

Kapitel 1

Inverse problems: an introduction

When the imprecise is preciser

Kirchgraber-Kirsch-Stoffer

A central problem in finance is the calibration of the models with the help of the market data. It is the aim of the following chapters to sketch the methods which can be used to compute/estimate the volatility from market data. This estimation problem is an inverse problem. In this first chapter we give an introduction to the theory of inverse problems.

1.1 Inverse terminology

First of all, we have to clarify the notation „inverse“ by mathematical terms.

1.1.1 Problem classification

A **model** is an image of reality formulated in terms which are well defined and can be handled by theoretical investigations. Usually, a model is developed by considering the relation between causes and effects in the process under consideration.

Suppose that we have a mathematical model of a physical (biological, . . .) process. We assume that this model gives a description of the **system behind the process** and its operating conditions and explains the principal quantities of the model:

input, system parameters, output

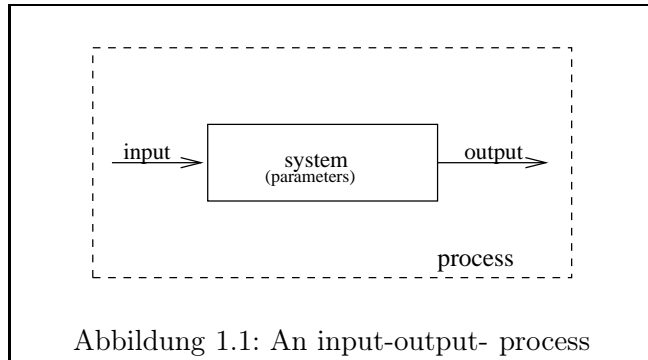
The analysis of the given process via the mathematical model may then be separated into three distinct types of problems; see Figure 1.1.

- (A) **The direct problem.** Given the input and the system parameter, find out the output of the model.
- (B) **The reconstruction problem.** Given the system parameters and the output, find out which input has led to this output.
- (C) **The identification problem.** Given the input and the output, determine the system parameters which are in agreement with the relation between input and output.

We call a problem of type (A) a direct (or forward) problem since it is oriented towards a cause-effect sequence. In this sense problems of type (B) and (C) are called inverse problems because they are problems of finding out unknown causes of known consequences. It is immediately clear

that the solution of one of the problems above involves a treatment of the other problems as well. A complete discussion of the model by solving the inverse problems is the main objective of **inverse modelling**.

Usually, the system is governed by a process operator transforming the input into the output. Moreover, the laws of nature are often expressed as systems of differential equations containing parameters describing the concrete physical situation; in a dynamical consideration the physical quantities are functions of time. These equations are **local** in the sense that the physical quantities (velocity, pressure, current,...) depend on a neighborhood of the current time only. Another typical feature of the laws is **causality**: later conditions depend on previous ones. Locality and causality are features typically associated with direct problems. Inverse problems on the other hand are most often **nonlocal** and **noncausal**. These properties contribute to the fact that inverse problems are sensitive to the input data and parameters.



Let us give a description of the input, the output, the parameters and the systems by mathematical terms:

- X : space of input quantities;
- Y : space of output quantities;
- \mathbb{P} : space of system parameters;
- $A(p)$: system operator from X into Y associated to $p \in \mathbb{P}$.

Later on we will use methods and results from analysis, linear algebra and functional analysis to describe the quantities $X, Y, \mathbb{P}, A(p)$ in various applications.

In these terms we may formulate the problems (A), (B), (C) above in the following way:

- (A) Given $x \in X$ and $p \in \mathbb{P}$, find $y := A(p)x$.
- (B) Given $y \in Y$ and $p \in \mathbb{P}$, find $x \in X$ such that $A(p)x = y$.
- (C) Given $x \in X$ and $y \in Y$, find $p \in \mathbb{P}$ such that $A(p)x = y$.

At first glance, the direct problem (A) seems to be solved much easier than the inverse problems (B), (C). However, for the computation of $y := A(p)x$ it may be necessary to solve a differential or integral equation, a task which may be of the same complexity as the solution of the equations in the inverse problems.

Inverse modelling involves the estimation of the solution of an equation from a set of observed data represented by y, p in (B) and y, x in (C) respectively. In applications, the knowledge of just one „observation“ y is not sufficient to determine p in problem (B). It is the subject of the *identification theory* to find out „how many data“ y are sufficient to determine p in a uniquely way (identifiability).

The theory to solve (B) or (C) falls into two distinct parts. One deals with the ideal case in which the data are supposed to be known exactly and completely (**perfect data**). The other treats the practical problems that are created by incomplete and imprecise data (**imperfect data**). It might be thought that an exact solution to an inverse problem with perfect data would

prove also useful for the practical case. But in general, it turns out in inverse problems that the solution obtained by the analytic formula obtained from the analysis of the forward problem is very sensitive to the way in which the data set is completed and to errors in it. Moreover, when the inverse problem is described by an operator-equation, the naive way to solve the equation by a fit of the data to the range of the related operator does not give necessarily good results: As a rule of thumb: *a good fit to the data does not mean a good fit to the unknown solution.*

There are three concepts that are often confused by modelers, in particular in financial and economic modeling: identification, simulation and extrapolation:

- **Identification** is building a model from the available (inexact) data.
- **Simulation** is experimenting with the model to produce new, additional data.
- **Extrapolation** is producing new data with a model beyond the reference period of the available data.

Example 1.1 *Let us consider the problem to finding the weighted integral of a given function. A mathematical description is given as follows:*

Forward Problem: *With a continuous function $x : [0, 1] \rightarrow \mathbb{R}$ and a parameter $p > 0$ compute*

$$y(t) := \int_0^t e^{-ps} x(s) ds, \quad t \in [0, 1].$$

Reconstruction Problem: *Given a differentiable function $y : [0, 1] \rightarrow \mathbb{R}$ and a parameter $p > 0$ determine $x : [0, 1] \rightarrow \mathbb{R}$ such that*

$$y(t) = \int_0^t e^{-ps} x(s) ds \quad \text{for all } t \in [0, 1].$$

Parameter Identification Problem: *Given a differentiable function $y : [0, 1] \rightarrow \mathbb{R}$ and a continuous function $x : [0, 1] \rightarrow \mathbb{R}$, find a parameter $p > 0$ such that*

$$y(t) = \int_0^t e^{-ps} x(s) ds \quad \text{for all } t \in [0, 1].$$

The term e^{-pt} may be considered as a discount factor.

As we know from integration theory, the forward problem can be solved analytically and numerically in a very stable way. Here we are mainly interested in the two inverse problems. Clearly, $y(0) = 0$ is a necessary condition for solvability of the inverse problems.

The reconstruction problem is equivalent to

$$y'(t) = e^{-pt} x(t) \quad \text{for all } t \in [0, 1].$$

Therefore, we get back the function x by

$$x(t) = y'(t) e^{pt}, \quad t \in [0, 1].$$

Notice that we have to differentiate the data y .

The parameter identification problem is equivalent to

$$y'(t) = e^{-pt} x(t) \quad \text{for all } t \in [0, 1].$$

Clearly, the parameter is overdetermined – we will come back to this classification – by this set of equations: we need just one $\hat{t} \in (0, 1)$ where

$$p = -\frac{1}{\hat{t}} \ln \left(\frac{y'(\hat{t})}{x(\hat{t})} \right)$$

makes sense. Again, we have to differentiate the data y . □

The prototype of an inverse problem will be an equation of the form

$$F(x) = y \quad (x \in X, y \in Y) \tag{1.1}$$

with a mapping F from the space X into the space Y . For such an equation, the unknown is x and the data are usually the right-hand side y . If the stability condition is violated, the numerical solution of the inverse problem by standard methods is difficult and often yields instability, even if the data are exact (since any numerical method has internal errors acting like noise). Therefore, special techniques, so called regularization methods have to be used in order to obtain a stable approximation of the solution. The appropriate construction and analysis of regularization methods and subsequently (or simultaneously) of numerical schemes is the major issue in the solution of inverse problems.

1.2 A list of inverse Problems

Let us start with a very short historical view on topics related to Inverse problems.

The term *inverse problem* is a well known mathematical term dating at least from the 19th century.

- For many centuries people have been searching for hiding places by tapping walls and analyzing echo; this is a particular case of an inverse problem.
- *Platon*¹ is talking in his *allegory of cave* in a philosophical context about the reconstruction of reality from observations of shadows (distorted blurred images from objects outside the cave) on a wall. Today we may consider Platon's problem as an inverse problem to solve.
- The problem which may be considered as one of the oldest inverse problem in applied science is the computation of the diameter of the earth by *Eratosthenes*² in 200 b. Chr.; see [6].
- In 1800 *Gauss*³ used the **method of least squares** to reconstruct the orbit of a comet from earlier orbit data. This method became fundamental in the applied sciences.
- When astronomer *Urbain Le Verrier* worked out the mathematics to successfully predict where the planet Neptune would be discovered in the night sky back in 1846, he was really solving an inverse problem. By that, he used the observations that had been recorded about Uranus' position in its orbit to infer how that orbit might have been affected by what was then a hypothetical eighth planet in the solar system.

¹Platon, 427 - 347

²Eratosthenes, 284 – 202

³C. F. Gauss, 1777 – 1855

- In 1917 *Radon*⁴ published a paper entitled by „Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten“; see [20]. In this paper a transformation is studied which is nowadays called the **Radon transform** and which is the basis of the X-ray tomography.
- It was *Heisenberg*⁵ who conjectured that quantum interaction was totally characterized by its scattering matrix which collects information of the interaction (at infinity). The discovery of neutrinos by measuring consequences of its existence is in the spirit of inverse problems too.
- The weakly electric nocturnal fish *Gnathonemus petersii* is well known for its ability to use its electro-sensory system in the dark for active electrolocation of objects; see [4].

Over the past 40 years, the number of publications on inverse problems has grown rapidly. Nowadays there are several mathematical journals devoted to this topic. Especially, the subject of image processing and non-invasively, non-destructively mapping became very important (*Making the unseen visible with mathematics*).

The following list of problems gives a good impression of the wide variety of applications. Several of the listed problems may be seen under the motto *making the invisible visible*.

- Estimation of a correlation matrix from data;
- the inverse problem of geomagnetic induction;
- X-ray tomography, ultrasound tomography, laser tomography;
- acoustic scattering, scattering in quantum mechanics;
- inverse problem of elastography;
- radio-astronomical imaging, image analysis;
- locating cracks or mines by electrical prospecting;
- seismic exploration, seismic tomography;
- the use of electrocardiography and magnetocardiography;
- evolution backwards in time, inverse heat conduction;
- the inverse problem of potential theory;
- „can you hear the shape of a drum/manifold?“
- deconvolution, reconstruction of truncated signals;
- compartmental analysis, parameter identification;
- data assimilation;
- determining the volatility in models for financial markets;
- discrete tomography, shape from probing.

1.3 Ill-posedness/Well-posedness

In a complete analysis of an inverse problems – we denote the collection of methods and results in this field by „inversion theory“ – the questions of **existence**, **uniqueness**, **stability** and **construction** of objects have to be considered.

⁴J. Radon, 1887 – 1956

⁵W. Heisenberg, 1901 – 1976

1.3.1 Definition of well-posedness

The question of existence and uniqueness is of great importance in testing the assumption behind any mathematical model. If the answer in the uniqueness question is no, then we know that even perfect data do not contain enough information to recover the physical quantity to be estimated. By questioning for stability we have to decide whether the solution depends continuously on the data. Stability is necessary if we want to be sure that a variation of the given data in a sufficiently small range leads to an arbitrarily small change in the solution. This concept was introduced by *Hadamard* in 1902 in connection with the study of boundary value problems for partial differential equations and he designated unstable problems **ill-posed**⁶ otherwise **well-posed**. The nature of inverse problems (irreversibility, causality, unmodelled structures, ...) leads to ill-posedness as a characteristic property of these problems.

When solving ill-posed problems (numerically), we must certainly expect some difficulties, since any error acts as a perturbation on the original equation and so may cause arbitrarily large variations in the solution. Since errors cannot be completely avoided, there may be a range of plausible solutions and we have to find out a reasonable solution. These ambiguities in the solution of inverse problems which are unstable can be reduced by incorporating some sort of **a-priori information** that limits the class of allowable solutions. By a-priori information we mean an information which has been obtained independently of the observed values of the data. This a-priori information may be given as a deterministic or a statistical information. We shall restrict ourselves mainly to deterministic considerations.

As we already have seen, an inverse problem may be formulated as the problem to solve an equation governed by an operator $A(p)$. Ill-posedness tells us that the inverse $A(p)^{-1}$ does not exist and/or is not continuous. The remedy is **regularization**. The idea of the regularization theory is to replace the inverse of such an operator by a one-parameter family of continuous operators and to choose the „best approximation“ in this family by a clever strategy. Such a **regularization strategy** is usually based on a *merit function* and uses two main ingredients for solving an inverse problem in a stable way: a-priori information and signal to noise ratio (SNR). The **signal to noise ratio** is a quantity which describes the relation of the size of the true solution (signal, image, ...) to the size of the noise contained in the measured quantity (right-hand side of the equation,...).

The Hilbert space l_2 and operators on l_2 (see Appendix 1.6 for informations concerning these objects) are used to formulate toy-problems; see the example below. In the course of the lecture we will see that these toy-problems may be seen at least from the theoretical point of view in a one-to-one correspondence to applied problems via the singular value decomposition. Let us present a first example using this elementary framework.

Example 1.2 Consider the problem of solving the equation

$$Ax = y$$

where A is a linear mapping from l_2 into l_2 ; see Appendix 1.6. Therefore, x and y are sequences in l_2 :

$$x = (x_1, x_2, \dots), Ax = y = (y_1, y_2, \dots);$$

⁶Hadamard believed – many mathematicians still do – that ill-posed problems are actually incorrectly posed and artificial in that they would not describe physical systems. He was wrong! Nowadays we know that such problems arise in a fundamental way in the modelling of complex (physical) systems.

Consider the specific case

$$Ax := (x_1, \frac{1}{2}x_2, \frac{1}{3}x_3, \dots), x \in \ell_2.$$

A is bounded since

$$\|Ax\| = \left(\sum_{k=1}^{\infty} \frac{1}{k^2} x_k^2 \right)^{\frac{1}{2}} \leq \left(\sum_{k=1}^{\infty} 1 x_k^2 \right)^{\frac{1}{2}} = \|x\|$$

and the candidate of the inverse of A is given by

$$A^{-1}y := (y_1, 2y_2, 3y_3, \dots).$$

Existence of a solution: For the right-hand side

$$y := (1, \frac{1}{2}, \frac{1}{3}, \dots)$$

no solution (in ℓ_2) exists!

Uniqueness of a solution: It is easy to see that uniqueness holds.

Stability: For the right-hand side

$$y^n := (0, 0, \dots, 0, \frac{1}{\sqrt{n}}, 0, 0, \dots)$$

the solution is given by

$$x^n := (0, 0, \dots, 0, \sqrt{n}, 0, 0, \dots)$$

and we have

$$\lim_n \|y^n\|_2 = 0, \lim_n \|x^n\|_2 = \lim_n \sqrt{n} = \infty.$$

Therefore A^{-1} cannot be a bounded operator!

Regularization: Approximate the mapping A by $A_\alpha := A + \alpha I$ where I is the identity operator from ℓ_2 into ℓ_2 . Then the inverse A_α^{-1} of A_α is given as follows:

$$A_\alpha^{-1}y = (\frac{1}{1+\alpha}y_1, \frac{2}{1+2\alpha}y_2, \frac{3}{1+3\alpha}y_3, \dots).$$

Clearly A_α^{-1} is a bounded operator. Now, A_α^{-1} should approximate the inverse A^{-1} . The question arises which is the best parameter α with respect to the approximation property. \square

1.3.2 The evaluation problem

A well-conditioned problem is, roughly speaking, one whose solution changes by only a small amount if the problem data are changed by a small amount; see above. In order to study this property in mathematical terms it is enough – for our purpose – to consider the problem of evaluation of a given mapping at a given point. When we want to find the solution of a (nonlinear) equation the solution can be found by evaluating the inverse of the mapping which governs the equation. From the numerical point of view it is enough to study the problem in the finite-dimensional context only. Therefore we consider:

Evaluation problem $E(f, \bar{x})$

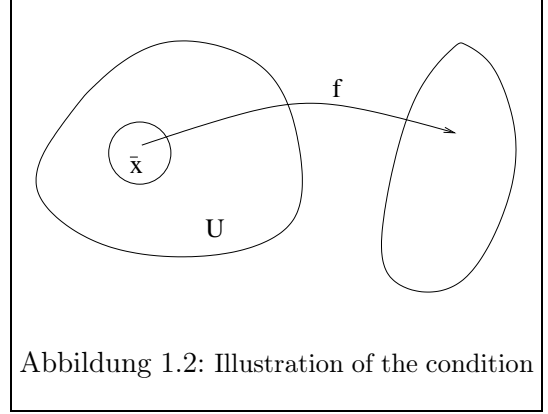
Given a mapping $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and \bar{x} in U .

Find the value $f(\bar{x})$.

Here the domain of definition U of f is an open subset⁷ of \mathbb{R}^n .

⁷A subset U of \mathbb{R}^n is called open when for each $x \in U$ there exists a ball $B_r(x)$ with $r > 0$ and $B_r(x) \subset U$. Here $B_r(x)$ is a ball with respect to any norm in \mathbb{R}^n ; see appendix ??.

Since the **input** \bar{x} may be available not as an exact quantity we have to study the evaluation of f in a neighborhood of \bar{x} . It is the goal to find out the size of the image of such a neighborhood; see Figure (1.2). Continuity of the mapping f in \bar{x} implies that the mapping f near \bar{x} is well-behaved. But this statement is of qualitative nature only and is not sufficient for numerical considerations. We need a quantitative formulation! This leads to Lipschitz-continuity. We study the related questions in more detail in the discussion of the solution of linear equations.



Well-posedness of $E(f, \bar{x})$ may be considered also quantitatively. We define

$$\kappa_{\text{abs}, U} f(\bar{x}) := \sup_{x \in U} \frac{|f(x) - f(\bar{x})|}{|x - \bar{x}|} \quad (1.2)$$

$\kappa_{\text{abs}, U} f(\bar{x})$ is called the *absolute condition number* of f . When $\bar{x} \neq \theta$ and $f(\bar{x}) \neq \theta$ then we call

$$\kappa_{\text{rel}, U} f(\bar{x}) := \sup_{x \in U} \frac{|f(x) - f(\bar{x})|}{|f(\bar{x})|} \bigg/ \frac{|x - \bar{x}|}{|\bar{x}|} \quad (1.3)$$

the *relative condition number* of f .

Remark 1.3 Suppose we want to solve a linear equation in the infinite-dimensional case governed by the operator A :

$$Ax = y$$

Then the evaluation problem for the inverse A^{-1} is unstable when the operator A^{-1} is unbounded. As a rule, this is the case when the problem to solve the equation is ill-posed. \square

Under the assumption that f in the evaluation problem is a differentiable function defined on \mathbb{R} – in the general case we would have to consider partial differentiability – we obtain an **infinitesimal relative condition number**:

$$\text{cond}_{\text{rel}} f(\bar{x}) := \frac{|\bar{x}| |f'(\bar{x})|}{|f(\bar{x})|} \quad (1.4)$$

This number could also be called the *elasticity* of the function value with respect to its argument. From this number the problems of evaluating the function f in \bar{x} can be read off.

Sensitivity analysis, which characterizes the change in model output due to variations in model input parameters, is of critical importance in simulation models; see Section 2.1.4. Sensitivity coefficients, defined as the partial derivatives of the model output with respect to the input parameters, are useful in assessing the reliability of the output from a complex model with many uncertainty parameters. Thus, we can state that condition numbers, as defined above, characterize the sensitivity of a mathematical problem to data perturbations.

1.3.3 Types of errors

Consider again the problem $E(f, \bar{x})$ of evaluating the mapping f in (a neighborhood of) \bar{x} . From the numerical point of view we have to take into account the presence of various sources of errors:

Model-error f may be considered as a model for a „physical process“: instead of f an approximation \tilde{f} is available only.

Errors in the data \bar{x} is known approximately only (due to measurement errors): instead of \bar{x} an approximation \tilde{x} is available only.

Truncation errors The value $\tilde{f}(\tilde{x})$ can be computed not exactly, arising from the fact that an infinite number of operations has to be replaced by a finite number of operations.

Rounding errors We have to expect errors due to the fact that in the computation process a finite number of real numbers is available only.

Example 1.4 Consider the evaluation of the function

$$f(t) := ce^{at} + de^{bt}, \quad t \in \mathbb{R} \quad (1.5)$$

in \bar{t} . The function f may be considered as a model for a chemical reaction: a, b are certain reaction constants, c, d describe the mixing relation.

A source of error in the evaluation of f in a point \bar{t} may be:

- a, b, c, d are known approximately: model error.
- \bar{t} is known approximately: error in the data.
- The exponential function can be computed via the potential series by truncating the infinite summation process only: truncation error.
- The evaluation of an exponential yields an irrational number in general; rounding is necessary.

□

Example 1.5 As we shall see, the price $C = C(S, t)$ at time t with value S of the underlying for an European call option in the Black-Scholes model is given by

$$C(S, t) = S\mathcal{N}(d_+(\sigma)) - Ke^{-r(T-t)}\mathcal{N}(d_-(\sigma)) \quad (1.6)$$

with

$$d_{\pm}(\sigma) = \frac{\ln(\frac{S}{K}) + (r \pm \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}, \quad \sigma \geq 0,$$

where \mathcal{N} denotes the density of the standard normal distribution:

$$\mathcal{N}(a) := \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} \exp(-\frac{s^2}{2}) ds, \quad a \in \mathbb{R}.$$

Here r, σ, K, T are the usual constants.

In the evaluation of this price formula various sources of errors should be regarded. Above all, the evaluation of the distribution function \mathcal{N} is a source of errors due to the fact that it contains the exponential function which is difficult to evaluate in a stable way, especially for large and small arguments. □

1.3.4 Over- and underdetermined linear systems

One of the most basic and important problems in science and engineering is the accurate and efficient simultaneous solution of a system of linear equations. Moreover, most numerical techniques in optimization and differential equations involve repeated linear equation solving.

A system of m linear equations in n unknowns consists of a set of algebraic relations of the form

$$\sum_{j=1}^n a_{ij}x_j = y_i, \quad i = 1, \dots, m \quad (1.7)$$

where x_j are the unknowns, a_{ij} are the coefficients of the system and y_i are the components of the right hand side. System (1.7) can be written more conveniently in matrix form as

$$Ax = y \quad (1.8)$$

where we have denoted $A = (a_{ij}) \in \mathbb{K}^{m \times n}$ the coefficient matrix, by $y = (y_i) \in \mathbb{K}^m$ the right hand side vector and by $x = (x_i) \in \mathbb{K}^n$ the unknown vector, respectively; here \mathbb{K} is the field of real or complex numbers. We call a solution of (1.8) any n -tuple of values x_j which satisfies (1.7). We call a system (1.8) **consistent** if there exists a solution otherwise **inconsistent**. The case that (1.8) is not consistent is not an unimportant situation. This situation occurs in virtually every branch of science and engineering.

We say that (1.8) is **overdetermined**⁸ if we have more equations than unknowns: $m > n$. Normally we do not expect that an overdetermined system has a solution. Clearly, we may decide by Gaussian elimination applied to the rectangular system whether a solution exists from the analytical point of view. But from the numerical point of view it may be difficult to decide this since rounding errors may destroy the solvability and vice versa.

We can reformulate the problem of solving the system (1.8) in another way to ask for the vector \bar{x} such that the residual $r := A\bar{x} - y$ is minimized in a given vector norm. In the case where we use the euclidean norm, $|\cdot|_2$, this becomes precisely the important statistical problem of finding the best *least squares solution* of (1.8). Under the assumption that the rank of A is n then this least squares solution x^\dagger can be found, using the transpose matrix A^t of A , in the following way:

$$x^\dagger = (A^t A)^{-1} A^t y \quad (1.9)$$

Here the matrix $(A^t A)^{-1} A^t$ is the so called **Moore-Penrose inverse** or **pseudoinverse**. Notice that this pseudoinverse exists since $A^t A$ is a $n \times n$ -matrix with rank n . As we know from numerical analysis, this solution x^\dagger can be computed by the QR-factorization.

When we restrict our consideration to the case of real-valued square systems of order n , that is to systems of the form (1.8) with $A \in \mathbb{K}^{n \times n}$ and $y \in \mathbb{K}^n$, then as we know from linear algebra existence, uniqueness and stability of the solution of (1.8) are ensured if one of the following (equivalent) hypotheses holds:

1. $\det(A) \neq 0$;
2. A is invertible, i.e. there exists a matrix B with $AB = E$;
3. $\text{range}(A) = \mathbb{R}^n$, i.e. $Ax = y$ has a solution for each $y \in \mathbb{R}^n$;
4. the columns of A are linearly independent;
5. $\text{rank}(A) = n$, i.e. the rows are linearly independent;

⁸„Die erfolgreichsten Theorien der Physik sind massiv von Daten überbestimmt“. C. Smeenk, University of Ontario, zur Diskussion der Modelle zur Entstehung des Universums

6. $\ker(A) = \{\theta\}$, i.e. $Ax = \theta$ implies $x = \theta$;
7. all eigenvalues of A are nonzero.

We say that (1.8) is **underdetermined** if we have less equations than unknowns: $m < n$. An underdetermined system of linear equations has infinitely many solutions (if there are any) and they form an affine space. All methods for solving underdetermined systems introduce some additional, artificial constraints. The advantage of the **maximum entropy method** is that it uses the most natural additional constraint. For most large underdetermined systems of linear equations the **minimal l_1 -norm solution** is also the **sparsest solution**. We shall come back to this property which is important in solving problems in image processing.

Suppose we have to solve the system (1.8) in the situation where A is invertible; see above. The solution of such a system may be stated as an evaluation problem as considered above.

Given a regular matrix $A \in \mathbb{K}^{n,n}$.
 Consider the mapping $f_A : \mathbb{K}^m \ni y \mapsto A^{-1}y \in \mathbb{K}^n$.
 Given $\bar{y} \in \mathbb{K}^n$, find $\bar{x} := f_A(\bar{y}) = A^{-1}\bar{y}$.

But in practical situations \bar{y} is not known exactly and instead of \bar{y} an approximation $y \in \mathbb{K}^n$ is available only. Then the question arises whether y is contained in $\text{ran}(A)$.

1.4 Tutorial example: numerical differentiation

Let us present a first example of an ill-posed problem: (numerical) differentiation of a function. The analysis of this problem can be done by using elementary considerations only. Here we consider just the elementary questions. In Chapter 2 we shall present the numerical differentiation based on Bernstein polynomials.

1.4.1 Numerical differentiation as an inverse problem

The differentiation of a (measured) function is involved in many inverse problems. In a mechanical system one may ask for hidden forces and since Newton's law relates forces to velocities and accelerations one has to differentiate observed data when we want to compute the forces. Moreover, one can see that the numerical differentiation is implicitly present in the problem of X-ray tomography, in parameter identification, in the determination of heat flow, and edge detection. We will see that this problem is important too in finding the volatility from market data. In the third chapter we present a method based on the approximation theorem of Weierstrass.

Let us consider the problem of finding the integral of a given function. This can be done analytically and numerically in a very stable way. When this problem is considered as a direct (forward) problem then to differentiate a given function is the related inverse problem. A mathematical description is given as follows:

Direct Problem: With a continuous function $x : [0, 1] \rightarrow \mathbb{R}$ compute

$$y(t) := \int_0^t x(s) ds, \quad t \in [0, 1].$$

Inverse Problem: Given a differentiable function $y : [0, 1] \rightarrow \mathbb{R}$ determine $x := y'$.

We are interested in the inverse problem. Since \mathbf{y} should be considered as the result of measurements the data \mathbf{y} are noisy and we may not expect that the noisy data $\tilde{\mathbf{y}}$ is presented by a continuously differentiable function. Therefore, the inverse problem has no obvious solution, especially when $\tilde{\mathbf{y}}(0) \neq 0$. (In practice we simply subtract $\tilde{\mathbf{y}}(0)$ from $\tilde{\mathbf{y}}$.)

Moreover, the problem should not be formulated in the space of continuous functions since perturbations due to noise lead to functions which are not continuous. But the analysis and the message of the results is not very different from the following in the more elementary case of continuous perturbations.

Suppose that instead of the continuous function $\mathbf{y} : [0, 1] \rightarrow \mathbb{R}$ a „measured“ function $\mathbf{y}^\varepsilon : [0, 1] \rightarrow \mathbb{R}$ is available only. We assume:

$$|\mathbf{y}^\varepsilon(\mathbf{t}) - \mathbf{y}(\mathbf{t})| \leq \varepsilon \text{ for all } \mathbf{t} \in [0, 1].$$

Clearly, ε is the level of noise in the measurement of \mathbf{y}^ε . It is reasonable to try to reconstruct the derivative $\mathbf{x} := \mathbf{y}'$ of \mathbf{y} at $\tau \in (0, 1)$ by

$$\mathbf{x}^{\varepsilon, \mathbf{h}}(\tau) := D_{\mathbf{h}} \mathbf{y}^\varepsilon(\tau) := \frac{\mathbf{y}^\varepsilon(\tau + \mathbf{h}) - \mathbf{y}^\varepsilon(\tau)}{\mathbf{h}}$$

where $\mathbf{h} \neq 0$ has to be chosen such that $\tau + \mathbf{h} \in [0, 1]$. We obtain

$$|\mathbf{x}^{\varepsilon, \mathbf{h}}(\tau) - \mathbf{x}(\tau)| \leq \left| \frac{\mathbf{y}(\tau + \mathbf{h}) - \mathbf{y}(\tau)}{\mathbf{h}} - \mathbf{x}(\tau) \right| + \left| \frac{(\mathbf{y}^\varepsilon - \mathbf{y})(\tau + \mathbf{h}) - (\mathbf{y}^\varepsilon - \mathbf{y})(\tau)}{\mathbf{h}} \right|.$$

Under the assumption that \mathbf{x} is continuously differentiable we have

$$\frac{\mathbf{y}(\tau + \mathbf{h}) - \mathbf{y}(\tau)}{\mathbf{h}} - \mathbf{x}(\tau) = \frac{1}{2} \mathbf{y}''(\eta) \mathbf{h} \text{ for some } \eta \in [0, 1].$$

Moreover, when we know a bound

$$|\mathbf{x}'(\mathbf{t})| \leq E \text{ for all } \mathbf{t} \in [0, 1], \quad (1.10)$$

then the estimate

$$|\mathbf{x}^{\varepsilon, \mathbf{h}}(\tau) - \mathbf{x}(\tau)| \leq \frac{1}{2} \mathbf{h} E + 2 \frac{\varepsilon}{\mathbf{h}} \quad (1.11)$$

follows. Now it is clear that the best what we can do is to balance the terms on the right hand side in (1.11). This is done by the choice

$$\mathbf{h} := \mathbf{h}_{\text{opt}} := 2 \sqrt{\frac{\varepsilon}{E}}. \quad (1.12)$$

This gives

$$|\mathbf{x}^{\varepsilon, \mathbf{h}(\varepsilon)}(\tau) - \mathbf{x}(\tau)| \leq 2 \sqrt{E \varepsilon}. \quad (1.13)$$

The Diagram 1.3 which is a graphical presentation of the bound (1.11) is typical for approximations in ill-posed problems: there are two terms in the error estimate, a term due to approximation of the inverse mapping and a term due to measurement error. The balance of these two terms gives an „optimal“ reconstruction result. Thus, in contrast to well-posed problems, it is not the best to discretize finer and finer. One may consider ill-posed problems under the motto „*When the imprecise is preciser*“.⁹ Moreover, the bound in (1.13) shows that halving the (measurement) error ε does not lead to a halving of the absolute error in the result (as it is usually the case in direct problems).

⁹This is the title of [11].

The requirement of the bound in (1.10) describes an information concerning the solution we want to find. Therefore this information is called **a priori information/a priori knowledge**. The quantity

$$\text{SNR} := \frac{E}{\varepsilon}$$

may be considered as the signal to noise ratio. The merit function we have used is the norm of the deviation of the approximation from the (unknown) solution. The regularization strategy to balance the merit function with respect to the a priori information (E) and the signal to noise ratio SNR is called an **a priori strategy** since it can be done without calculating some approximations.

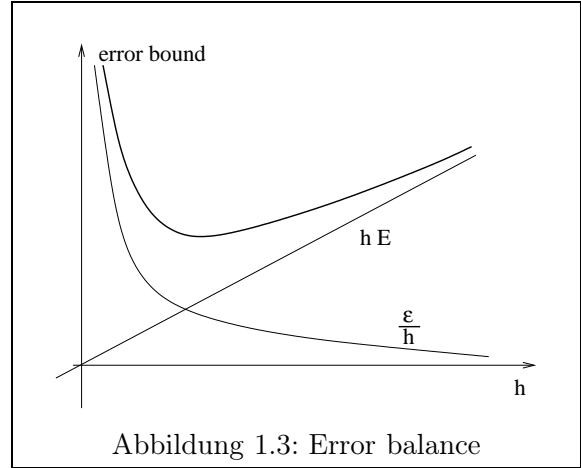


Abbildung 1.3: Error balance

Remark 1.6 *The equation*

$$\int_0^s x(t) dt = y(s), \quad s \in [0, 1]$$

is the simplest example of a Volterra equation of the first kind which is an equation of the form

$$\int_a^b \kappa(t, s) x(t) dt = y(s), \quad s \in [c, d],$$

where κ is a kernel function from $[a, b] \times [c, d]$ into \mathbb{R} . □

1.4.2 Numerical differentiation realized with floating point numbers

We consider the problem of the numerical differentiation realized on a computer with floating point numbers. Floating numbers are elements x of

$$\mathbb{F} := \mathbb{F}(g, t, e_{\min}, e_{\max}) := \left\{ \sigma [d_1 \dots d_t] g^{e-t} \mid \begin{array}{l} d_1, \dots, d_t \in \{0, 1, \dots, g-1\}, d_1 \neq 0, \\ e \in \mathbb{Z}, e_{\min} \leq e \leq e_{\max}, \sigma \in \{+, -\} \end{array} \right\} \cup \{0\}.$$

These numbers are called normalized since $d_1 \neq 0$ is required. The parameters of this set are:

$g \in \mathbb{N}, g \geq 2$	base of the numbers
$t \in \mathbb{N}$	number of significant digits $d_i \in \{0, 1, \dots, g-1\}$ (precision)
$m := d_1 \dots d_t \in \mathbb{N}$	mantissa of x
$e \in \mathbb{Z}$	exponent of x
$[e_{\min}, e_{\max}]$	range of exponents

The consequences of the use of these numbers in the differentiation is the subject of this subsection. We restrict us to the case $g = 10$.

Consider the approximation of the derivative $f'(x)$ by difference quotients

$$f'(x) = \frac{f(x+h) - f(x)}{h} + O(h);$$

the term $O(h)$ is from of the remainder of the approximation. If the step with h is given as $h = 10^{-s}$ then the difference quotient is an approximation of $f'(x)$ with approximately s significant

digits. In the evaluation of the nominator in the difference quotient cancellation may occur. This can lead to a loss of s digits. Therefore the precision of the result has

$$\min(s, t - s) , \text{ i.e. } t/2$$

significant digits. If we choose the approximation method

$$f'(x) = \frac{f(x + h) - f(x - h)}{2h} + O(h^2)$$

with step width $h = 10^{-s}$ then the precision is

$$\min(2s, t - s) , \text{ i.e. } 2t/3 .$$

A further improvement results from

$$f'(x) = \frac{4}{3} \frac{f(x + h) - f(x - h)}{2h} - \frac{1}{3} \frac{f(x + 2h) - f(x - 2h)}{2h} + O(h^4)$$

for a step width $h = 10^{-s}$ since the precision is now

$$\min(4s, t - s) , \text{ i.e. } 4t/5 .$$

The observation is that going ahead in the same way the gain-rate of significant digits is poor.

Each of the formulae above may be used to approximate the derivative. An improvement of the accuracy can be achieved by halving the stepsize h . Notice that in halving the stepsize the cost is different: one new evaluation for the first formula, 2 additional evaluation for the second and 4 evaluations for the third.

1.5 Numerical algorithms

Algorithms are the main tools to formulate methods to solve problems. Here we formulate the main goals in the analysis of algorithms. In Chapter 3 we shall consider the questions again under more realistic assumptions.

1.5.1 Algorithms

The solution of a problem by numerical methods is usually formulated by an ordered list of steps where each of which uses data of the problem and results from an earlier step; the evaluation problem may be considered as such a problem when we are forced to decompose the mapping f as follows: $f = f_l \circ \dots \circ f_1$. Such a list of instructions (of length l) is called an **algorithm** when certain requirements are fulfilled:

An algorithm is a finite set of well-defined feasible ordered instructions for accomplishing some task which, given an initial state, will terminate after a finite number of operations in a corresponding recognizable end-state.

Different algorithms may complete the same task with a different set of instructions in more or less time, space or effort than others. The analysis of an algorithm is concerned with the following characteristics:

Accuracy: Accuracy refers to the absolute or relative error of an approximate quantity. The numerical approximation should be as accurate as possible. This requires the algorithm to be numerically stable.

Robustness/Reliability: The algorithm should solve many problems well and it should warn the user, if the result is inaccurate. Hence it should be possible to estimate the error.

Efficiency/Complexity: The amount of operations and the size of the memory required should be as small as possible.

Precision: Precision is the accuracy with which the basic arithmetic operations are performed. In all our analyses there is an implicit assumption that the given arithmetic is of arbitrarily high precision just in the more principal consideration of numerical algorithms precision may play a role.

To discuss whether a numerical solution approximates the solution of a problem, we need a measure of the distance between a numerical solution and the exact solution to the problem and a method parameter which we can use to vary the numerical method.

A numerical method is **convergent** if the distance between the numerical solution and the exact solution goes to zero as the method parameter approaches some limit. In order to obtain this property, the numerical method must be **stable**. This means that the error growth factor is bounded independent of the method parameter. Stability is purely a property of the numerical method and is independent of the problem. To have any hope of establishing convergence, it is necessary to establish some kind of connection between the problem and the numerical method. This connection is **consistency**. In the following we will discuss the properties *convergence*, *stability*, *consistency* in mathematical terms.

1.5.2 Analysis of numerical algorithms

Let X be a space endowed with the metric d . Let x^\dagger in X be the object which we want to compute. We assume that we have computed by an algorithm a sequence $(x_n)_{n \in \mathbb{N}}$ in X which should be an approximating sequence of x^\dagger .

Definition 1.1 *The pair $(x^\dagger, (x_n)_{n \in \mathbb{N}})$ is called **consistent** iff*

$$\lim_n x_n = x^\dagger.$$

□

In practice, an algorithm which is designed to construct the approximating sequence $(x_n)_{n \in \mathbb{N}}$ is never exact due to

- **rounding errors** during the calculations,
- **truncation errors** due to approximating certain variables and results,
- **data errors** which come in when certain data of the problem to solve are available by measurements only.

Therefore, instead of $(x_n)_{n \in \mathbb{N}}$ we obtain a sequence $(x_{n,\tau})_{n \in \mathbb{N}}$ where $\tau \in [0, \infty)$ is the quantity which collects the errors in the computation. Clearly, we don't have $\lim_n x_{n,\tau} = x^\dagger$ when $\tau > 0$.

In Chapter 4 we will discuss methods which are appropriate to extract from the sequence $(x_{n,\tau})_{n \in \mathbb{N}}$ good approximations for x^\dagger . The main idea is to search for a stopping-index $n(\tau)$ such that $x_{n(\tau),\tau}$ becomes such a good approximation.

1.5.3 Numerical algorithms: inverse crime

Often, tests of the exact or approximate theoretical models employed in inversion schemes are made with synthetic data. Generating the latter also requires a theoretical model, which, mathematically speaking, can be identical to, or different from, the one employed in the inversion scheme. In [3] the authors coin the expression *inverse crime* to denote the act of employing the same model to generate, as well as to invert, synthetic data. Moreover, they warn against committing the inverse crime, *in order to avoid trivial inversion* and go on to state: *it is crucial that the synthetic data be obtained by a forward solver which has no connection to the inverse solver*. These assertions raise the following questions:

- What does the term „no connection“ mean?
- Provided a definition can be given, what kind of reconstructions of the unknown parameters can one obtain when there is „no connection“ between the forward and inverse solvers?
- Should the inverse crime always be avoided?
- Are inverse crime inversions always trivial?

Shortly:

The term „**inverse crime**“ is saying that sampling the data in the forward problem and solving the inverse problem are done in the same manner. Inverse crime arises when

- the numerically produced simulated data is produced by the same model that is used to invert the data;
- the discretization in the numerical simulation is the same as the one used in the inversion.

When inverse crimes are not avoided a numerical algorithm may lead to unrealistically optimistic results. Note that inverse crimes are not possible in situations where actual real-world measured data are used, they are only a problem of computational simulation studies.

1.6 Appendix: Several facts in Hilbert and Banach spaces

Here we collect some functional analytic results which are useful for the analysis of inverse problems.

Let X be a \mathbb{K} -vector space where $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ is the scalar field. In the following, we denote the null vector by θ .

Definition 1.2 A mapping $\|\cdot\| : X \longrightarrow \mathbb{R}$ is called a **norm** if the following properties are satisfied:

$$(1) \quad \|x\| \geq 0 \text{ for all } x \in X \text{ and } \|x\| = 0 \text{ iff } x = \theta.$$

$$(2) \quad \|ax\| = |a|\|x\| \text{ for all } x \in X, a \in \mathbb{K}.$$

$$(3) \quad \|x + y\| \leq \|x\| + \|y\| \text{ for all } x, y \in X.$$

If $\|\cdot\|$ is a norm then the pair $(X, \|\cdot\|)$ is called a **normed space**.

□

If $(X, \|\cdot\|)$ is a normed space then (X, d) is a metric space where d is defined by

$$d : X \times X \ni (x, y) \longmapsto \|x - y\| \in \mathbb{R}.$$

Definition 1.3 Let $(X, \|\cdot\|)$ be a normed space. A sequence $(x_n)_{n \in \mathbb{N}}$ in X is **convergent** with limit $x \in X$ iff

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n \geq N (\|x_n - x\| < \varepsilon).$$

We use the notation $\lim_n x_n := x$. □

Let $(X, \|\cdot\|)$ be a normed space and let A be a subset of X . Then A is called closed iff for each convergent sequence $(x_n)_{n \in \mathbb{N}}$ with $x_n \in A$ for all $n \in \mathbb{N}$ we have $\lim_n x_n \in A$.

Definition 1.4 Let $(X, \|\cdot\|)$ be a normed space.

(a) A sequence $(x_n)_{n \in \mathbb{N}}$ is called a **Cauchy sequence** in X iff

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n, m \geq N (\|x_n - x_m\| < \varepsilon).$$

(b) $(X, \|\cdot\|)$ is called **complete** or a **Banach space** if each Cauchy sequence in X is convergent. □

Example 1.7 We define

$$\begin{aligned} l_\infty(\mathbb{K}) &:= \{x = (x_k)_{k \in \mathbb{N}} | x_k \in \mathbb{K} \text{ for all } k \in \mathbb{N}, \sup_{k \in \mathbb{N}} |x_k| < \infty\}. \\ c(\mathbb{K}) &:= \{x = (x_k)_{k \in \mathbb{N}} | x_k \in \mathbb{K} \text{ for all } k \in \mathbb{N}, (x_k)_{k \in \mathbb{N}} \text{ convergent}\}. \\ c_0(\mathbb{K}) &:= \{x = (x_k)_{k \in \mathbb{N}} | x_k \in \mathbb{K} \text{ for all } k \in \mathbb{N}, (x_k)_{k \in \mathbb{N}} \text{ is a null sequence}\}. \\ c_c(\mathbb{K}) &:= \{x = (x_k)_{k \in \mathbb{N}} | x_k \in \mathbb{K} \text{ for all } k \in \mathbb{N}, x_k = 0 \text{ for almost all } k \in \mathbb{N}\}. \end{aligned}$$

Obviously,

$$c_c(\mathbb{K}) \subset c_0(\mathbb{K}) \subset c(\mathbb{K}) \subset l_\infty(\mathbb{K}).$$

Additionally, we set

$$\|x\|_\infty := \sup_{k \in \mathbb{N}} |x_k|, \quad x = (x_k)_{k \in \mathbb{N}} \in l_\infty.$$

One can show that $(l_\infty(\mathbb{K}), \|\cdot\|_\infty)$ is complete. $c(\mathbb{K})$ is a closed subspace of $l_\infty(\mathbb{K})$ and $c_0(\mathbb{K})$ is a closed subspace of $c(\mathbb{K})$. This implies that $(c(\mathbb{K}), \|\cdot\|_\infty)$, $(c_0(\mathbb{K}), \|\cdot\|_\infty)$ are Banach spaces. $c_c(\mathbb{K})$ is not a closed subspace of $l_\infty(\mathbb{K})$; we have that $c_0(\mathbb{K})$ is the closure of $c_c(\mathbb{K})$. □

Example 1.8 We set for $1 \leq p < \infty$

$$l_p(\mathbb{K}) := \{x = (x_k)_{k \in \mathbb{N}} | x_k \in \mathbb{K} \text{ for all } k \in \mathbb{N}, \sum_{k \in \mathbb{N}} |x_k|^p < \infty\}$$

and

$$\|x\|_p := \left(\sum_{k \in \mathbb{N}} |x_k|^p \right)^{\frac{1}{p}}, \quad x = (x_k)_{k \in \mathbb{N}} \in l_p(\mathbb{K}).$$

Then $(l_p(\mathbb{K}), \|\cdot\|_p)$ is a Banach space. □

Definition 1.5 Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ be normed spaces. A mapping $f : D \rightarrow Y, D \subset X$, is called **continuous** in $x^0 \in D$ iff

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x \in D (\|x - x^0\|_X < \delta \implies \|f(x) - f(x^0)\|_Y < \varepsilon).$$

The mapping f is called **continuous** iff f is continuous in each $x^0 \in D$. □

Example 1.9 Let $(X, \|\cdot\|)$ be a normed space. Then the norm considered as a mapping from X into \mathbb{K} is continuous. This follows from

$$|\|\mathbf{x}\| - \|\mathbf{y}\|| \leq \|\mathbf{x} - \mathbf{y}\|.$$

Moreover, the mappings

$$X \ni \mathbf{x} \mapsto \mathbf{x} + \mathbf{w} \in X, \mathbb{R} \ni \mathbf{a} \mapsto \mathbf{a}\mathbf{w} \in X, X \ni \mathbf{x} \mapsto r\mathbf{x} \in X,$$

where $\mathbf{w} \in X$ and $r \in \mathbb{R}$ are fixed elements are continuous too. This follows from

$$\|(\mathbf{x} + \mathbf{w}) - (\mathbf{x}^0 + \mathbf{w})\| = \|\mathbf{x} - \mathbf{x}^0\|, \|\mathbf{a}\mathbf{w} - \mathbf{a}^0\mathbf{w}\| = |\mathbf{a} - \mathbf{a}^0|\|\mathbf{w}\|, \|r\mathbf{x} - r\mathbf{x}^0\| = |r|\|\mathbf{x} - \mathbf{x}^0\|.$$

□

The most interesting mappings in linear algebra are the linear mappings. Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ be normed spaces. As we know, a mapping $L : X \rightarrow Y$ is linear iff

$$L(\mathbf{a}\mathbf{u} + \mathbf{b}\mathbf{v}) = \mathbf{a}L(\mathbf{u}) + \mathbf{b}L(\mathbf{v}) \text{ for all } \mathbf{a}, \mathbf{b} \in \mathbb{K}, \mathbf{u}, \mathbf{v} \in X.$$

The following theorem is easy to prove.

Theorem 1.6 Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ be normed spaces and let $L : X \rightarrow Y$ be a linear mapping. Then the following conditions are equivalent:

- (a) L is bounded, i.e. there exists $c > 0$ with $\|L(\mathbf{x})\|_Y \leq c\|\mathbf{x}\|_X$ for all $\mathbf{x} \in X$.
- (b) L is continuous in each $\mathbf{x}^0 \in X$.
- (c) L is continuous in $\mathbf{x}^0 := \theta$.

Definition 1.7 Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ be normed spaces and let $L : X \rightarrow Y$ be continuous. Then the number $\|L\| := \sup\{\|L(\mathbf{x})\|_Y \mid \|\mathbf{x}\|_X \leq 1\}$ is called the **operator norm**. □

Important is the estimate

$$\|L(\mathbf{x})\|_Y \leq \|L\| \|\mathbf{x}\|_X, \mathbf{x} \in X, \tag{1.14}$$

for all $\mathbf{x} \in X$. It is easy to prove

Corollary 1.8 Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y), (Z, \|\cdot\|_Z)$ be normed spaces and let $L : X \rightarrow Y, K : Y \rightarrow Z$ be continuous. Then

$$\|K \circ L\| \leq \|K\| \|L\|.$$

Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y), (Z, \|\cdot\|_Z)$ be normed spaces. We set

$$L(X, Y) := \{T : X \rightarrow Y \mid T \text{ linear and continuous}\}.$$

The space $L(X, Y)$ endowed with the operator norm is a Banach space.

Definition 1.9 Let X be a vector space with scalar space $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. A mapping $\sigma : X \times X \rightarrow \mathbb{K}$ is called an **inner product** on X if the following conditions are satisfied.¹⁰

¹⁰Here, as usual, \bar{z} denotes the complex conjugate of z

- (a) $\sigma(x, x) \in \mathbb{R}, \sigma(x, x) > 0$ for all $x \in X \setminus \{0\}$;
- (b) $\sigma(x, y) = \overline{\sigma(y, x)}$ for all $x, y \in X$;
- (c) $\sigma(ax + by, z) = a\sigma(x, z) + b\sigma(y, z)$ for all $x, y, z \in X, a, b \in \mathbb{R}$.

A vector space X endowed with the inner product σ is called a **pre-Hilbert space**. \square

Let (X, σ) be a pre-Hilbert space. Then we have for all $x, y \in X$

$$|\sigma(x, y)| \leq \sqrt{\sigma(x, x)} \sqrt{\sigma(y, y)},$$

with equality when x, y are linear dependent. This inequality is called the **Cauchy–Schwarz inequality**. A consequence of this inequality is the fact that the pre-Hilbert space (X, σ) becomes a normed space since $\|\cdot\|_\sigma$, defined by

$$\|x\|_\sigma := \sqrt{\sigma(x, x)}, \quad x \in X,$$

is a norm in X .

Definition 1.10 The pre-Hilbert space (X, σ) is called a **Hilbert space** if the normed space $(X, \|\cdot\|_\sigma)$ is complete. \square

Let (X, σ) be a pre-Hilbert space. In the following we want to use the notation which is common in physics:

$$\langle \cdot, \cdot \rangle := \sigma(\cdot, \cdot).$$

Cauchy-Schwarz' inequality becomes now

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

Here are three further identities in a pre-Hilbert space $(X, \langle \cdot, \cdot \rangle)$.

Parallelogram identity $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2, \quad x, y \in X.$

Polarization identity¹¹

$$\langle x, y \rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2) + \frac{i}{4}(\|x+iy\|^2 - \|x-iy\|^2), \quad x, y \in X.$$

Theorem of Pythagoras $\|x+y\|^2 = \|x\|^2 + \|y\|^2$ for $x, y \in X$ with $\langle x, y \rangle = 0$.

A subset K in a vector space X is **convex** iff

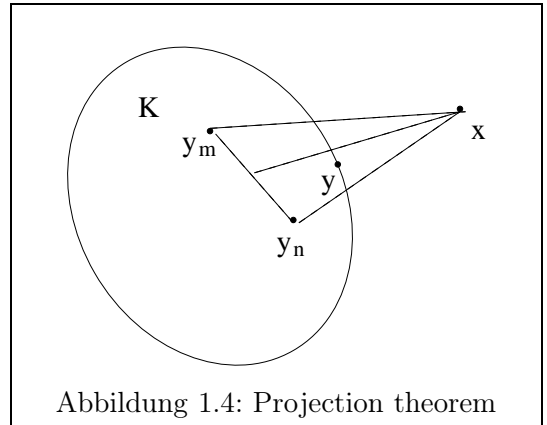
$$ax + (1-a)y \in K \text{ for all } x, y \in K, a \in [0, 1].$$

Theorem 1.11 (Variational inequality) Let X be a pre-Hilbert space and let $K \subset X$ be non empty and convex. Then for all $y \in K$ the following conditions are equivalent:

(a)

$$\|x - y\| = \inf_{z \in K} \|x - z\|. \quad (1.15)$$

¹¹This is the case of complex scalars.



(b)

$$\Re\langle x - y, z - y \rangle \leq 0 \text{ for all } z \in K. \quad (1.16)$$

Theorem 1.12 (Projection theorem) *Let X be a Hilbert space and let $K \subset X$ be non empty and convex. Then there exists for each $x \in X$ a uniquely determined $y \in K$ with*

$$\|x - y\| = \inf_{z \in K} \|x - z\|. \quad (1.17)$$

The vector y in Theorem 1.12 is called the projection of x with respect to K . We denote this y as $P_K x$. The result of the theorem above is the fact that the mapping

$$X \ni x \longmapsto P_K x \in K$$

is well defined. This mapping is called the **projection** of X onto K .

Definition 1.13 *Let X be a pre-Hilbert space.*

(a) *A set $O \subset X$ is called an **orthonormal system** (in X) iff*

$$\|e\| = 1 \text{ for all } e \in O, \langle e, f \rangle = 0 \text{ for all } e, f \in O, e \neq f.$$

(b) *An orthonormal system is called an **orthonormal basis** iff there exists no orthonormal system O' with $O \subset O', O' \neq O$.*

□

Clearly, every orthonormal system is a linearly independent set. Due to Zorn's lemma, each pre-Hilbert space possesses an orthonormal basis. If the underlying pre-Hilbert space X is separable, i.e. when X is the closure of a denumerable subset A , then there exists an orthonormal basis in X which is denumerable. Such an orthonormal system can be constructed with the help of the Gram-Schmidt orthogonalization.

1.7 Bibliographical comments

Inverse problems are considered for instance in [2, 8, 12, 15, 21, 24, 25]. Early sources for the treatment of inverse and ill-posed problems are [5, 7, 14, 13, 16, 19, 22, 23].

A monography concerning the important subject *tomography* is [17]; see also [18]. The theme „inverse crime“ is discussed for example in [3, 9, 10, 26]. A comprehensive bibliography for numerical differentiation is presented in

`file:///C:/Users/gruppe/Desktop/NumericalDiffBib_lnk_3.html`

and

`http://sourcedb.cas.cn/sourcedb-igg-cas/cn/zjrck/200907/W020100801406250190094.pdf`

A useful source for the tools in functional analysis (normed spaces, continuous operators, function spaces, ...) is [1]. The results concerning Banach and Hilbert spaces in the appendix can be found in [1] too.

1.8 Exercises

1.1 Show that under the assumption

$$|x''(t)| \leq E \text{ for all } t \in [0, 1]$$

with

$$x^{\varepsilon, h}(\tau) := D_h y^\varepsilon(\tau) := \frac{y^\varepsilon(\tau + h) - y^\varepsilon(\tau)}{h}$$

combined with a step-width rule $h := h(\varepsilon)$ leads to the error estimate

$$|x^{\varepsilon, h(\varepsilon)}(\tau) - x(\tau)| \leq cE^{1/3}\varepsilon^{2/3}.$$

Here c is a constant independent of ε, E .

1.2 Suppose that $f : [0, 1] \rightarrow \mathbb{R}$ is three times continuous differentiable. Show for $\tau \in [0, 1]$

$$f'(\tau) = \frac{1}{2h}(-3f(\tau) + 4f(\tau + h) - f(\tau + 2h)) + O(h^2).$$

$O(\cdot)$ is the big-O Landau-Symbol.

1.3 Suppose that $f : [0, 1] \rightarrow \mathbb{R}$ is twice continuous differentiable. Show for $\tau \in (0, 1)$

$$\lim_{h \rightarrow 0} \left| f''(\tau) - \frac{f(\tau + h) - 2f(\tau) + f(\tau - h))}{h^2} \right| = 0.$$

1.4 Suppose that $f : [0, 1] \rightarrow \mathbb{R}$ is four times continuous differentiable. Show for $\tau \in (0, 1)$ and $h > 0$ such that $\tau + h, \tau - h \in [0, 1]$,

$$\left| f''(\tau) - \frac{f(\tau + h) - 2f(\tau) + f(\tau - h))}{h^2} \right| \leq \frac{h^2}{12} \max_{s \in [0, 1]} |f^{(4)}(s)|.$$

$O(\cdot)$ is the big-O Landau-Symbol.

1.5 A simple model for the growth of a population is given by

$$u' = qu$$

where u denotes the size of the population and where $q : [0, 1] \rightarrow \mathbb{R}$ is a „parameter“. Find a method to reconstruct the function $q : [0, 1] \rightarrow \mathbb{R}$ from observations of $u : [0, 1] \rightarrow \mathbb{R}$.

1.6 The distance s of a runner from a fixed mark is measured at certain times t :

t	0.0	0.5	1.0	1.5	2.0
s	0.00	3.65	6.80	9.90	12.15

- Use central differences in order to reconstruct the velocity of the runner at times $t = 0.5$ and $t = 1.25$.
- Describe a method in order to reconstruct the acceleration of the runner and apply this method for $t = 1.5$.

1.7 Derive a formula for approximating the second derivative based on five equidistant nodes symmetric to the evaluation point. What is the error of the approximation?

1.8 Let X be the Hilbert space ℓ_2 and define the linear operator $A : X \rightarrow X$ by the formula $A(x)_k = \omega_k x_k, k \in \mathbb{N}$, where $(\omega_k)_{k \in \mathbb{N}}$ is a given sequence of real numbers. Show: Then λ is an eigenvalue of A if and only if there is an index k such that $\lambda = \omega_k$.

1.9 Consider the Volterra integral equation

$$\lambda y(t) + \int_0^t \sin(t-s)y(s)ds = 1, \quad t \in \mathbb{R},$$

with $\lambda \neq 0$. (Then the integral equation is said to be of the 2nd kind and - as opposed to the case $\lambda = 0$ (integral equation of the 1st kind) a well-posed equation.)

- (a) Show that this equation can be transformed into an initial value problem of 2nd order.
- (b) In order to solve the integral equation, solve the initial value problem.

1.10 Let X be the Hilbert space $L_2([0, 1])$ of all equivalence classes of square-integrable functions on the interval $[0, 1]$. Define an operator A onto X by declaring $A(x)(t) := g(t)x(t)$, $t \in [0, 1]$, for some bounded and measurable function g .

Show: A possesses no eigenvalues if g is chosen appropriate.

1.11 Consider the problem of calculating the fractional derivative of a function f given in $L_2(\mathbb{R})$:

$$D^\alpha f(x) := \frac{1}{\Gamma(n+1-\alpha)} \frac{d^{n+1}}{dx^{n+1}} \int_{-\infty}^x \frac{f(t)}{(x-t)^{\alpha-n}} dt$$

for $n \in \mathbb{N}, n < \alpha < n+1$. Such problems are frequently encountered in many practical contexts. It is well known that if $0 < \alpha \leq 1$, then $D^\alpha f(x)$ is a formal solution of the Abel integral equation

$$I^\alpha u(x) = \frac{1}{\Gamma(\alpha)} \int_{-\infty}^x \frac{u(t)}{(x-t)^{1-\alpha}} dt = f(x), \quad -\infty < x < \infty.$$

Compute $D^\alpha f$ for

$$f(x) := e^{-x^2}, \quad f(x) := \begin{cases} 0 & , x \leq -1 \\ 1+x & , -1 < x \leq 0 \\ 1-x & , 0 < x \leq 1 \\ 0 & , 1 < x \end{cases}, \quad f(x) := \begin{cases} 0 & , x \leq -1 \\ 1 & , -1 < x \leq 1 \\ 0 & , 1 < x \end{cases}, \quad f(x) := 0.$$

1.12 *Can you hear the length of string?* Consider the boundary value problem

$$u'' = f, \quad u(0) = u(l) = 0$$

where $f: \mathbb{R} \rightarrow \mathbb{R}$ is a given continuous function. Suppose that the solution u and f are known. Find the length l of the interval.

1.13 Consider the boundary value problem

$$u'' + \sigma u = f, \quad u(0) = u(l) = 0$$

where $f: \mathbb{R} \rightarrow \mathbb{R}$ is a given continuous function. Find sufficient conditions on f such that the constant $\sigma > 0$ can be computed from an observation $u(\tau)$ for some point $\tau \in (0, l)$.

1.14 The computation of eigenvalues of a given matrix is a standard problem in (numerical) linear algebra.

- (a) If this problem is considered as a direct problem what should be the formulation of the inverse problem?

- (b) Let $\lambda_1 := 1 - i, \lambda_2 := 1 + i$. Compute a matrix $A \in \mathbb{R}^{2,2}$ with eigenvalues λ_1, λ_2 .
- (c) Is the matrix A in (b) uniquely determined? Give an argument for your answer.

1.15 Consider the real Hilbert space $(\ell_2, \|\cdot\|_2)$. Let $x = (x_k)_{k \in \mathbb{N}} \in \ell_2$ and let $x^\varepsilon = (x_k^\varepsilon)_{k \in \mathbb{N}}, \varepsilon \geq 0$, be a sequence of real numbers. Suppose

$$|x_k - x_k^\varepsilon| \leq \varepsilon, k \in \mathbb{N}, \sum_{k=1}^{\infty} \omega_k x_k^2 \leq E^2 < \infty$$

where $(\omega_k)_{k \in \mathbb{N}}$ is a sequence of reals with the properties $0 < \omega_k \leq \omega_{k+1}, k \in \mathbb{N}, \lim_k \omega_k = \infty$.

Define for $N \in \mathbb{N}, \varepsilon \geq 0$, the sequence $x^{N,\varepsilon} \in \ell_2$ by

$$x_k^{N,\varepsilon} := \begin{cases} x_k^\varepsilon & , \text{ if } k \leq N-1 \\ 0 & , \text{ if } k \geq N \end{cases}.$$

Show:

- (a) $\|x - x^{N,\varepsilon}\|_2^2 \leq f(N, \varepsilon) := N\varepsilon^2 + \omega_N^{-1}E^2$.
- (b) Minimize (balance) $f(N, \varepsilon)$ with respect to N for the choice $\omega_k := k^2, k \in \mathbb{N}$.
- (c) Minimize (balance) $f(N, \varepsilon)$ with respect to N for the choice $\omega_k := e^k, k \in \mathbb{N}$.
- (d) Compare and comment the results in (b),(c).

1.16 The problem of computing the eigenvalues of given matrix is well known. If this problem is considered as a direct problem what can be the formulation of the inverse problem?

1.17 Given is a well of depth H . How much time T passes until one hears the impact of a stone which falls freely into the shaft, starting with zero velocity at time 0? This is the direct problem. The associated inverse problem is: Is it feasible to determine the depth of the shaft from T ? Due to Galilei, the distance s the stone has fallen after a period t of time is given by

$$s = \frac{1}{2}gt^2.$$

Here g denotes the gravitational constant. Let c denote the speed of sound, which has to be taken into consideration when calculating T .

- (a) Determine T in dependence of H .
- (b) Determine H in dependence of T .
- (c) What is the precision in H under the assumption that the time T which has passed can be measured with a precision of ε ?

1.18 Let f be a real function defined over the interval $[0, \infty)$. The Laplace transform $\mathcal{L}(f)$ is defined as the integral

$$\mathcal{L}(f)(s) = \int_0^{\infty} e^{-st}f(t)dt$$

provided the integral exists.

Given the function $g : (0, \infty) \ni s \longmapsto \frac{1}{2s^2}(2 - 3e^{-s} + e^{-3s}) \in \mathbb{R}$. Find a function f with $\mathcal{L}(f) = g$.

1.19 Consider the integral equation

$$\int_0^1 \exp(st)x(t)dt = y(s), \quad 0 \leq s \leq 1 \quad (1.18)$$

which admits the solution

$$x(t) := \exp(t), \quad 0 \leq t \leq 1,$$

when the right-hand side is given by

$$y(s) := \frac{(\exp(s+1) - 1)}{(s+1)}, \quad 0 \leq s \leq 1.$$

If we approximate (1.18) by the trapezoidal rule, we get

$$\int_0^1 \exp(st)x(t)dt \approx h \left(\frac{1}{2}x(0) + \sum_{j=1}^{n-1} \exp(jhs)x(jh) + \frac{1}{2} \exp(s)x(1) \right), \quad 0 \leq s \leq 1,$$

where $h := 1/n$. With nodes $s = ih$ we get

$$h \left(\frac{1}{2}x_0 + \sum_{j=1}^{n-1} \exp(jhs)x_j + \frac{1}{2} \exp(s)x_n \right) = y_i := y(ih), \quad i = 0, \dots, n, \quad (1.19)$$

where the numbers x_j are to be considered as approximations of $x(jh)$.

Tackle the following problems for the cases $n = 4, 8, 16$ with the help of MATLAB.

- Find the matrix A which governs the system of linear equations (1.19). beschreibt.
- Determine the LU decomposition of A .
- Calculate the approximation $x_{1+n/2}$ and compare with the value of the exact solution $x(1/2)$.
- Calculate a condition number of A .

1.20 Consider the following sets of sequences of real numbers:

$$\begin{aligned} l_\infty &:= \{(x_n)_{n \in \mathbb{N}} \mid \| (x_n)_n \|_\infty := \sup_n |x_n| < \infty\}, \\ l_p &:= \{(x_n)_{n \in \mathbb{N}} \mid \| (x_n)_n \|_p := (\sum_{i=1}^\infty |x_n|^p)^{\frac{1}{p}} < \infty\}, \quad 1 \leq p < \infty \\ c &:= \{(x_n)_{n \in \mathbb{N}} \mid \lim_n x_n \text{ exists}\}, \\ c_0 &:= \{(x_n)_{n \in \mathbb{N}} \mid \lim_n x_n = 0\}. \end{aligned}$$

Show for $1 \leq p < \infty$:

$$l_p \subset c_0 \subset c \subset l_\infty$$

1.21 Consider the left-shift operator defined for a sequence of real numbers $x := (x_n)_{n \in \mathbb{N}}$ by

$$A(x) := (x_2, x_3, x_4, \dots).$$

Show:

- A is a linear and continuous operator which maps l_p to l_p and has operator norm $\|A\| = 1$ for each $1 \leq p \leq \infty$. Here the norm $\|\cdot\|_p$ is defined as above.
- Is there a sequence e with the property $Ae = e$ other than the sequence $(0, 0, \dots)$?

1.22 Consider the mapping

$$T: c(\mathbb{K}) \longrightarrow c_0(\mathbb{K}), (x_n)_{n \in \mathbb{N}} \longmapsto (y_n)_{n \in \mathbb{N}} \text{ with } y_n := \begin{cases} \lim_k x_k & , \text{ if } n = 1 \\ x_n - \lim_k x_k & , \text{ else} \end{cases}.$$

- (a) Show that T is bijective and continuous.
- (b) Compute $\|T\|, \|T^{-1}\|$.

1.23 In statistics the sample variance of n numbers x_1, \dots, x_n is defined as

$$s_n^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{where } \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i \text{ (two-pass-formula)} \quad (1.20)$$

$$s_n^2 := \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) \text{ (one-pass-formula)} \quad (1.21)$$

The computational behavior is very poor in the presence of (rounding-)errors. Instead consider the updating formulae

$$\begin{aligned} M_1 &:= x_1, & M_k &:= M_{k-1} + (x_k - M_{k-1})/k, & k &= 2, \dots, n, \\ Q_1 &:= 0, & Q_k &:= Q_{k-1} + (k-1)(x_k - M_{k-1})^2/k, & k &= 2, \dots, n \end{aligned} \quad (1.22)$$

- (a) Show that the updating formulae (1.22) result in $s_n^2 = Q_n/(n-1)$.
- (b) Compare the results for s_n^2 in the case

$$x_1 := 10000, x_2 := 10001, x_3 := 10002, x_4 := 10003$$

via the formulae in (1.20), (1.21), (1.22) respectively. Compute in single precision arithmetic.

1.24 Consider the Hilbert matrix of order 3:

$$H_3 := \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}.$$

- (a) Compute the inverse T_3 of H_3 .
- (b) Compute the inverse \tilde{T}_3 of H_3 by Gaussian elimination by using the floating-point number system $\mathbb{F}(10, 3, -\infty, \infty)$. Compare \tilde{T}_3 with $\text{rd}(T_3)$.

Literaturverzeichnis

- [1] K. Atkinson and W. Han. *Theoretical Numerical Analysis. A Functional Analysis Framework*. Texts in applied mathematics, vol. 39. Springer, New York, 2001.
- [2] J. Baumeister. *Stable Solution of Inverse Problems*. Vieweg, Braunschweig, 1987.
- [3] D. Colton and R. Kress. *Integral Equation Methods in Scattering Theory*. Wiley, New York, 1983.
- [4] G. Emde. Active electrolocation of objects in weakly electric fish. *The Journal of Experimental biology*, 202:1205–1215, 1999.
- [5] V.M. Fridman. Method of successive approximations for fredholm integral equations of the first kind. *Uspekhi Mat. Nauk*, 11:233–234, 1956.
- [6] C.W. Groetsch. *Inverse Problems in Mathematical Sciences*. Vieweg, Braunschweig, 1993.
- [7] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52, 1902.
- [8] H.W. Engl, M. Hanke and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.
- [9] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, Berlin, 2005.
- [10] J. Kaipio and E. Somersalo. Statistical inverse problems: discretization, model reduction and inverse crimes. *J. of Computational and Applied Mathematics*, 198:493–504, 2007.
- [11] A. Kirchgraber, D. Stoffer und A. Kirsch. Schlecht gestellte Probleme – oder Wenn das Ungenaue genauer ist. *Mathematische Semesterberichte*, 51:175–206, 2005.
- [12] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Springer, New York, 1996.
- [13] C. Lanczos. Evaluation of noisy data. *J. SIAM Numer. Anal.*, 1:76–85, 1964.
- [14] L. Landweber. An iteration formula for fredholm integral equations of the first kind. *Amer. J. Math.*, 73:615–624, 1951.
- [15] A.K. Louis. *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart, 1989.
- [16] R.W. McKelvey. An analysis of approximative methods for fredholm’s integral equation of the first kind. Technical report, 1956.
- [17] F. Natterer. *The Mathematics of Computerized Tomography*. John Wiley, New York, 1986.

- [18] F. Natterer. Imaging and inverse problems of partial differential equations. *Jahresbericht der DMV*, 109:31–48, 2006.
- [19] D.L. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9:84–97, 1962.
- [20] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Ber. Verh. Sächs. Akad. Wiss.*, 69:262–277, 1917.
- [21] A. Rieder. *Keine Probleme mit Inversen Problemen*. Vieweg, Braunschweig, 2003.
- [22] A.N. Tikhonov. On the stability of inverse problems. *Doklady Akademii Nauk SSSR*, 39:195–198, 1943.
- [23] A.N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 4:1035–1038, 1963.
- [24] A.N. Tikhonov and V.Y. Arsenin. *Solutions of Ill-Posed Problems*. Winston & Sons, Washington, D.C., 1977.
- [25] L. von Wolfersdorf. Inverse und schlecht gestellte Probleme, 1993. Sitzungsberichte der sächsischen Akademie der Wissenschaften zu Leipzig.
- [26] A. Wirgin. The inverse crime. *arXiv:math-ph/0401050v1*, 2004.