing.* Vieweg, Wiesbaden, 2000.

[Bos10] K. Bosch. *Elementare Einfürhung in die angewandte Statistik.* Vieweg+Teubner Verlag, Wiesbaden, 2010.

[Bro91] R. Bronson. *Matrix Methods. An introduction.* Academic Press, San Diego, 1991.

[Dem97] J.W. Demmel. *Applied Numerical Linear Algebra.* Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[Fra99] M.W. Frazier. *An Introduction to Wavelets Through Linear Algebra.* Springer-Verlag, New York, 1999.

[FW79] B. Fritz, F.-J.; Huppert and W. Willems. *Stochastische Matrizen.* Springer-Verlag, Berlin, Heidelberg, New York, 1979.

[Goc10] M.S. Gockenbach. *Finite-Dimensional Linear Algebra.* CRC Press, Boca Raton, London, New York, 2010.

[Gra04] G. Gramlich. *Anwendungen der Linearen Algebra.* Carl Hanser Verlag, München, Wien, 2004.

[GW99] W. Gasquet and P. Witomski. *Fourier Analysis and Applications. Filtering, Numerical Computation, Wavelets.* Springer-Verlag, New York, 1999.

[HR95] J. Humenberger and H.C. Reichel. *Fundamentale Ideen der Angewandten Mathematik und ihre Umsetzung im Unterricht.* Wissenschaftsverlag, Mannheim, Leipzig, Wien, Zürich, 1995.

[Hum97] H. Humenberger. Anwendungsorientierung im mathematikunterricht - erste resultate eines forschungsprojekts. *Journal für Mathematik-Didaktik*, 18:3–50, 1997.

[MMP07] Misiti M., G. Misiti, Y.; Oppenheim, and J.-M. Poggi, editors. *Wavelets and their Applications.* ISTE Ltd, London, 2007.

[Mut06] H.J. Muthsam. *Lineare Algebra und ihre Anwendungen.* Elsevier GmbH, München, 2006.

[OS06] P.J. Olver and C. Shakiban. *Applied Linear Algebra.* Pearson Prentice Hall, New Jersey, 2006.

[Pau07] S. Paukner. Foundations of gabor analysis for image processing, 2007.

[Sad08] L. Sadun. *Applied Linear Algebra. The Decoupling Principle. Second Edition.* American Mathematical Society, Rhode Island, 2008.

[SB97] G. Strang and K. Borre. *Linear Algebra, Geodesy, and GPS.* Wellesley-Cambridge Press, Wellesley, MA, 1997.

[Sen81] E. Seneta. *Non-negative Matrices and Markov Chains. Second Edition.* Springer-Verlag, Berlin, Heidelberg, New York, 1981.

[Str86] G. Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Wellesley, MA, 1986.

[Str03] G. Strang. *Introduction to Linear Algebra.* Wellesley-Cambridge Press, Wellesley, MA, 2003.

[Str07] G. Strang. *Computational Sciene and Engineering.* Wellesley-Cambridge Press, Wellesley, MA, 2007.

[SV04] W. Strampp and E.V. Vorozhtsov. *Mathematische Methoden der Signalverarbeitung.* Oldenbourg Verlag, München, Wien, 2004.

[TK05] W. Timischl and G. Kaiser. *Ingenieur-Mathematik 4.* E. Dorner GmbH, Wien, 2005.

[Usm87] R.A. Usmani. *Applied Linear Algebra.* Marcel Dekker, New York, Basel, 1987.

Austrian Federal Ministry for Education, Arts and Culture:

`http://www.bmukk.gv.at`

`http://www.bmukk.gv.at/medienpool/11859/lp_neu_ahs_07.pdf`

`http://www.bmukk.gv.at/schulen/bw/bbs/Berufsbildende_Schulen_H1745.xml`

List of Figures

List of Tables

Abstract

This thesis combines Linear Algebra with Applied Mathematics. The author presents mathematical applications from physics and electrical engineering, statistics and demography, signal and image processing by means of linear algebra.

The first chapter presents three fundamental laws of Kirchhoff and Ohm, respectively, which combine into the *fundamental network equation*. Therefore, the author considers the four subspaces, graphs and incidence matrices.

The second chapter is all about applications to eigenvalues and eigenvectors. The author introduces *Markov chains*, the *singular value decomposition* and the method of *(minimal norm) least squares*. Statistics and demographic processes, image compression and (minimal norm) least squares applications, respectively, are among mathematical applications introduced in this chapter.

The third chapter deals with basics of signal and image processing. The author provides an introduction to Fourier analysis and discusses some important tools for signal processing, e.g., *filters*.

The last chapter constitutes an attempt to review this thesis with respect to mathematics education, teaching and, in general, didactics.

Zusammenfassung

Die vorliegende Diplomarbeit kombiniert elementare Methoden der Linearen Algebra mit der Angewandten Mathematik. Die Lineare Algebra wird dabei als unverzichtbares Werkzeug der modernen Anwendungsgebiete der Mathematik verstanden.

Das erste Kapitel beschäftigt sich mit drei fundamentalen Gesetzen der Physik und Elektrotechnik von Kirchhoff beziehungsweise Ohm, die sich zu einer Gleichung zusammenfassen lassen, nämlich der *Netzwerkgleichung*. Hierfür werden die vier Teilräume, Graphen und Inzidenzmatrizen in Betracht gezogen.

Das zweite Kapitel widmet sich der Eigenwerte und Eigenvektoren. *Markov Ketten*, die *Singulärwertzerlegung* und die Methode der *kleinsten Quadrate (minimaler Länge)* werden vorgestellt. Statistik und demographische Prozesse, Bildkompression beziehungsweise Anwendungen der kleinsten Quadrate (minimaler Länge) sind unter den mathematischen Anwendungen, die hier zutrage kommen.

Das dritte Kapitel beschäftigt sich mit Grundlagen der Bild- und Signalverarbeitung. Es wird in die Fourier-Analysis eingeleitet und einige wichtige Elemente der Signalverarbeitung, wie *Filter*, besprochen.

Das letzte Kapitel unternimmt einen Versuch diese Arbeit in die Diskussion über Angewandte Mathematik in der Schule einzubinden.

Curriculum Vitae

Name:	Bayram ÜLGEN
Born:	August 17th, 1986
Nationality:	Austria
E-mail:	bayram.uelgen@hotmail.de

1992–1996:	Primary school (Gasserg. 46)
1996–2005:	Secondary academic school (Haydn Bundesrealgymnasium), high school diploma (Matura) with distinction
2005–2006:	Community service
2006–2011:	Study of *Secondary School Teacher Accreditation Program* in Mathematics (major) and History (minor) at the University of Vienna