

Vladimir N. Troyan

INVERSE GEOPHYSICAL PROBLEMS

Course of lectures

Saint Petersburg University
2001

Introduction

The solution of a geophysical inverse problem is usually obtained by a combination from observed data, the theoretical relation between data and earth parameters (forward modeling) and prior information on data and models. In geophysical analysis, it is very important to choose a forward modeling procedure which will adequately describe the observations. In addition to the choice of an appropriate mathematical model, it is also important to know how many models parameters should be used and which parameters are significant. The appropriateness of modeling choices will depend on the exploration problem at hand and on the geologic area of interest. For a given data set, inversion seeks to define a geologic model which agrees with the observations. Inherent in the inversion process is an attempt to determine rock property parameters which allow model responses to fit the available data. Hence, the choice of an appropriate model is as important for inversion as for forward modeling, and the geophysicist should always be concerned with the physical basis of the inversion model. Due to uncertainties in the data and model, probability-statistical theory and methods are used as a tool to describe the inverse problems.

Inverse problems are very important field of a human activity not only in Geophysics. For example, detectives and doctors solve inverse problems every day in their practice. A detective should reconstruct a crime on some kind of delicts and convict the criminal. A doctor should do a diagnosis on the heart illness by using a cardiogram. Recent decades widely practice has become using the tomography methods in medicine to reconstruct images of the different parts of a human body. Recently, geophysical tomography methods are well known as very effective approaches for the reconstruction of an internal structure of the Earth by using geophysical data measured on a surface of the Earth and in wells.

Inverse problems have common features and very similar mathematical methods of solutions independently from the physical nature of the measurement data and reconstructed parameters:

- These problems are ill-posed or incorrect, i.e. small variation in the measurement data can lead to large oscillations in the solution.
- This problem is nonunique, we have not got enough information from observed data to provide the uniqueness of the solution of the inverse problem.
- The very important feature of inverse problems is a presence of a random noise in the real geophysical data, which leads to the necessity of the application of probability-statistical methods to the solution of the inverse geophysical problems.

For the reasonable application of the inversion theory and methods it is very important to understand:

How do you construct of measurement data?

What role does a priori information play by the solution of inverse problem?

What method should you use to decide your particular geophysical problem?

What is the quality of your solution?

On these questions and many others we would like to answer in our course.

We shall remain the basic definitions and principles of the statistical interpretation of geophysical data and give the most important theoretical results of the solution of the inverse geophysical problems and consider its practical applications.

The main milestones of the course are the following:

Introduction to the probability theory. Definition of probability, Bayes formula, distribution of probability (Gaussian, uniform, Laplace etc.), numerical characteristics of distribution (mean value, median, variance, correlation coefficient, covariance matrix, confidence interval).

Information and entropy. Shannon and Fisher information and its properties, Fisher information matrix.

Random functions and its properties. Random processes and time series, auto-correlation and cross-correlation functions, power spectrum and correlation function.

Elements of mathematical statistics. Sampling, estimation, robust estimator, Rao-Kramer inequality.

Models of measurement data. Model space for the description of an unknown parameter field, for example, velocities of P- and S-waves, density, conductivity and etc; construction of a measurement data model; examples of forward modeling and inverse setting problems; the influence of modelization and measurement errors; a priori information about unknown parameters.

Solution of inverse geophysical problem: case of linear model. Parameter estimation methods by the solution of the linear inverse problem: minimum least squares method, maximum likelihood method, singular analysis method, least modulus method, Bayes method, regularization methods and a priori information.

Solution of inverse geophysical problem: case of nonlinear model. Parameter estimation methods by the solution of the nonlinear problem: Newton method, Newton-Le Came method, linearization procedure, a choice of an initial parameter vector, genetic algorithm.

Statistical criteria for choice of models. Parametric hypothesis test, criterion of a posteriori probability ration, the signal resolution problem.

Ray geophysical tomography method. Ray tomography: Radon transform, Gilbert transform, back projection, algebraic method, recurrent procedure, recon-

struction of slowness and absorption coefficient.

Diffraction tomography method. Perturbation method for scalar wave equation, Born approximation, tomography functionals, reconstruction algorithm.

Multifactor analysis and processing of time series. Analysis and selection of principal components.

The suggested course of lectures is illustrated by computer examples and computer exercises.

1

Introduction to the probability

The geophysical observations are the results of random experiments and are random events. Because of the random nature of the observations the question arises as to the probability that these random events occur. By means of geophysical observations unknown parameters of the Earth structure and confidence regions for the parameters will be estimated and the hypotheses for the parameters will be tested. This is the problem, which the theory of probability deals with.

1.1 The definition of probability

A central concept in the theory of probability is the *random event*. These random events are results from measurements or experiments, which are uncertain. It is important to know *the probability* with which random events occur.

1.1.1 Set of elementary events

The outcome of a certain experiment ω_i is called an *elementary event*. All events are collected the *set Ω of elementary events*:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}.$$

Example 1: Coin-tossing game: “head and tails” $\Omega = \{H, T\}$.

Example 2: n -coin-tossing case:

$$\Omega = \{\omega : \omega = (a_1, a_2, \dots, a_n), a_i = \{H, T\}\}.$$

The total number of outcomes $N(\Omega) = 2^n$.

Set operations

A set A is a *subset* of Ω if every element of A is also an element of Ω , $A \subset \Omega$ (see Figure 1.1). The set that contains no elements is called *null set* or *empty set* and is denoted \emptyset .

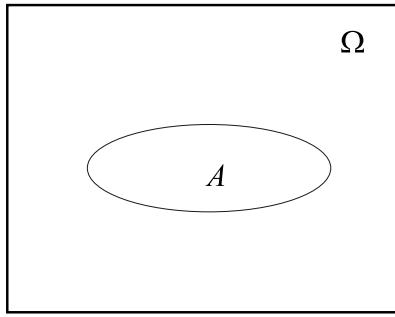


Figure 1.1: The set A is subset of Ω .

Equality. The set A_1 equals to the set A_2 if and only if (iff) every element of A_1 is an element of A_2 and every element of A_2 is an element of A_1 (see Figure 1.2):

$$A_1 = A_2 \quad \text{iff} \quad A_1 \subset A_2 \quad \text{and} \quad A_2 \subset A_1.$$

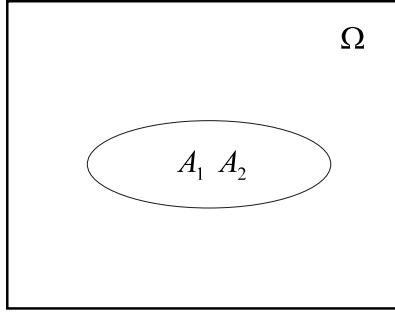


Figure 1.2: The subset A_1 is equal to subset A_2 .

Product. The *product* or *intersection*

$$A_1 \cdot A_2 = A_1 \cap A_2 = \{\omega \in \Omega : \omega \in A_1 \text{ and } \omega \in A_2\}$$

of two sets A_1 and A_2 is a set consisting of all elements that are common to set A_1 and set A_2 (see Figure 1.3)

$$A_1 \cdot A_2 = A_2 \cdot A_1, \quad A_1 \cdot \emptyset, \quad A_1 \cdot \Omega = A_1.$$

The sets A_1 and A_2 are called *mutually exclusive* (or *disjoint*) if they have no common elements

$$A_1 \cdot A_2 = \emptyset.$$

The sets A_1, A_2, \dots are called *mutually exclusive* if any two of them have no common elements i.e. if

$$A_i \cdot A_j = \emptyset \text{ for every } i, j, \quad i \neq j.$$

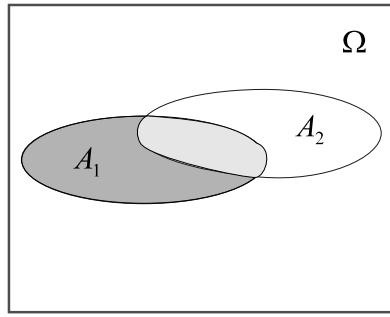


Figure 1.3: The intersection of two subsets.

Sums. The *union*

$$A_1 \cup A_2 = \{\omega \in \Omega : \omega \in A_1 \text{ or } \omega \in A_2\}$$

of two sets A_1 and A_2 is a set whose elements are all the elements of A_1 or of A_2 or of both (see Figure 1.4). The next relations are valid

$$A_1 \cup A_2 = A_2 \cup A_1, \quad A_1 \cup A_1 = A_1, \quad A_1 \cup \emptyset = A_1, \quad A_1 \cup \Omega = \Omega.$$

If $A_1 \cap A_2 = \emptyset$ then in the case of the union operation a sign + instead of \cup is used and this operation is named the *sum*.

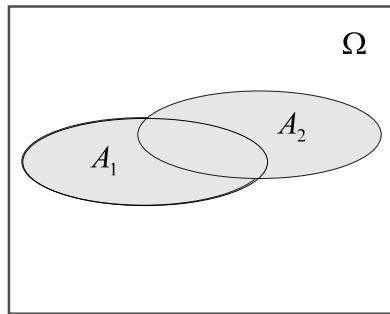


Figure 1.4: The union of two subsets.

Complement. The *complement* \bar{A} of a set A is defined as the set consisting of all elements of Ω that are not in A (see Figure 1.5). The next relations are valid

$$\begin{aligned} \bar{A} &= \Omega \setminus A = \Omega - A, & \bar{A} \cdot A &= \bar{A} \cap A = \emptyset, \\ \bar{\emptyset} &= \Omega, & \bar{\Omega} &= \emptyset, & A + \bar{A} &= \Omega. \end{aligned}$$

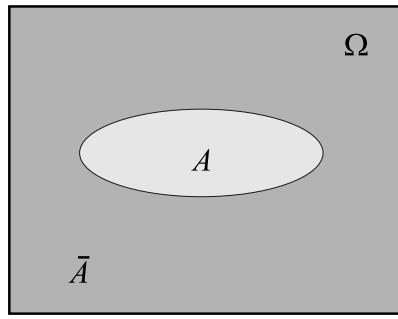


Figure 1.5: The complement of subset A is subset \bar{A} .

Difference. The *difference* $A_1 - A_2$ is a set consisting of the elements of A_1 that are not in A_2 (see Figure 1.6). The next relations are valid

$$A - \emptyset = A, \quad A - \bar{A} = \emptyset, \quad \Omega - A = \bar{A}.$$

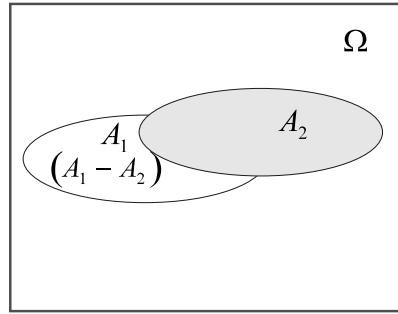


Figure 1.6: The difference of the two subsets.

De Morgan law

It is easy to see that

$$\overline{A_1 \cup A_2} = \bar{A}_1 \cdot \bar{A}_2, \quad (\overline{A_1 \cdot A_2}) = \bar{A}_1 \cup \bar{A}_2.$$

If in an identity we replace sums by products, products by sums, and sets by their complements, the identity is preserved (see Figure 1.7). The set Ω is a *certain event*. The empty set \emptyset is an *impossible event*.

Events algebra

The set \mathcal{A} satisfying to conditions:

1. $\Omega \in \mathcal{A}$.

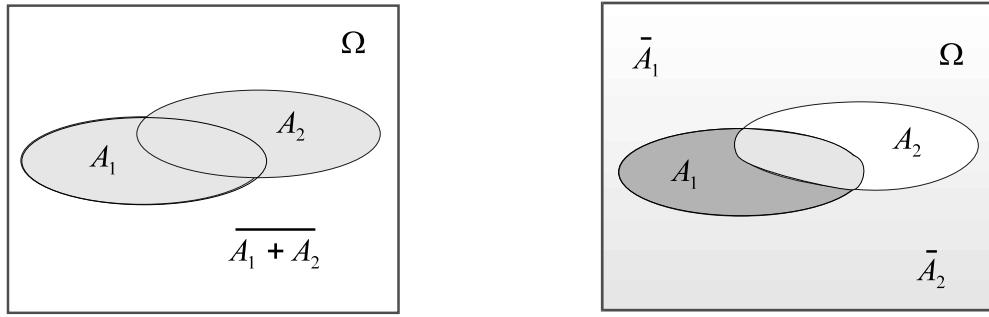


Figure 1.7: An illustration of the Morgan law.

2. $\emptyset \in \mathcal{A}$.
3. If $A \in \mathcal{A}$ then $\bar{A} \in \mathcal{A}$.
4. If $A_1 \in \mathcal{A}$ and $A_2 \in \mathcal{A}$ then $A_1 \cup A_2 \in \mathcal{A}$ and $A_1 \cap A_2 \in \mathcal{A}$; then \mathcal{A} is called an *algebra of events*.

1.1.2 Axiomatic definition, probability space

1. Space of elementary events – Ω .
2. Subset system is an algebra of events – \mathcal{A} .
3. Probability of outcomes (Axioms of probability).

Axiom 1: A real number $P(\omega_i) \geq 0$ is assigned to every event $\omega_i \in \Omega$ ($i = 1, 2, \dots, n$). $P(\omega_i)$ is called the *probability* of ω_i .

Axiom 2: The probability of the certain event is equal to one, $P(\Omega) = 1$.

Axiom 3: If $\omega_1, \omega_2, \dots$ is a sequence of events of Ω , which are countable and which are mutually exclusive, that is $\omega_i \cap \omega_j = \emptyset$ for $i \neq j$, then

$$P(\omega_1 \cup \omega_2 \cup \dots) = P(\omega_1) + P(\omega_2) + \dots .$$

Normalization: $p(\omega_1) + p(\omega_2) + \dots + p(\omega_n) = 1$.

The probability of an event $A \in \mathcal{A}$ is

$$P(A) = \sum_{\{i: \omega_i \in A\}} p(\omega_i).$$

The triplet (Ω, \mathcal{A}, P) is called a *probability space*.

1.1.3 Basic properties of probability

From the above axioms we easily conclude that:

1. $P(\emptyset) = 0.$
2. $P(\Omega) = 1.$
3. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$
If $(A_1 \cap A_2) = \emptyset$ then

$$P(A_1 + A_2) = P(A_1) + P(A_2).$$

If A_1, A_2, \dots, A_n are mutually exclusive, then

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

In general case:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i=1}^n \sum_{j=1}^n P(A_i A_j) + \\ &+ \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n P(A_i A_j A_k) + \dots + (-1)^{n-1} P(A_1 A_2 \dots A_n). \end{aligned}$$

4. If $A_1 \subset A_2$, then $P(A_1) \leq P(A_2)$, $P(A_2) = P(A_1) + P(A_2 \bar{A}_1) \geq P(A_1).$
5. $P(\bar{A}) = 1 - P(A)$, $A + \bar{A} = \Omega$, $A\bar{A} = \emptyset$,
 $P(A + \bar{A}) = P(A) + P(\bar{A}) - P(A\bar{A}) = 1.$
6. For arbitrary events A_1, A_2, \dots, A_n an inequality

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

is satisfied.

1.1.4 Relative-frequency definition

The experiment is repeated n times. If the event A occurs n_A times, then its probability $P(A)$ is defined as the limit of the relative frequency n_A/n of A :

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}.$$

For example, if a coin is tossed n times and heads shows n_h times, then the probability of heads equals the limit n_h/n .

1.1.5 Classical definition

The probability $P(A)$ of an event A is found a priori without actual experimentation. This is done by counting the total number n of the possible outcomes. If in n_A of these outcomes the event A occurs, the $P(A)$ is given by

$$P(A) = \frac{n_A}{n},$$

where n_A is named the favorable outcomes to A . Probability $P(A)$ equals the ratio of the favorable to the total number of outcomes.

Example: In the single-die experiment the possible outcomes are 1, 2, 3, 4, 5, 6. For the event “even”, the favorable outcomes are 2, 4, 6; hence $P(\text{“even”}) = 3/6$.

1.1.6 Geometrical definition

An important probability space is the set of numbers on the real axis or on the portion of it.

In general case, the set has geometrical form 1-D, 2-D and 3-D (see Figure 1.8).

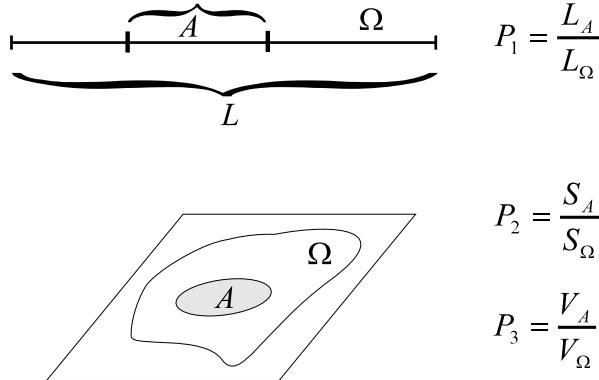


Figure 1.8: Geometrical definition of probability.

1.1.7 Conditional probabilities and independent events

There are situations when the probability of an event is required under the condition that a different event has already occurred.

Let the probability $P(A_2)$ of the random event A_2 be unequal to zero, then the *conditional probability* $P(A_1/A_2)$ of the A_1 , given that the event A_2 has occurred is the ratio

$$P(A_1/A_2) = \frac{P(A_1A_2)}{P(A_2)} \quad \text{with} \quad P(A_2) > 0$$

and we get

$$P(A_1A_2) = P(A_2)P(A_1/A_2).$$

If event A_1 is independent on event A_2 then

$$P(A_1/A_2) = P(A_1)$$

and

$$P(A_1A_2) = P(A_1)P(A_2).$$

For independent A_1, A_2, \dots, A_n we have

$$P(A_1A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n).$$

1.1.8 Total probability

We are given n mutually exclusive events

$$B_1, B_2, \dots, B_n,$$

whose sum equals the certain event Ω

$$\begin{aligned} B_iB_j &= \emptyset, \quad i \neq j, \quad i, j = 1, 2, \dots, n, \\ B_1 + B_2 + \dots + B_n &= \Omega. \end{aligned}$$

Statement: With A an arbitrary event, we shall show that

$$P(A) = P(A/B_1)P(B_1) + \dots + P(A/B_n)P(B_n).$$

Proof: Since $A = A\Omega$, we obtain from

$$A = A(B_1 + \dots + B_n) = AB_1 + \dots + AB_n \quad \text{and} \quad (AB_i)(AB_j) = \emptyset$$

the next equality

$$P(A) = P(AB_1) + \dots + P(AB_n).$$

But

$$P(AB_i) = P(A/B_i)P(B_i)$$

and we have

$$P(A) = \sum_{i=1}^n P(A/B_i)P(B_i).$$

1.1.9 Bayes theorem

We shall show that

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{\sum_{i=1}^n P(B_i)P(A/B_i)},$$

where the events B_i satisfy the conditions $B_iB_j = \emptyset$ if $B_i \neq B_j$, $\sum_{i=1}^n B_i = \Omega$. This formula permits us to evaluate the so-called *a posteriori probabilities* $P(B_i/A)$ of the event B_i in the terms of the (*a priori*) probabilities $P(B_i)$ and the conditional probabilities $P(A/B_i)$.

Proof: We have

$$P(B_i A) = P(A)P(B_i/A) = P(B_i)P(A/B_i).$$

Hence

$$P(B_i/A) = \frac{P(B_i)P(A/B_i)}{P(A)}.$$

1.2 Distribution functions

1.2.1 Random variable

In general, it is preferable to associate the real numbers with the elementary events.

A unique real-valued function $\xi = \xi(\omega_i)$, defined in the probability space (Ω, \mathcal{A}, P) for the set Ω of the elementary events ω_i , is called a *random variable*, if for every $\xi \in \mathcal{R}$ the event, which is characterized by $\xi(\omega_i) < x$, belongs to the random events of \mathcal{A} .

Example: Coin-tossing game: $\omega_1 = HH$, $\omega_2 = HT$, $\omega_3 = TH$, $\omega_4 = TT$, $\xi(\omega_1) = 2$, $\xi(\omega_2) = 2$, $\xi(\omega_3) = 1$, $\xi(\omega_4) = 0$.

$$P_\xi(x_i) = P\{\omega : \xi(\omega) = x_i\}, \quad x_i \in X.$$

The discrete function $\{P_\xi(x_1), \dots, P_\xi(x_n)\}$ is called a *distribution of probabilities* of random variable ξ (see Figure 1.9).

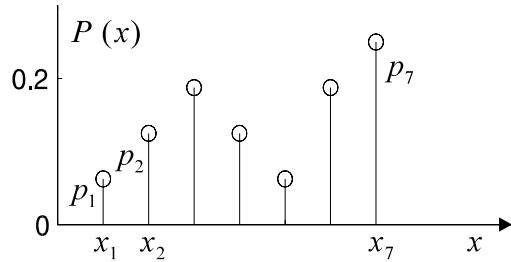


Figure 1.9: Graphic presentation of the distribution of the probabilities.

1.2.2 Definition of distribution function

Given a real number x , the set $\{\xi(\omega) \leq x\}$, consist of all outcomes ω such that $\xi(\omega) \leq x$. Its probability $P\{\xi(\omega) \leq x\}$ is a number depending on x ; that is, it is a function of x . This function will be denoted by

$$F_\xi(x) = P\{\omega : \xi(\omega) \leq x\},$$

defined for any number x from $-\infty$ to ∞ , is called the *distribution function* of the random variable ξ (see Figure 1.10).

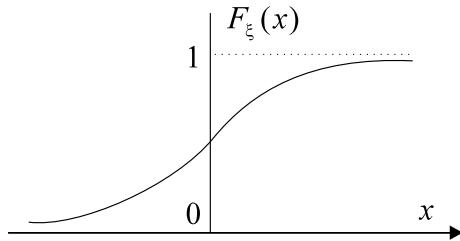


Figure 1.10: An example of the distribution function (continuous case).

The distribution function of the discrete random variable is defined by (see Figure 1.11)

$$F_\xi(x) = \sum_{\{i: x_i \leq x\}} P_\xi(x_i),$$

where

$$P_\xi(x_i) = F_\xi(x_i) - F_\xi(x_i^-), \quad F_\xi(x_i^-) = \lim_{y \rightarrow x_i} F_\xi(y)$$

and for $x_1 < x_2 < \dots < x_{i-1} < x_i$

$$F_\xi(x_0)|_{x_0 < x_1} = 0, \quad P_\xi(x_i) = F_\xi(x_i) - F_\xi(x_{i-1}).$$

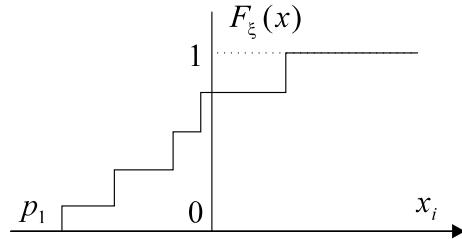


Figure 1.11: An example of the distribution function (discrete case).

1.2.3 Properties of the distribution function

We shall now show that a distribution function $F(x)$ has the following properties:

1. $F_\xi(-\infty) = 0, F_\xi(\infty) = 1.$

2. It is nondecreasing function of x

$$F(x_1) \leq F(x_2) \quad \text{for } x_1 < x_2.$$

3. If $\alpha < \beta$, then

$$P_\xi(\alpha < \xi \leq \beta) = F_\xi(\beta) - F_\xi(\alpha).$$

1.2.4 The density function

The derivative

$$f_\xi(x) = \frac{dF_\xi(x)}{dx}$$

of the distribution function $F_\xi(x)$ is called the *density function* of the random variable ξ .

The properties of the density function.

1. From the monotonicity of $F_\xi(x)$ follows that $f_\xi(x)$ is nonnegative

$$f_\xi(x) \geq 0.$$

2. The distribution function is an integral of the density function

$$F_\xi(x) = \int_{-\infty}^x f_\xi(y) dy,$$

hence

$$F_\xi(x_2) - F_\xi(x_1) = \int_{x_1}^{x_2} f_\xi(x) dx.$$

- 3.

$$\int_{-\infty}^{\infty} f_\xi(x) dx = F_\xi(\infty) - F_\xi(-\infty) = 1.$$

4. The probability of event $x_1 < \xi \leq x_2$ is the integral of the density function (see Figure 1.12)

$$P(x_1 < \xi \leq x_2) = \int_{x_1}^{x_2} f_\xi(x) dx.$$

If Δx is sufficiently small,

$$P(x < \xi \leq x + \Delta x) \simeq f(x)\Delta x.$$

The density function can be defined as a limit

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x < \xi \leq x + \Delta x)}{\Delta x}$$

and $P(\xi = x) = 0$.

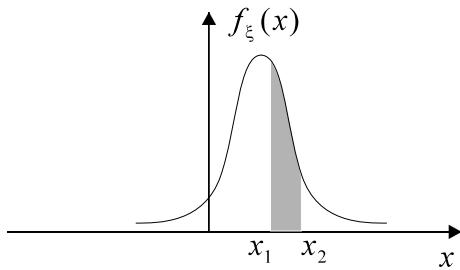


Figure 1.12: Geometrical presentation of the probability as an integral of the density function.

1.2.5 The distribution and density of function of one random variable

We shall now determine the density $f_\eta(y)$ of the random value

$$\eta = \varphi(\xi)$$

in the terms of the density $f_\xi(x)$ of ξ . The function $\varphi(\xi)$ is a continue differentiable and monotone on ξ .

Let ψ be an inverse function to the function φ , then the density function is

$$f_\eta(y) = f_\xi(\psi(y))|\psi'(y)|.$$

For $\eta < y \{\Delta_1(y), \Delta_2(y), \dots, \Delta_n(y)\}$ distribution function has the following presentation (see Figure 1.13)

$$\begin{aligned} F_\eta(y) &= P\{\omega : \eta(\omega) \leq y\} = \\ &= P\{\omega : (\xi(\omega) \in \Delta_1(y) + \xi(\omega) \in \Delta_2(y) + \dots)\} = \\ &= \sum_i P\{\omega : \xi(\omega) \in \Delta_i(y)\}. \end{aligned}$$

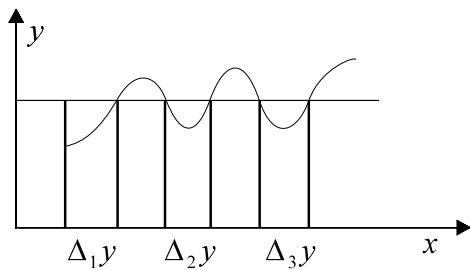


Figure 1.13: An illustration for the change of variables.

Example: If the random variable ξ has the normal distribution

$$\xi \in N(m_\xi, \sigma_\xi^2), \quad f_\xi(x) = \frac{\exp\left(-\frac{(x-m_\xi)^2}{2\sigma_\xi^2}\right)}{\sqrt{2\pi}\sigma_\xi},$$

then the linear function $\eta = a_1 + a_2\xi$ has the density function:

$$f_\eta(y) = \frac{1}{|a_2|\sigma_\xi\sqrt{2\pi}} \exp\left(-\frac{[y - (a_2m_\xi + a_1)]^2}{2|a_2|^2\sigma_\xi^2}\right),$$

$$m_\eta = a_2m_\xi + a_1, \quad \sigma_\eta = |a_2|\sigma_\xi \quad (f_\eta(y) = N(m_\eta, \sigma_\eta^2)),$$

because of

$$f_\eta(y) = f_\xi(\psi(y))|\psi'(y)|, \quad \psi(y) = (y - a_1)/a_2, \quad |\psi'(y)| = 1/|a_2|.$$

1.2.6 Random vectors

A unique function $\vec{\xi}(\omega_i) = (\xi_1(\omega_i), \dots, \xi_m(\omega_i))$ with values in \mathcal{R}^n , defined in the probability space (Ω, \mathcal{A}, P) for the set Ω of the elementary events ω_i is called *m-dimensional random vector*. Distribution function of the random vector is

$$F_\xi(x_1, \dots, x_m) = P\{\omega : \xi_1(\omega) \leq x_1, \dots, \xi_m(\omega) \leq x_m\}.$$

Let be given two random variables (ξ, η) . The joint distribution function of the random variables ξ and η is defined by

$$F_{\xi\eta}(x, y) = P\{\omega : \xi(\omega) \leq x, \eta(\omega) \leq y\}.$$

Properties of the distribution function:

1. $x_2 > x_1, F_{\xi\eta}(x_2, y) \geq F_{\xi\eta}(x_1, y), \quad y_2 > y_1, F_{\xi\eta}(x, y_2) \geq F_{\xi\eta}(x, y_1).$
2. $F_{\xi\eta}(x, -\infty) = F_{\xi\eta}(-\infty, y) = F_{\xi\eta}(-\infty, -\infty) = 0.$
3. $F_{\xi\eta}(x, +\infty) = F_\xi(x), F_{\xi\eta}(+\infty, y) = F_\eta(y), F_{\xi\eta}(+\infty, +\infty) = 1.$

The quantity

$$f_{\xi\eta}(x, y) = \frac{\partial^2 F_{\xi\eta}(x, y)}{\partial x \partial y}$$

is known as the *joint density function*. The probability of the event $(\xi(\omega), \eta(\omega)) \in D$ is

$$P\{\omega : (\xi(\omega), \eta(\omega)) \in D\} = \iint_D f_{\xi\eta}(x, y) dx dy.$$

and the distribution function can be presented as

$$F_{\xi\eta}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{\xi\eta}(p, q) dp dq.$$

Properties of density function:

1. $f_{\xi\eta}(x, y) \geq 0$.
2. $\iint_{-\infty}^{\infty} f_{\xi\eta}(x, y) dx dy = 1$.

1.2.7 Marginal and conditional distributions

Marginal distributions are

$$\begin{aligned} F_\xi(x) &= F_{\xi\eta}(x, \infty), \quad F_\eta(y) = F_{\xi\eta}(\infty, y), \\ F_\xi(x) &= \int_{-\infty}^x \int_{-\infty}^{\infty} f(p, q) dp dq, \quad F_\eta(y) = \int_{-\infty}^{\infty} \int_{-\infty}^y f(p, q) dp dq \end{aligned}$$

and the *marginal densities*:

$$f_\xi(x) = F'_\xi(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_\eta(y) = F'_\eta(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Conditional densities (see Figure 1.14):

$$\begin{aligned} f_{\xi/\eta}(x/y) &= \frac{f_{\xi\eta}(x, y)}{f_\eta(y)} = \frac{f_{\xi\eta}(x, y)}{\int_{-\infty}^{\infty} f_{\xi\eta}(x, y) dx}, \\ f_{\eta/\xi}(y/x) &= \frac{f_{\xi\eta}(x, y)}{f_\xi(x)} = \frac{f_{\xi\eta}(x, y)}{\int_{-\infty}^{\infty} f_{\xi\eta}(x, y) dy}, \\ f_{\xi\eta}(x, y) &= f_\eta(y) f_{\xi/\eta}(x/y) = f_\xi(x) f_{\eta/\xi}(y/x). \end{aligned}$$

The random variables ξ and η are independent, if the expressions are satisfied the next relations

$$f_{\xi/\eta}(x/y) = f_\xi(x), \quad f_{\eta/\xi}(y/x) = f_\eta(y), \quad f_{\xi\eta}(x, y) = f_\xi(x) f_\eta(y).$$

The analog of the Bayes formula:

$$f_{\xi/\eta}(x, y) = \frac{f_\xi(x) f_{\eta/\xi}(y/x)}{\int_{-\infty}^{\infty} f_{\xi\eta}(x, y) dx}.$$

1.2.8 Joint distribution and density functions

Let two random variables (ξ, η) be described by the density $f_{\xi\eta}(x, y)$ and random value $\zeta = \varphi(\xi, \eta)$ has the distribution function $F_\zeta(z)$ (see Figure 1.15):

$$\begin{aligned} F_\zeta(z) &= P\{\omega : \zeta(\omega) < z\} = P\{\omega : \varphi(\xi(\omega), \eta(\omega)) < z\}, \\ F_\zeta(z) &= P\{\omega : (\xi, \eta) \in \mathcal{D}\} = \iint_{\mathcal{D}(z)} f_{\xi\eta}(x, y) dx dy. \end{aligned}$$

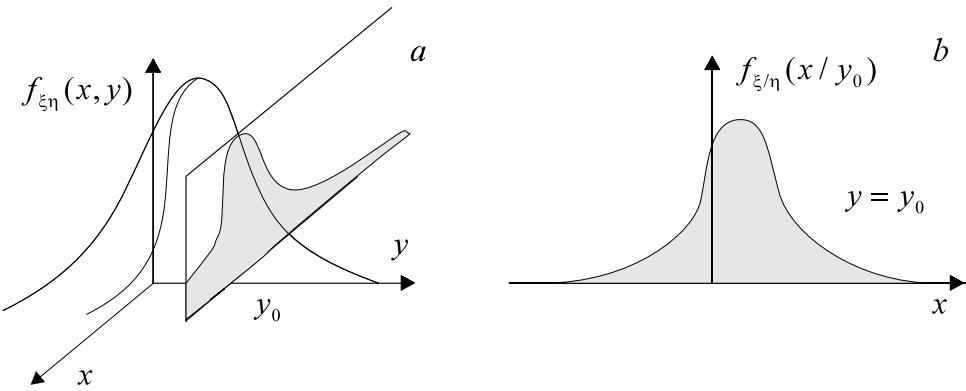


Figure 1.14: Graphic presentation of the joint and conditional density functions. (a) – the joint density function; (b) – the conditional density function.

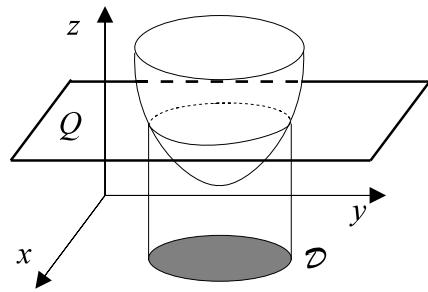


Figure 1.15: An illustration of calculation of the joint distribution function.

Example 1: Let us $\zeta = \xi + \eta$. This is a most important example of a function of two random variables. To determine $F_\zeta(z)$ we note that the region $\mathcal{D}(z)$ of the (x, y) plane, such that

$$x + y < z,$$

(see Figure 1.16) is the half plane to the left of the line $x + y = z$. Differentiating the distribution function

$$\begin{aligned} F_\zeta(z) &= \iint_{\mathcal{D}(z)} f_{\xi\eta}(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{\xi\eta}(x, y) dx dy = \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{z-x} f_{\xi\eta}(x, y) dy \right\} dx \end{aligned}$$

with respect to z we obtain

$$f_\zeta(z) = \int_{-\infty}^{\infty} f_{\xi\eta}(x, z-x) dx \quad \text{or} \quad f_\zeta(z) = \int_{-\infty}^{\infty} f_{\xi\eta}(z-y, y) dy.$$

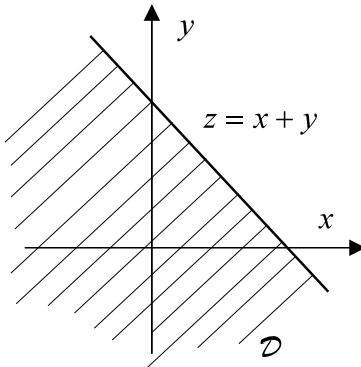


Figure 1.16: The region for integration in the (x, y) plane.

If ξ and η are independent random variables, then the density of their sum $\zeta = \xi + \eta$ equals the convolution of their respective densities

$$f_\zeta(z) = \int_{-\infty}^{\infty} f_\xi(x) f_\eta(z - x) dx, \quad f_\zeta(z) = \int_{-\infty}^{\infty} f_\xi(z - y) f_\eta(y) dy,$$

or

$$f_\zeta(z) = f_\xi * f_\eta.$$

Example 2: We assume that the random variables ξ and η have the normal distributions

$$\begin{aligned} \xi &\in N(m_\xi, \sigma_\xi^2), \quad f_\xi(x) = \frac{1}{\sigma_\xi \sqrt{2\pi}} \exp\left(-\frac{(x - m_\xi)^2}{2\sigma_\xi^2}\right), \\ \eta &\in N(m_\eta, \sigma_\eta^2), \quad f_\eta(y) = \frac{1}{\sigma_\eta \sqrt{2\pi}} \exp\left(-\frac{(y - m_\eta)^2}{2\sigma_\eta^2}\right). \end{aligned}$$

Let $\zeta = \xi + \eta$. What is the function $f_\zeta(z)$?

The density function $f_\zeta(z)$ is the convolution

$$f_\zeta(z) = \int_{-\infty}^{\infty} f_\xi(x) f_\eta(z - x) dx,$$

$$f_\zeta(z) = \frac{1}{2\pi\sigma_\xi\sigma_\eta} \int_{-\infty}^{\infty} \exp\left(-\frac{(x - m_\xi)^2}{2\sigma_\xi^2} - \frac{(z - x - m_\eta)^2}{2\sigma_\eta^2}\right) dx,$$

$$f_\zeta(z) = \frac{1}{2\pi\sigma_\xi\sigma_\eta} \int_{-\infty}^{\infty} \exp(-a_2x^2 + 2a_1x - a_0) dx,$$

$$a_2 = \frac{1}{2} \frac{\sigma_\xi^2 + \sigma_\eta^2}{\sigma_\xi \sigma_\eta}, \quad a_1 = \frac{m_\xi}{2\sigma_\xi^2} + \frac{z - m_\eta}{2\sigma_\eta^2}, \quad a_0 = \frac{m_\xi^2}{2\sigma_\xi^2} + \frac{(z - m_\eta)^2}{2\sigma_\eta^2},$$

$$\int_{-\infty}^{\infty} \exp(-a_2x^2 + 2a_1x - a_0) dx = \sqrt{\frac{\pi}{a_2}} \exp\left(-\frac{a_2a_0 - a_1^2}{a_2}\right)$$

and we get

$$f_\zeta(z) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_\xi^2 + \sigma_\eta^2}} \exp\left(-\frac{[z - (m_\xi + m_\eta)]^2}{2(\sigma_\xi^2 + \sigma_\eta^2)}\right),$$

i.e.

$$\zeta \in N(m_\zeta, \sigma_\zeta^2), \quad m_\zeta = m_\xi + m_\eta, \quad \sigma_\zeta^2 = \sigma_\xi^2 + \sigma_\eta^2.$$

1.3 Expected values and moments of random variables

Let (Ω, \mathcal{A}, P) be the probability space and $\xi(\omega)$ be a random variable $X = \{x_1, \dots, x_n\}$. If the random event $A_i = \{\omega : \xi(\omega) = x_i\}$ ($i = 1, \dots, n$) then

$$\xi(\omega) = \sum_{i=1}^n x_i I(A_i); \quad I(A_i) = \begin{cases} 1 & \omega \in A_i, \\ 0 & \omega \notin A_i. \end{cases}$$

1.3.1 Mathematical expectation

The *expectation* or the *expected value* of a random variable is obtained, if the average is formed of all possible values of a random variable by taking their probabilities into account. If ξ is of discrete type, taking the value x_i with probability $P(A_i)$, we introduce a definition of the expectation as

$$M\xi = \sum_{i=1}^n x_i P(A_i), \quad A_i = \{\omega : \xi(\omega) = x_i\}, \quad P_\xi(x_i) = P(A_i),$$

$$M\xi = \sum_{i=1}^n x_i P_\xi(x_i); \quad \xi \in R^1.$$

The expected value of a real random variable ξ is the integral

$$M\xi = \int_{-\infty}^{\infty} x f_\xi(x) dx,$$

where $f_\xi(x)$ is the density function of ξ (see Figure 1.17).

Properties of the mathematical expectation.

- $\xi \geq 0, M\xi \geq 0$.

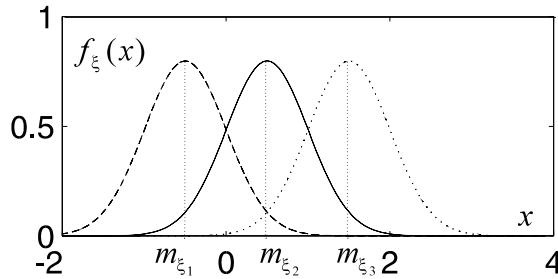


Figure 1.17: An example of graphic presentation of the normal densities with different mathematical expectations. Standard deviations are $\sigma_{\xi_1} = \sigma_{\xi_2} = \sigma_{\xi_3} = 0.5$, Mathematical expectations are $m_{\xi_1} = -0.5$, $m_{\xi_2} = 0.5$, $m_{\xi_3} = 1.5$.

- $M(a\xi + b\eta) = aM\xi + bM\eta$.
- $\xi \geq \eta$, $M\xi \geq M\eta$.
- $|M\xi| \leq M|\xi|$.
- If ξ and η are independent r.v., then $M(\xi \cdot \eta) = M\xi \cdot M\eta$.
- $(M|\xi\eta|)^2 \leq M\xi^2 \cdot M\eta^2$ — Cauchy-Bunyakovskii inequality.
- $M\varphi(\xi) = \sum_i \varphi(x_i)P_\xi(x_i)$, $M\varphi(\xi) = \int_{-\infty}^{\infty} \varphi(x)f_\xi(x)dx$.

1.3.2 Variance

The mean $M\xi$ of a random variable places the center of gravity of $f(x)$. Another important parameter is its *variance* or *dispersion* σ^2 , defined by

$$\xi(\omega) : D\xi = M(\xi - M\xi)^2, \quad \sigma_\xi = \sqrt{D\xi}.$$

This quantity equals the moment of inertia of the probability masses and gives some notion of their concentration near $M\xi$. Its positive square root σ_ξ is called *standard deviation* (an example of the random uncorrelated Gaussian time series with different standard deviations is presented in Figure 1.18). If ξ is of discrete type, then

$$\sigma_\xi^2 = \sum_i (x_i - M\xi)^2 P_\xi(x_i).$$

1.3.3 Properties of variance and correlation coefficients

- $D\xi = M\xi^2 - (M\xi)^2$,

$$D\xi = M(\xi - M\xi)^2 = M\xi^2 - 2(M\xi)^2 + (M\xi)^2 = M\xi^2 - (M\xi)^2$$
.
- $D(\xi) \geq 0$.

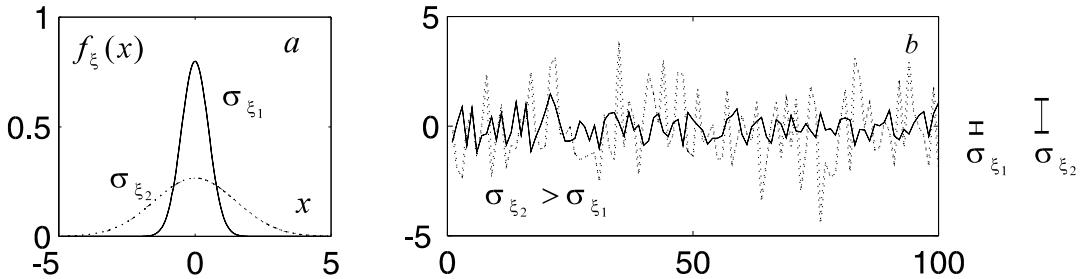


Figure 1.18: An example of uncorrelated Gaussian time series (*b*) with two different standard deviations and their Gaussian density functions (*a*). Standard deviations are $\sigma_{\xi_1} = 0.5$ $\sigma_{\xi_2} = 1.5$. Mathematical expectations are $m_{\xi_1} = m_{\xi_2} = 0$.

- $D(a_1 + a_2\xi) = a_2^2 D\xi$,
 $D(a_1 + a_2\xi) = M(a_1 + a_2\xi - a_1 - a_2M\xi)^2 = a_2^2 M(\xi - M\xi)^2 = a_2^2 D\xi$.
- $D(\xi + \eta) = D\xi + D\eta + 2R_{\xi\eta}$,
 $D(\xi + \eta) = M((\xi - M\xi) + (\eta - M\eta))^2 =$
 $= M(\xi - M\xi)^2 + M(\eta - M\eta)^2 + 2M(\xi - M\xi)(\eta - M\eta) = D\xi + D\eta + 2R_{\xi\eta}$,
where $R_{\xi\eta} = \text{cov}(\xi, \eta) = M(\xi - M\xi)(\eta - M\eta)$ is a *covariation coefficient* and
 $r(\xi, \eta) = r_{\xi\eta} = \text{cov}(\xi, \eta)/(\sqrt{D\xi}\sqrt{D\eta}) = R_{\xi\eta}/\sigma_\xi\sigma_\eta$ is a *correlation coefficient* ($-1 \leq r_{\xi\eta} \leq 1$) (see Figure 1.19).

Let r.v. η be linear function of r.v. ξ

$$\eta = a_1\xi + a_2.$$

By definition of correlation coefficient we can write

$$\begin{aligned} r_{\xi\eta} &= M(\xi - M\xi)(a_1\xi + a_2 - a_1M\xi - a_2)/\sigma_\xi|a_1|\sigma_\eta = \\ &= \frac{a_1}{|a_1|} \frac{D\xi}{\sigma_\xi^2} = \frac{a_1}{|a_1|}. \end{aligned}$$

If $a_1 > 0$, then $r_{\xi\eta} = 1$. If $a_1 < 0$ then $r_{\xi\eta} = -1$.

If r.v. ξ and η are independent, then

$$\begin{aligned} \text{cov}(\xi, \eta) &= M(\xi - M\xi) \cdot M(\eta - M\eta) = (M\xi - M\xi)(M\eta - M\eta) = 0 \\ \text{and in this case } D(\xi + \eta) &= D\xi + D\eta. \end{aligned}$$

- $M(\xi \cdot \eta) = M\xi \cdot M\eta + R_{\xi\eta}$,
 $R_{\xi\eta} = M(\xi - M\xi)(\eta - M\eta) = M(\xi\eta) - M\xi M\eta - M\eta M\xi + M\xi M\eta =$
 $= M(\xi\eta) - M\xi M\eta$. If r.v. ξ and η are independent, then $M(\xi\eta) = M\xi M\eta$.
- $D(\sum_{i=1}^n a_i\xi_i + a_0) = \sum_{i=1}^n a_i^2 D\xi_i + 2 \sum_{i < j} a_i a_j R_{\xi_i \xi_j}$,
 $D(\sum_{i=1}^n a_i\xi_i + a_0) = M(\sum_{i=1}^n a_i\xi_i + a_0 - \sum_{i=1}^n a_i M(\xi_i) + a_0)^2 =$

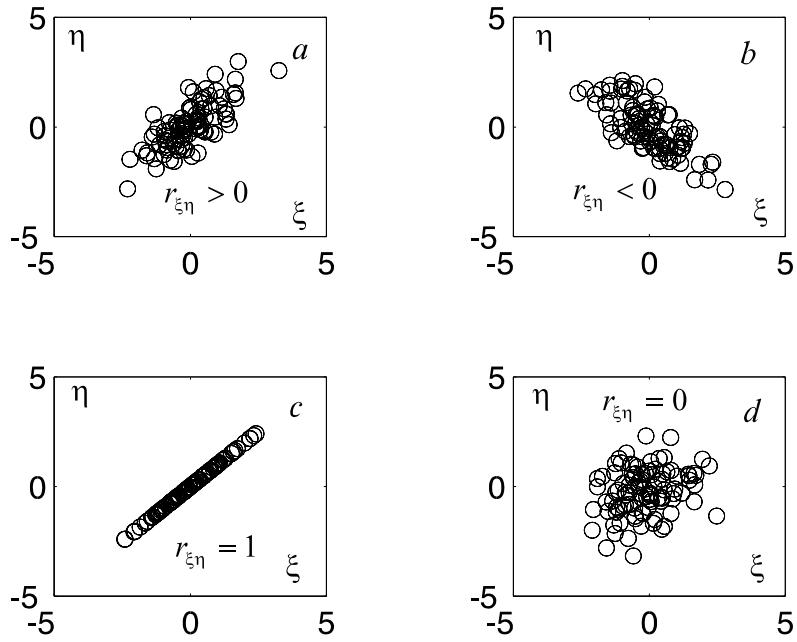


Figure 1.19: Graphic presentation of the samples from 2-D Gaussian distribution of the random variables ξ and η . $M\xi = M\eta = 0$; $r_{\xi\eta} = 0.75$ – (a); $r_{\xi\eta} = -0.75$ – (b); $r_{\xi\eta} = 1$ – (c); $r_{\xi\eta} = 0$ – (d). Sample size is 100.

$$= M \left(\sum_{i=1}^n a_i (\xi_i - M\xi_i)^2 \right) = \sum_{i=1}^n a_i^2 D\xi_i + 2 \sum_{i < j} a_i a_j R_{\xi_i \xi_j}.$$

If r.v. ξ_i and ξ_j are uncorrelated, then $D \left(\sum_{i=1}^n a_i \xi_i + a_0 \right) = \sum_{i=1}^n a_i^2 D\xi_i$.

- Let ξ and η be independent r.v. $D(\xi\eta) = D\xi D\eta + (M\xi)^2 D\eta + (M\eta)^2 D\xi$,
 $D(\xi\eta) = M(\xi\eta - M(\xi\eta))^2 = M(\xi\eta - M\xi M\eta)^2 =$
 $= M(\xi^2\eta^2) - 2M(\xi\eta)M\xi M\eta + (M\xi)^2(M\eta)^2 = M\xi^2 M\eta^2 - (M\xi)^2(M\eta)^2 =$
 $= (D\xi + (M\xi)^2)(D\eta + (M\eta)^2) - (M\xi)^2(M\eta)^2 = D\xi D\eta + (M\xi)^2 D\eta + (M\eta)^2 D\xi$.

1.3.4 Quantiles

The value x_p defined by

$$P(\xi < x_p) = F_\xi(x_p) = p, \quad 0 < p < 1$$

is called the *lower p-percentage point* or *p-th quantile* of the distribution (see Figure 1.20). Correspondingly, the *upper p-percentage point* is defined by

$$F_\xi(x_{1-p}) = 1 - p = 1 - F_x(x_p).$$

Quantity $x_{1/2}$ is called a *median* (see Figure 1.21). Quantities $x_{1/4}$, $x_{1/2}$, $x_{3/4}$ are called *quartiles* and $x_{0.1}$, $x_{0.2}$, ..., $x_{0.9}$ are called *deciles* (see Figure 1.22).

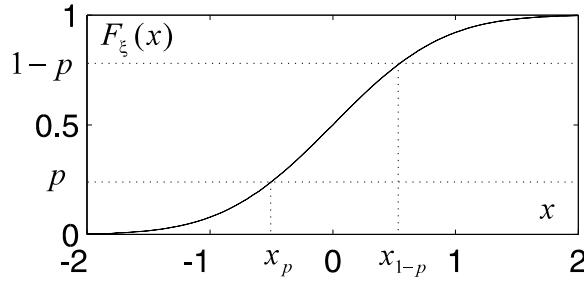


Figure 1.20: The distribution function and p -percentage point. Gaussian distribution: $m_\xi = 0$, $\sigma_\xi = 1$.

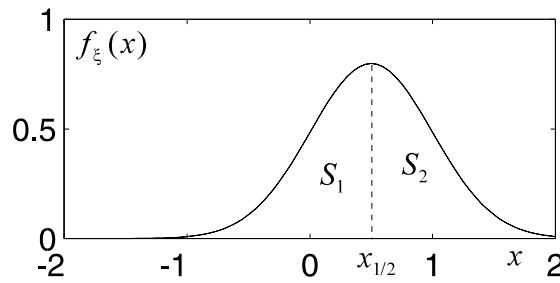


Figure 1.21: The density function and median. Gaussian distribution: $m_\xi = 0.5$, $\sigma_\xi = 0.5$.

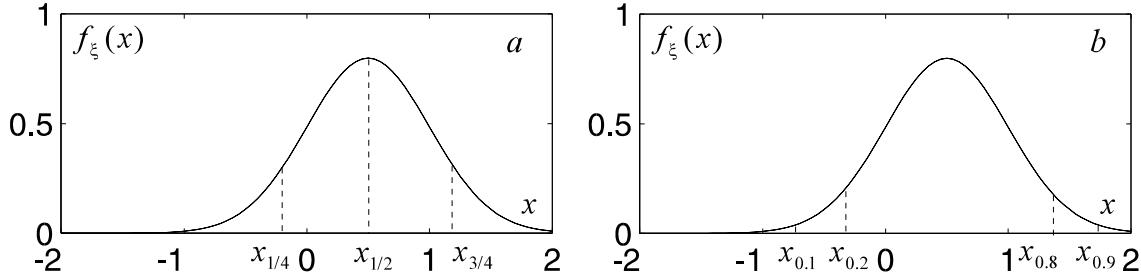


Figure 1.22: The density function and its characteristics: quartiles (a) and deciles (b). Gaussian distribution: $m_\xi = 0.5$, $\sigma_\xi = 0.5$.

1.3.5 Characteristics of a distribution density function

Quantity x_{max} is called a maximum of the density function (see Figure 1.23).

The ratio $\sigma_\xi/M\xi$ is called a *variation coefficient*. The value $M|\xi - M\xi|$ is called a *mean absolute deviation*. The difference of the quartiles $x_{3/4} - x_{1/4}$ is called an *interquartile width* (see Figure 1.24).

The modulus of the difference $|x_{max} - x_{min}|$ is called a range of the probability series $P_\xi(x_i)$ (see Figure 1.25).

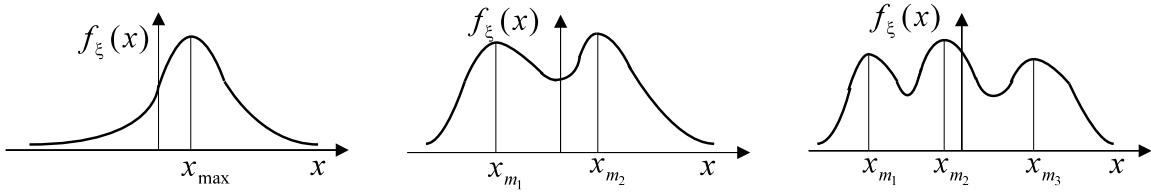


Figure 1.23: The density functions with the diffrent numbers of the extremum points.

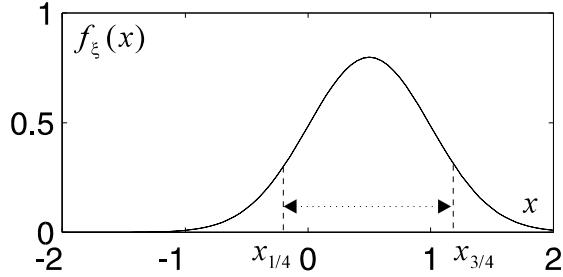


Figure 1.24: A density function and interquartile width. Gaussian distribution: $m_\xi = 0.5$, $\sigma_\xi = 0.5$.

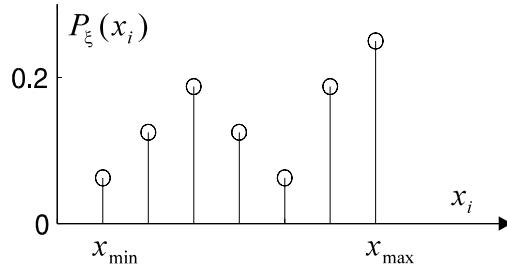


Figure 1.25: Probability series and its range.

Moments

The expected value of ξ determines only the center of gravity of its density. A more complete specification of the statistics of ξ is possible if one knows its *moments* α_s defined by

$$\alpha_s = M\xi^s = \int_{-\infty}^{\infty} x^s f(x) dx.$$

It is clearly, that $\alpha_0 = 1$, $\alpha_1 = M\xi = m_\xi$. The constants

$$\mu_s = M(\xi - M\xi)^s = \int_{-\infty}^{\infty} (x - M\xi)^s f(x) dx$$

are called *central moments*. We have

$$\mu_0 = 1, \quad \mu_1 = 0, \quad \mu_2 = D\xi = \sigma^2 = \alpha_2 - m_\xi^2, \quad \mu_3 = \alpha_3 - 3m_\xi\alpha_2 + 2m_\xi^2.$$

The constant

$$\gamma_1 = \mu_3/\sigma^3$$

is called *asymmetry coefficient* (see Figure 1.26). If the density function is symmetric,

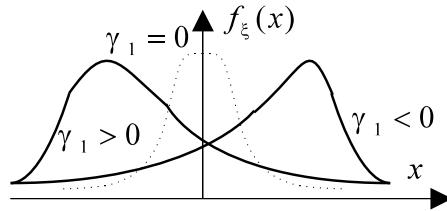


Figure 1.26: A density function and asymmetry coefficient.

then the coefficient γ_1 is equal to zero. The constant

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

is called *excess* (see Figure 1.27). The excess is a measure of the deviation of the density function $f_\xi(x)$ from the normal distribution. If the density function is a normal one, then the excess is equal to zero.

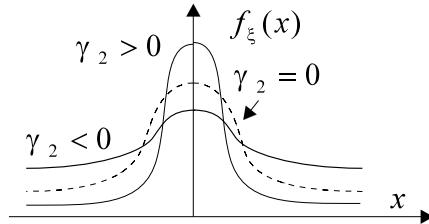


Figure 1.27: A density function and axcess.

1.4 Characteristic and generating functions

1.4.1 Characteristic functions

The *characteristic function* $g(t)$ of a r.v. ξ is the Fourier transform (with the reversal in sign) of its density function $f_\xi(x)$.

Definition: The *characteristic function* of a r.v. ξ is defined by

$$g(t) = M(\exp(it\xi)).$$

If ξ is of continuous type, then

$$g(t) = \int_{-\infty}^{\infty} \exp(itx) f_{\xi}(x) dx$$

and if ξ is of discrete type, taking the values x_k , then

$$g(t) = \sum_{k=1}^n \exp(itx_k) P_{\xi}(x_k).$$

The density $f_{\xi}(x)$ can be expressed in terms of $g(t)$ by the integral

$$f_{\xi}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) g(t) dt,$$

known as inverse Fourier transform or inversion formula.

Properties of characteristic functions

1. If ξ and η are random variables such that $\eta = a\xi$, then $g_{\eta}(t) = g_{\xi}(at)$.

Proof: $g_{\eta}(t) = M \exp(it\eta) = M \exp(ita\xi) = M \exp(i(at)\xi) = g_{\xi}(at)$.

2. Let $\xi_1, \xi_2, \dots, \xi_n$ are independent random variables with the characteristic functions $g_{\xi_1}(t), g_{\xi_2}(t), \dots, g_{\xi_n}(t)$ and $\eta = \sum_{k=1}^n \xi_k$, then $g_{\eta}(t) = \prod_{k=1}^n g_{\xi_k}(t)$.

Proof: $g_{\eta}(t) = M(\exp(it\eta)) = M(\exp(it \sum_{k=1}^n \xi_k)) = M \left[\prod_{k=1}^n \exp(it\xi_k) \right] = \prod_{k=1}^n g_{\xi_k}(t)$.

3. If ξ and η are independent random variables with density functions $f_{\xi}(x)$ and $f_{\eta}(y)$ respectively, then the density function and characteristic function of r.v. $\zeta = \xi + \eta$ are equal to

$$f_{\zeta}(z) = f_{\xi}(x) * f_{\eta}(y)$$

and (using Fourier transform)

$$g_{\zeta}(t) = g_{\xi}(t) \cdot g_{\eta}(t),$$

where

$$f_{\zeta}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g_{\zeta}(t) \exp(-itz) dt.$$

1.4.2 Moment generating function

Moments of random variables and the relation between their distributions are frequently derived by the *moment generating function*. Let ξ be a r.v. with the density $f_\xi(x)$. Then $\alpha_\xi(t)$ with

$$\alpha_\xi(t) = M(\exp(\xi t)) = \int_{-\infty}^{\infty} \exp(xt) f_\xi(x) dx$$

is called the *moment generating function* of the r.v. ξ .

The moments of a r.v. ξ is obtained by the moment generating function as follows. We have, expanding the exponential,

$$\alpha_\xi(t) = M\left[1 + \xi t + \frac{(\xi t)^2}{2!} + \frac{(\xi t)^3}{3!} + \dots\right] = 1 + \alpha_1 t + \alpha_2 \frac{t^2}{2!} + \alpha_3 \frac{t^3}{3!} + \dots .$$

The derivatives of the function $\alpha_s(t)$ are related to its moments α_s . By setting $t = 0$ we obtain

$$\frac{\partial^s \alpha_\xi(t)}{\partial t^s} = \int_{-\infty}^{\infty} x^s \exp(xt) f_\xi(x) dx, \quad \alpha_s = \left. \frac{\partial^s \alpha_\xi(t)}{\partial t^s} \right|_{t=0} = \int_{-\infty}^{\infty} x^s f_\xi(x) dx.$$

For the central moments, we have

$$\begin{aligned} \mu_\xi(t) &= M(\exp\{(\xi - m_\xi)t\}), \\ \mu_\xi(t) &= M\left(1 + (\xi - m_\xi)t + (\xi - m_\xi)^2 \frac{t^2}{2!} + \dots\right) = \\ &= 1 + \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \dots \end{aligned}$$

and

$$\mu_s = \left. \frac{\partial^s \mu_\xi(t)}{\partial t^s} \right|_{t=0} = \int_{-\infty}^{\infty} (x - m_\xi)^s f_\xi(x) dx.$$

1.4.3 Semi-invariants or cumulants

Let ξ be a r.v. with the moment generating function $\alpha_\xi(t)$. Taking the logarithm of the $\alpha_\xi(t)$ and expanding its, we have

$$\log \alpha_\xi(t) = \kappa_1 t + \kappa_2 \frac{t^2}{2!} + \kappa_3 \frac{t^3}{3!} + \dots,$$

where constants

$$\kappa_s = \left. \frac{\partial^s}{\partial t^s} \log \alpha_\xi(t) \right|_{t=0}$$

are called *semi-invariants* or *cumulants* ($\kappa_1 = m_\xi$, $\kappa_2 = D\xi$, $\kappa_3 = \mu_3$, $\kappa_4 = \mu_4 - 3\mu_2^2$).

1.5 Limit theorems

1.5.1 Convergence in probability

Consider now the sequence of r.v. $\xi_1, \xi_2, \dots, \xi_n \dots$. We form the probability

$$P(|\xi_n - a| < \varepsilon) > 1 - \delta \quad \text{for } n \rightarrow \infty,$$

that $|\xi_n - a|$ is less than the given number $\varepsilon > 0$. If it converges to one for every $\varepsilon > 0$ and $\delta > 0$, then we say that the sequence ξ_n tends to a in probability.

1.5.2 Chebyshev inequality

Suppose now that ξ is arbitrary r.v. with density $f_\xi(x)$, finite variance $D\xi$ and mean value m_ξ , then the inequality

$$P(|\xi - m_\xi| \geq \alpha) \leq D\xi / \alpha^2$$

is realized.

We have

$$P(|\xi - m_\xi| \geq \alpha) = \int_{|\xi - m_\xi| \geq \alpha} f_\xi(x) dx$$

and we can write down the variance $D\xi$ as

$$\begin{aligned} D\xi &= \int_{-\infty}^{\infty} (x - m_\xi)^2 f_\xi(x) dx = \int_{-\infty}^{\infty} |x - m_\xi|^2 f_\xi(x) dx \geq \\ &\geq \int_{|x - m_\xi| \geq \alpha} |x - m_\xi|^2 f_\xi(x) dx \geq \alpha^2 \int_{|x - m_\xi| \geq \alpha} f_\xi(x) dx = \alpha^2 P(|\xi - m_\xi| \geq \alpha). \end{aligned}$$

1.5.3 Chebyshev theorem

Let ξ be a r.v. with mean m_ξ and variance $D\xi$. It is produced n -trials $\xi_1, \xi_2, \dots, \xi_n$. Introduce a r.v. $\eta = \sum_{i=1}^n \xi_i / n$ (mean-arithmetic value)

$$m_\eta = M\eta = \frac{1}{n} \sum_{i=1}^n M\xi_i = \frac{1}{n} nm_\xi = m_\xi$$

and variance

$$D\eta = \frac{1}{n^2} \sum_{i=1}^n D\xi_i = \frac{D\xi}{n}.$$

We show that η converges to m_ξ in probability:

$$P\left(\left|\frac{\sum_{i=1}^n \xi_i}{n} - m_\xi\right| < \varepsilon\right) > 1 - \delta.$$

Proof: We write down Chebyshev inequality for $\alpha = \varepsilon$

$$P(|\eta - m_\eta| \geq \varepsilon) \leq \frac{D\eta}{\varepsilon^2} = \frac{D\xi}{n\varepsilon^2}.$$

Introducing δ such that

$$\frac{D\xi}{n\varepsilon^2} < \delta,$$

we have

$$P\left(\left|\frac{\sum_{i=1}^n \xi_i}{n} - m_\xi\right| \geq \varepsilon\right) \leq \delta$$

or

$$P\left(\left|\frac{\sum_{i=1}^n \xi_i}{n} - m_\xi\right| < \varepsilon\right) > 1 - \delta.$$

1.5.4 Central-limit theorem

Consider a sequence $\xi_1, \xi_2, \dots, \xi_n$ of independent r.v. with equal densities $f_\xi(x)$, mean m_ξ and variance σ_ξ^2 , then $\eta_n = \sum_{i=1}^n \xi_i$ has a normal distribution.

Proof: We assume that r.v. ξ_i are of continuous type. The characteristic function of a r.v. η is

$$g_\eta(t) = [g_\xi(t)]^n, \quad \text{where } g_\xi(t) = \int_{-\infty}^{\infty} \exp(itx) f_\xi(x) dx.$$

Expanding $g_\xi(t)$, we have

$$g_\xi(t) \approx g_\xi(0) + g'_\xi(0)t + [g''_\xi(0)/2 + \alpha(t)]t^2,$$

where $\alpha(t) \rightarrow 0$ for $t \rightarrow 0$, $g_\xi(0) = \int_{-\infty}^{\infty} f_\xi(x) dx = 1$,

$$g'_\xi(0) = \left[\int_{-\infty}^{\infty} ix \exp(itx) f_\xi(x) dx \right]_{t=0} = i \int_{-\infty}^{\infty} xf_\xi(x) dx = im_\xi.$$

We assume that $m_\xi = 0$, then $g'_\xi(0) = 0$, and

$$\begin{aligned} g''_\xi(0) &= - \int_{-\infty}^{\infty} x^2 f_\xi(x) dx = -\sigma^2, \\ g_\xi(t) &= 1 - \left[\frac{\sigma_\xi^2}{2} - \alpha(t) \right] t^2. \end{aligned}$$

We introduce $\zeta_n = \eta_n / \sigma \sqrt{n}$, then we have

$$g_\zeta(t) = g_\eta\left(\frac{t}{\sigma_\xi \sqrt{n}}\right) = \left[g_\xi\left(\frac{t}{\sigma_\xi \sqrt{n}}\right)\right]^n$$

and

$$g_\zeta(t) = \left\{1 - \left[\frac{\sigma_\xi^2}{2} - \alpha\left(\frac{t}{\sigma_\xi \sqrt{n}}\right)\right] \frac{t^2}{\sigma_\xi^2 n}\right\}^n.$$

Denoting

$$\beta = \left[\frac{\sigma_\xi^2}{2} - \alpha\left(\frac{t}{\sigma_\xi \sqrt{n}}\right)\right] \frac{t^2}{\sigma_\xi^2 n}$$

and taking the logarithm of the function $g_\zeta(t)$, we have

$$\ln g_\zeta(t) = n \ln(1 - \beta).$$

Expanding

$$\ln(1 - \beta) = -\left[\beta + \frac{\beta^2}{2!} + \frac{\beta^3}{3!} + \dots\right]$$

for $n \rightarrow \infty$, we can write

$$\ln(1 - \beta) \approx -\beta.$$

Thus we obtain

$$\lim_{n \rightarrow \infty} \ln g_{\zeta_n}(t) = \lim_{n \rightarrow \infty} n(-\beta) = \lim_{n \rightarrow \infty} \left\{-\frac{t^2}{2} + \alpha\left(\frac{t}{\sigma_\xi \sqrt{n}}\right) \frac{t^2}{\sigma_\xi^2}\right\}.$$

By definition $\alpha(t) \rightarrow 0$ for $t \rightarrow 0$, hence

$$\lim_{n \rightarrow \infty} \alpha\left(\frac{t}{\sigma_\xi \sqrt{n}}\right) = 0,$$

and

$$\lim_{n \rightarrow \infty} \ln g_{\zeta_n}(t) = -\frac{t^2}{2},$$

hence

$$\lim_{n \rightarrow \infty} g_{\zeta_n}(t) = \exp\left(-\frac{t^2}{2}\right)$$

and

$$f_\xi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) \exp(-itx) dt = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

It is a normal distribution with $m_\zeta = 0$ and $\sigma_\zeta = 1$.

1.6 Discrete distribution functions

Let ξ be discrete random variable. It takes on a countable number of values x_1, x_2, \dots, x_n with probabilities p_1, p_2, \dots, p_n ($\sum_{i=1}^n p_i = 1$, $P(\xi = x_i) = p_i$) (see Figure 1.28 and Table 1.1).

ξ	x_1	x_2	\cdots	x_n
P	p_1	p_2	\cdots	p_n

Table 1.1: A table presentation of the probabilities for a discrete random variable.

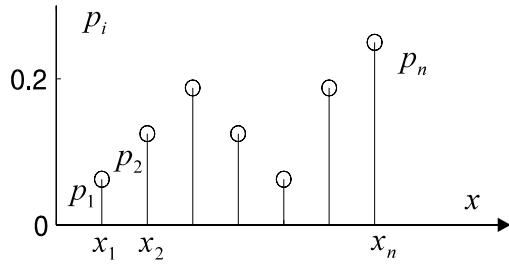


Figure 1.28: Graphic presentation of the probabilities of a discrete random variable.

1.6.1 Binomial distribution

Consider the experiment, which results either in a success or a failure. Let p be the probability of success. If this experiment is repeated in n independent trials, then the *binomial distribution* is valid and it gives the probability that m successes occur.

Definition: The random variable ξ has the *binomial distribution*, if its probability density function is given by

$$P_m = P(\xi = m) = C_n^m p^m q^{n-m}, \quad 0 < p < 1, \quad q = 1 - p, \quad m \in \{0, 1, \dots, n\}$$

(see Figure 1.29).

The first two moments of the distribution are

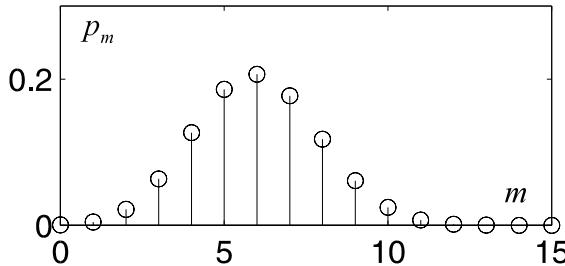
$$M\xi = np, \quad D\xi = np(1 - p).$$

Asymmetry coefficient and excess correspondingly read as

$$\gamma_1 = \frac{1 - 2p}{[np(1 - p)]^{1/2}} \quad \text{and} \quad \gamma_2 = \frac{1 - 6p(1 - p)}{np(1 - p)}.$$

Example: The probability of random event A equals p in one experiment. How many experiments n are necessary to produce, that the probability of the occurrence even one event will be α

$$\begin{aligned} P(\xi \geq 1) &\geq \alpha, \\ 1 - P(\xi = 0) &\geq \alpha, \quad P(\xi = 0) \leq 1 - \alpha, \\ m = 0, \quad (1 - p)^n &\leq 1 - \alpha, \quad n \geq \ln(1 - \alpha) / \ln(1 - p). \end{aligned}$$

Figure 1.29: Binomial distribution ($p = 0.4$, $n = 15$).

1.6.2 Poisson distribution

The discrete random variable $\xi \{0, 1, \dots, m, \dots\}$ has the *Poisson distribution* with the real-valued parameter a , if the probability density function is given by

$$p_m = P(\xi = m) = \frac{a^m}{m!} e^{-a}, \quad a > 0$$

(see Figure 1.30). To derive the Poisson distribution from the binomial distribution, we substitute $p = a/n$, thus $p \rightarrow 0$ with $n \rightarrow \infty$. The numerical characteristics of the Poisson distribution are

$$M\xi = a, \quad D\xi = a, \quad \gamma_1 = 1/\sqrt{a}, \quad \gamma_2 = 1/a.$$

If the random events from the Poisson flux, then number of events ξ occur on an arbitrary time interval $(t_0, t_0 + \tau)$ has the Poisson distribution

$$p_m = \frac{a^m}{m!} e^{-a},$$

where a is a mathematical expectation of the events number. If $\lambda(t)$ is a flux intensity, then a can be represented as

$$a = \int_{t_0}^{t_0+\tau} \lambda(t) dt.$$

Poisson flux well describes the arrival-times of seismic waves on seismograms.

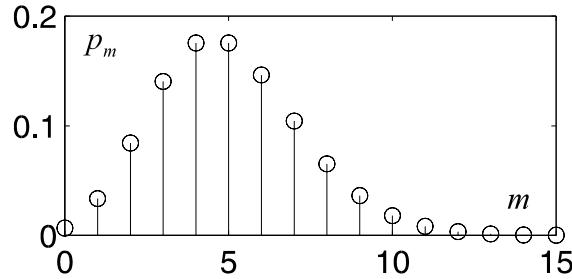
Example: It is registered a sequence of reflected waves from the stratified medium.

The waves are described by the stationary Poisson flux with the intensity λ .

We shall find that during the time τ :

- a) it will be not registered seismic wave (event A);
- b) it will be registered not less than 3 waves (event B);
- c) it will be registered 3 waves (event C).

$$\begin{aligned} P(A) &= p_0 = e^{-\lambda\tau}, \quad P(C) = \frac{(\lambda\tau)^3}{3!} e^{-\lambda\tau}, \\ P(B) &= 1 - (p_0 + p_1 + p_2) = 1 - e^{-\lambda\tau}[1 - \lambda\tau - 0.5(\lambda\tau)^2]. \end{aligned}$$

Figure 1.30: Poisson distribution ($a = 5$).

1.6.3 Geometrical distribution

The discrete random values $\xi \{0, 1, 2, \dots, m, \dots\}$ has the *geometrical distribution*, if its probability is given by

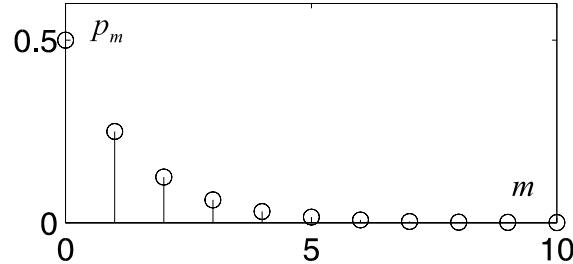
$$p_m = P(\xi = m) = q^m p, \quad 0 < p < 1, \quad q = 1 - p,$$

where p is the probability of success and q is the probability of failure (see Figure 1.31).

The geometrical distribution gives the probability of m failure trials before first success trial.

The numerical characteristics of the geometrical distribution are

$$M\xi = (1-p)/p, \quad D\xi = (1-p)/p^2, \quad \gamma_1 = (2-p)/(1-p)^{1/2}, \quad \gamma_2 = (p^2-6p+6)/(1-p).$$

Figure 1.31: Geometrical distribution ($p = 0.5$).

1.7 Continuous distribution functions

Let ξ be a continuous random variable. The probability of the event $\xi < x$ is described by the distribution function $F_\xi(x)$ and probability density function $f_\xi(x)$

$$P(\xi < x) = F_\xi(x) = \int_{-\infty}^x f_\xi(t)dt, \quad f_\xi(x) \geq 0, \quad \int_{-\infty}^{\infty} f_\xi(x)dx = 1.$$

The probability of the event $\alpha < \xi \leq \beta$ is computed by

$$P(\alpha \leq \xi < \beta) = F_\xi(\beta) - F_\xi(\alpha).$$

1.7.1 Univariate normal distribution

The univariate normal distribution is the most important distribution of the continuous random variables.

Definition: The random variable ξ is said to be *normally distributed* with the parameters m_ξ and σ_ξ^2 , which is written $\xi \in N(m_\xi, \sigma_\xi^2)$, if its density function $f_\xi(x)$ is given by

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma_\xi} \exp\left\{-\frac{(x-m_\xi)^2}{2\sigma_\xi^2}\right\} \quad \text{for } -\infty < x < \infty.$$

The normal distribution function is given by

$$F_\xi(x) = \Phi\left(\frac{x-m_\xi}{\sigma_\xi}\right),$$

where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left\{-\frac{1}{2}y^2\right\} dy$$

is the Laplace function (see Figure 1.32).

Properties of the Laplace function.

1. $\Phi(-\infty) = (1/\sqrt{2\pi}) \int_{-\infty}^{-\infty} \exp\{-\frac{1}{2}y^2\} dy = 0$.
2. $\Phi(-x) = 1 - \Phi(x)$.
3. $\Phi(\infty) = (1/\sqrt{2\pi}) \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}y^2\} dy = 1$.

Characteristic function is

$$g(t) = \exp\{itm_\xi - t^2\sigma_\xi^2/2\}.$$

The asymmetry coefficient and excess are equal to zero $\gamma_1 = \gamma_2 = 0$. The moments are

$$\mu_{2s} = \frac{(2s)!}{2^s s!} \sigma_\xi^{2s}, \quad \mu_{2s+1} = 0, \quad s \geq 1.$$

Example: For the random variable ξ with $\xi \in N(m_\xi, \sigma_\xi^2)$ the probabilities $P = 0.9$, $P = 0.95$ and $P = 0.99$ shall be computed:

$$\begin{aligned} P(-1.64 < \frac{\xi - m_\xi}{\sigma_\xi} < 1.64) &= 0.90, \\ P(-1.96 < \frac{\xi - m_\xi}{\sigma_\xi} < 1.96) &= 0.95, \\ P(-2.58 < \frac{\xi - m_\xi}{\sigma_\xi} < 2.58) &= 0.99. \end{aligned}$$

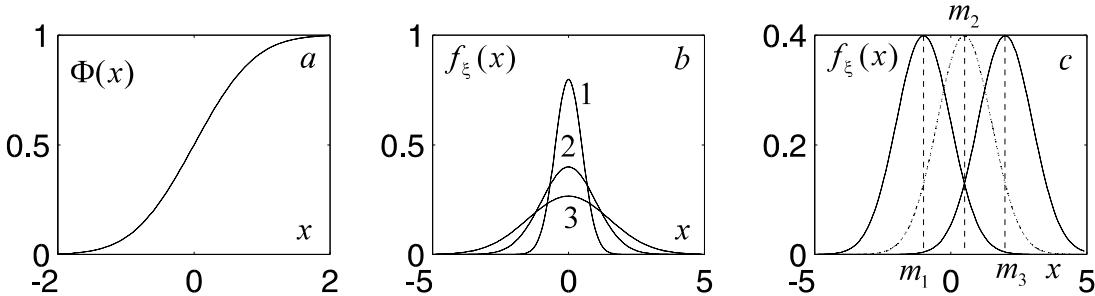


Figure 1.32: The Laplace function and the normal density function. (a) – the Laplace function: $m = 0$, $\sigma = 0.5$. (b) – the normal density functions: $m_1 = m_2 = m_3 = 0$; $\sigma_1 = 0.5$ – (1); $\sigma_2 = 1.0$ – (2); $\sigma_3 = 1.5$ – (3). (c) the normal density functions: $m_1 = -1$; $m_2 = 0.5$; $m_3 = 2$; $\sigma_1 = \sigma_2 = \sigma_3 = 1$.

We define a *standard normal density function* ($N(0, 1)$):

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}t^2\right\},$$

a *standard normal distribution function* ($\Phi(x)$):

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{1}{2}t^2\right\} dt.$$

If the random variables $\xi_1, \xi_2, \dots, \xi_n$ are normal and independent, then the arbitrary linear combination $\eta = \sum_{i=1}^n a_i \xi_i$ of ξ_i is normal.

Suppose, that ξ_1 and ξ_2 are normal and independent r.v. with the means m_1 and m_2 , variances σ_1^2 and σ_2^2 respectively, then $a_1 \xi_1$ and $a_2 \xi_2$ are normal distributed with the characteristic functions

$$g_{a_1 \xi_1}(t) = \exp(it a_1 m_1 - t^2 a_1^2 \sigma_1^2 / 2), \quad g_{a_2 \xi_2}(t) = \exp(it a_2 m_2 - t^2 a_2^2 \sigma_2^2 / 2).$$

The characteristic function of r.v. $\eta = a_1 \xi_1 + a_2 \xi_2$ equals

$$g_\eta(t) = g_{a_1 \xi_1}(t) g_{a_2 \xi_2}(t) = \exp[it(a_1 m_1 + a_2 m_2) - t^2(a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)/2],$$

i.e. r.v. η belongs to the normal distribution with the mean $m_\eta = a_1 m_1 + a_2 m_2$ and the variance $\sigma_\eta^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2$.

Analogously, it is possible to show that

$$\eta = \sum_{i=1}^n a_i \xi_i$$

will have the normal distribution with the mean

$$m_\eta = \sum_{i=1}^n a_i m_i$$

and variance

$$\sigma_\eta^2 = \sum_{i=1}^n a_i^2 \sigma_i^2.$$

1.7.2 Multivariate normal distribution

The random vector $\vec{\xi} = [\xi_1, \xi_2, \dots, \xi_n]^T$ is said to have a multivariate normal distribution $N(\vec{m}_\xi, D_\xi)$ with the vector \vec{m}_ξ and $n \times n$ positive definite matrix D_ξ , if the density function $f_\xi(\vec{x})$ of \vec{x} is given by

$$f_\xi(\vec{x}) = (2\pi)^{-n/2} |D_\xi|^{-1/2} \exp \left\{ -(1/2)(\vec{x} - \vec{m}_\xi)^T D_\xi^{-1} (\vec{x} - \vec{m}_\xi) \right\},$$

where \vec{m}_ξ denotes the vector of the expected values of $\vec{\xi}$: $M\vec{\xi} = \vec{m}_\xi$ and D_ξ denotes the covariance matrix of $\vec{\xi}$: $[D_\xi]_{ij} = M[(\xi_i - m_{\xi i})(\xi_j - m_{\xi j})]$. The normal characteristic function is given by

$$g_\xi(\vec{t}) = \exp[i\vec{t}^T \vec{m}_\xi - (1/2)\vec{t}^T D_\xi \vec{t}].$$

Example: Two r.v. ξ_1 and ξ_2 are said to be jointly normal if their density is of the form

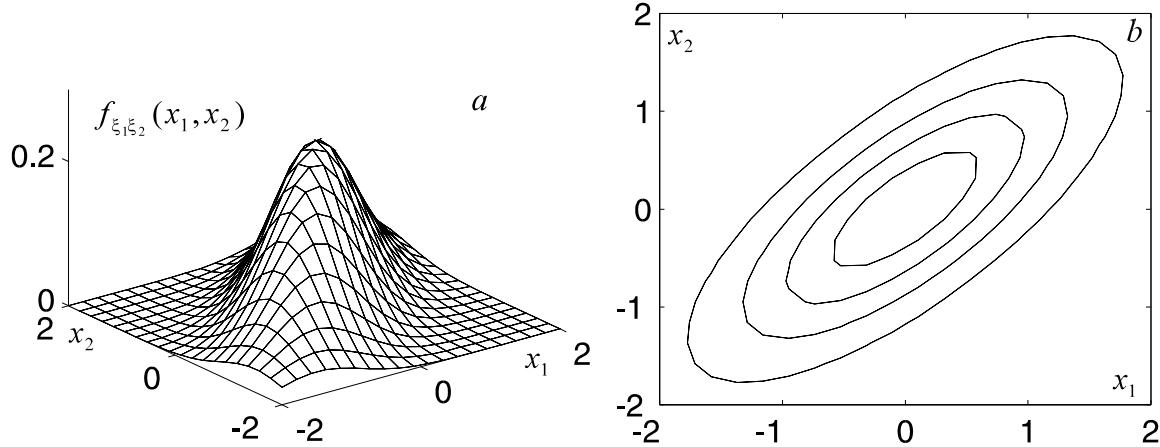


Figure 1.33: Gaussian density function for two random variables ξ_1 and ξ_2 : $m_1 = m_2 = 0$; $\sigma_1 = \sigma_2 = 1$; $r = 0.75$. (a) – surface function; (b) – its isolines at the (x_1, x_2) plane.

$$\begin{aligned} f_{\xi_1 \xi_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[\frac{(x_1 - m_1)^2}{\sigma_1^2} - \right. \right. \\ &\quad \left. \left. - 2r \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} \right] \right\}, \end{aligned}$$

(see Figure 1.33) with covariance matrix

$$D = \begin{vmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix}, \quad r = \frac{M[(\xi_1 - m_1)(\xi_2 - m_2)]}{\sigma_1\sigma_2}.$$

The marginal densities of ξ_1 and ξ_2 are normal and are given by

$$\begin{aligned} f_{\xi_1}(x_1) &= \int_{-\infty}^{\infty} f_{\xi_1 \xi_2}(x_1, x_2) dx_2, \quad f_{\xi_2}(x_2) = \int_{-\infty}^{\infty} f_{\xi_1 \xi_2}(x_1, x_2) dx_1, \\ f_{\xi_1}(x_1) &= \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left\{-\frac{(x_1 - m_1)^2}{2\sigma_1^2}\right\}, \\ f_{\xi_2}(x_2) &= \frac{1}{\sigma_2 \sqrt{2\pi}} \exp\left\{-\frac{(x_2 - m_2)^2}{2\sigma_2^2}\right\}. \end{aligned}$$

The conditional density of ξ_2 , assuming $\xi_1 = x_1$ can be given by

$$\begin{aligned} f_{\xi_2/\xi_1}(x_2/x_1) &= \frac{f_{\xi_1 \xi_2}(x_1, x_2)}{f_{\xi_1}(x_1)} = \\ &= \frac{1}{\sqrt{2\pi} \sigma_2 \sqrt{1-r^2}} \exp\left\{\frac{-1}{2(1-r^2)} \left[\frac{(x_2 - m_2)}{\sigma_2} - r \frac{(x_1 - m_1)}{\sigma_1}\right]^2\right\} = \\ &= \frac{1}{\sqrt{2\pi} \sigma_2 \sqrt{1-r^2}} \exp\left\{\frac{-1}{2(1-r^2)\sigma_2^2} \left[(x_2 - m_2) - r \frac{\sigma_2}{\sigma_1} (x_1 - m_1)\right]^2\right\}. \end{aligned}$$

The conditional density of ξ_1 , assuming $\xi_2 = x_2$ can be given by

$$\begin{aligned} f_{\xi_1/\xi_2}(x_1/x_2) &= \frac{1}{\sqrt{2\pi} \sigma_1 \sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)\sigma_1^2} \left[(x_1 - m_1) - r \frac{\sigma_1}{\sigma_2} (x_2 - m_2)\right]^2\right\}. \end{aligned}$$

These conditional densities are normal with mean and variance given by

$$\begin{aligned} M[\xi_2/\xi_1 = x_1] &= m_{\xi_2/\xi_1} = m_2 + r \frac{\sigma_2}{\sigma_1} (x_1 - m_1), \quad \sigma_{\xi_2/\xi_1} = \sigma_2 \sqrt{1-r^2}, \\ M[\xi_1/\xi_2 = x_2] &= m_{\xi_1/\xi_2} = m_1 + r \frac{\sigma_1}{\sigma_2} (x_2 - m_2), \quad \sigma_{\xi_1/\xi_2} = \sigma_1 \sqrt{1-r^2} \end{aligned}$$

(see Figure 1.34).

$$\begin{aligned} \frac{(x_1 - m_1)^2}{\sigma_1^2} - 2r \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} &= \lambda^2, \\ \tan 2\alpha &= \frac{2r\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}. \end{aligned}$$

1.7.3 Uniform distribution

If the density function of a r.v. ξ is a rectangular pulse

$$f_{\xi}(x) = \begin{cases} (b-a)^{-1} & x \in (a, b), \\ 0 & x \notin (a, b), \end{cases}$$

we say, that ξ is uniformly distributed in the interval (a, b) (see Figure 1.35) with the next characteristics

$$M\xi = (a+b)/2, \quad D\xi = (b-a)^2/12, \quad \gamma_1 = 0, \quad \gamma_2 = -1.2.$$

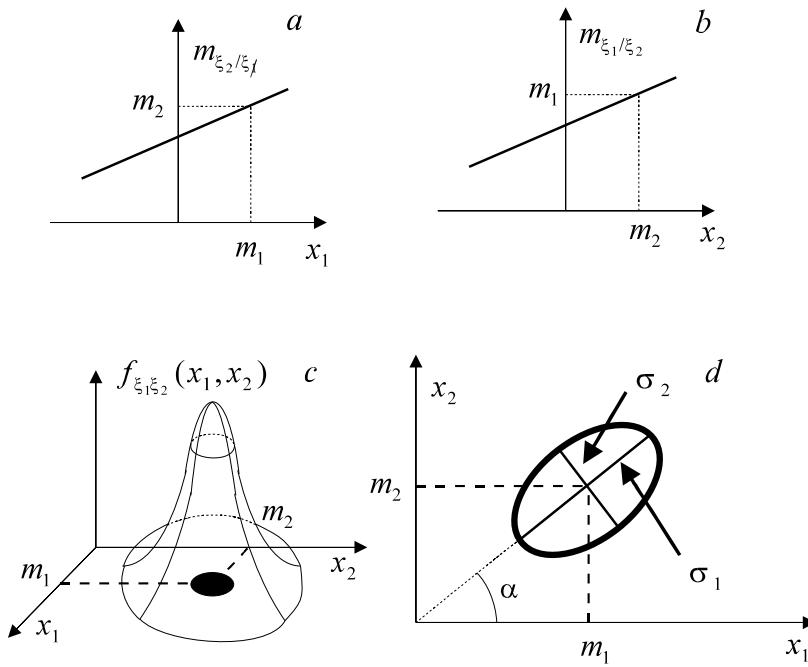


Figure 1.34: Two dimensional normal distribution. (a) and (b) – the conditional mathematical expectations of ξ_1 and ξ_2 correspondingly; (c) – graphic presentation of the density function; (d) – the contour of the density function at the (x_1, x_2) plane.

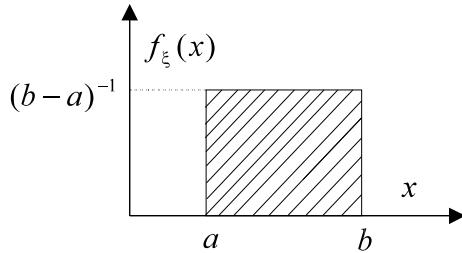


Figure 1.35: Uniform distribution.

1.7.4 χ^2 — distribution

Let the $n \times 1$ random vector $\vec{\xi} = [\xi_1, \dots, \xi_n]^T$ be normally distributed according to $\vec{\xi} = N(0, 1)$, then the sum of squares $\eta = \sum_{i=1}^n \xi_i^2$ is said to have χ_n^2 -distribution with n degrees of freedom, whose density function is given by

$$\eta = \sum_{i=1}^n \xi_i^2, \quad \xi_i \in N(0, 1), \quad \eta \in \chi_n^2,$$

$$f_\eta(x) = \begin{cases} \left[(x/2)^{(n/2)-1} \exp(-x/2) \right] / (2\Gamma(n/2)) & x > 0, \\ 0 & x \leq 0 \end{cases}$$

(see Figure 1.36).

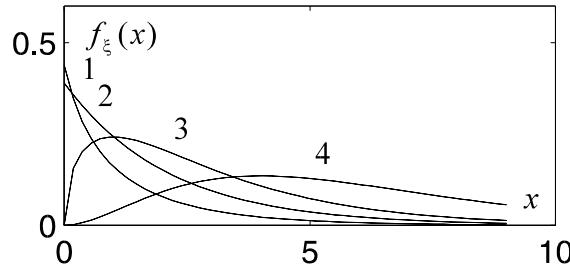


Figure 1.36: χ^2 -distribution. 1 – $n = 1$; 2 – $n = 2$; 3 – $n = 3$; 4 – $n = 6$.

The moments of the χ_n^2 -distribution are:

$$M\xi = n, \quad D\xi = 2n, \quad \gamma_1 = 2\sqrt{2/n}, \quad \gamma_2 = 12/n.$$

Characteristic function:

$$g(t) = (1 - 2it)^{-n/2}.$$

Let $\vec{\xi} = [\xi_1, \dots, \xi_n]^T$ be a random vector with normal distributed components $\xi_i \in N(m, \sigma^2)$. The sample mean is

$$\bar{\xi} = (1/n) \sum_{i=1}^n \xi_i.$$

The sample variance is

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2.$$

The random variable

$$\eta = \frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (\xi_i - \bar{\xi})^2 \in \chi_{n-1}^2$$

has the χ^2 -distribution.

1.7.5 Student's t-distribution

Let the random variable α and ζ be independently distributed according to $\alpha \in N(0, 1)$ and $\zeta \in \chi_n^2$, then the random variable η with

$$\eta = \frac{\alpha}{(\zeta/n)^{1/2}}$$

is said to have the t -distribution, also called *Student's t-distribution*, with n degrees of freedom with the density function

$$f_\eta(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

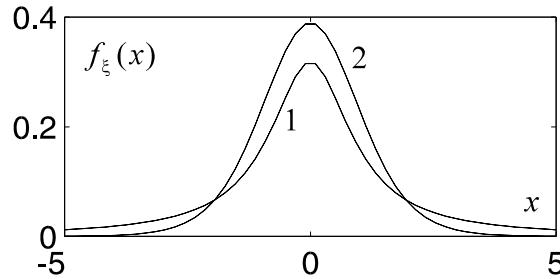


Figure 1.37: Student distribution. 1 — $n = 1$; 2 — $n = 10$.

(see Figure 1.37).

Let ξ_1, \dots, ξ_n be a random vector with normal distribution $\xi_i \in N(m, \sigma^2)$.

Sample mean:

$$\bar{\xi} = (1/n) \sum_{i=1}^n \xi_i.$$

The sample variance:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2.$$

We can write down

$$\eta = \sqrt{n}(\bar{\xi} - m_\xi)/s.$$

Moments:

$$\begin{aligned} M\eta &= 0, & D\eta &= n/(n-2), & n > 2, \\ \gamma_1 &= 0, & \gamma_2 &= 6/(n-4), & n > 4. \end{aligned}$$

1.7.6 Fisher distribution

Let r.v. $\xi_1, \xi_2, \dots, \xi_{n_1}$, and $\eta_1, \eta_2, \dots, \eta_{n_2}$ be independently distributed according to

$$\xi_i \in N(m_1, \sigma_1^2), \quad \eta_i \in N(m_2, \sigma_2^2).$$

Let the sample variance s_1 and s_2 of the r.v. $\vec{\xi}$ and $\vec{\eta}$ be independent

$$s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (\xi_i - \bar{\xi})^2, \quad s_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (\eta_i - \bar{\eta})^2.$$

Let the random variables ζ_1 and ζ_2 be independently distributed accordingly to

$$\zeta_1 = \frac{n_1 s_1^2}{\sigma_1^2} \in \chi_{n_1}^2, \quad \zeta_2 = \frac{n_2 s_2^2}{\sigma_2^2} \in \chi_{n_2}^2,$$

then the random variable

$$\zeta = \frac{\zeta_1/n_1}{\zeta_2/n_2} = \frac{s_1^2 \sigma_2^2}{\sigma_1^2 s_2^2} \in F(n_1, n_2)$$

is said to have a Fisher distribution $F(n_1, n_2)$ with n_1 and n_2 degrees of freedom, thus ζ with the density function

$$f_\zeta(F) = \frac{n_1^{n_2/2} n_2^{n_1/2} F(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2})} \frac{F^{(n_1/2)-1}}{(n_1 + n_2 F)^{(n_1+n_2)/2}}$$

(see Figure 1.38).

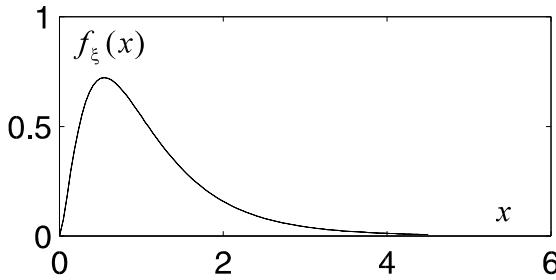


Figure 1.38: Fisher distribution ($n_1 = 5, n_2 = 50$).

The two moments of the Fisher distribution are

$$M\zeta = \frac{n_2}{n_2 - 2}, \quad n_2 > 2; \quad D\zeta = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}, \quad n_2 > 4.$$

The Fisher density $f_\zeta(F)$ tends to the normal distribution by large enough numbers n_1 and n_2 ($n_1, n_2 > 50$)

$$f_\zeta(F) \xrightarrow{n_1, n_2 \rightarrow \infty} N(m, \sigma).$$

1.7.7 Exponential distribution

The random variable has the exponential distribution density if

$$f_\xi(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{for } x \geq 0, \\ 0 & \text{for } x < 0 \end{cases}$$

where $\lambda > 0$ (see Figure 1.39).

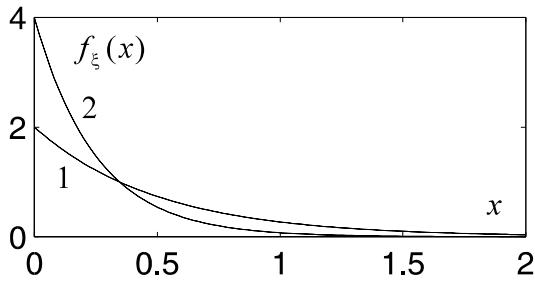
Moments:

$$M\xi = \lambda^{-1}, \quad D\xi = \lambda^{-2}, \quad \gamma_1 = 2, \quad \gamma_2 = 6.$$

Characteristic function:

$$g(t) = (1 - it/\lambda)^{-1}.$$

The exponential distribution plays a very important role in the theory of Markov's process. If it is the stationary Poisson flux with the intensity λ , then the time interval between neighboring events has the exponential distribution with the parameter λ .

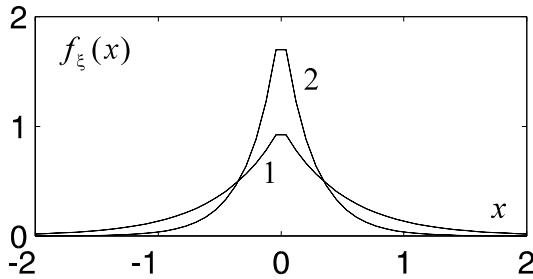
Figure 1.39: Exponential distribution. $1 - \lambda = 2$, $2 - \lambda = 4$.

1.7.8 Laplace distribution

The random variable ξ has the Laplace distribution, if the density is given by

$$f_\xi(x) = (\lambda/2) \exp(-\lambda|x|)$$

(see Figure 1.40).

Figure 1.40: Laplace distribution. $1 - \lambda = 2$, $2 - \lambda = 4$.

Moments:

$$M\xi = 0, \quad D\xi = 2/\lambda^2.$$

1.7.9 Cauchy distribution

The random variable ζ has the Cauchy distribution if the density is given by

$$f_\zeta(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

(see Figure 1.41).

We can obtain the Cauchy density as the ratio of the r.v. with the standard normal distribution

$$\zeta = \xi/\eta, \quad \xi \in N(0, 1), \quad \eta \in N(0, 1).$$

Characteristic function: $g(t) = \exp(-|t|)$.

The quantity $M\xi$ is not determined and other moments are nonconvergent.

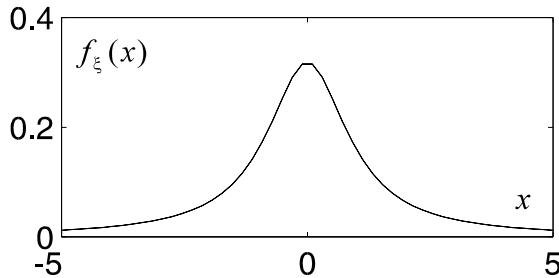


Figure 1.41: Cauchy distribution.

1.7.10 Logarithmic normal distribution

If the random variable $\eta = \ln \xi$, has the normal distribution, then the r.v. ξ is given by lognormal distribution

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{1}{2\sigma^2}(\ln x - m)^2\right\}$$

(see Figure 1.42).

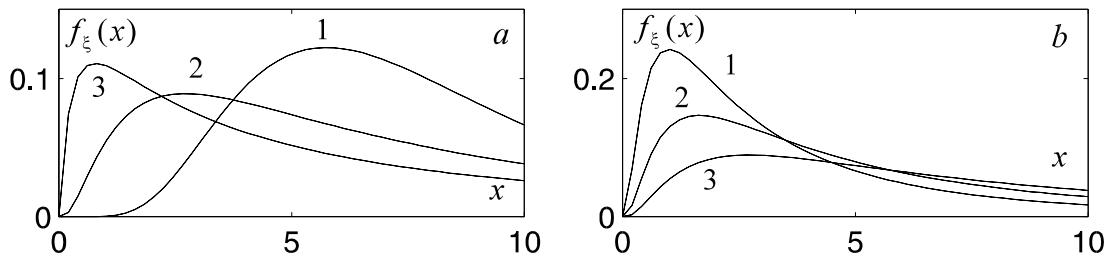


Figure 1.42: Logarithmic normal distribution. (a) $m = 2.0$, $1 - \sigma = 0.5$, $2 - \sigma = 1.0$, $3 - \sigma = 1.5$; (b) $\sigma = 1.0$, $1 - m = 1.0$, $2 - m = 1.5$, $3 - m = 2.0$.

Moments:

$$M\xi = \exp(m + \sigma^2/2), \quad D\xi = \exp(2m + \sigma^2)(\exp(\sigma^2) - 1).$$

1.7.11 Significance of the normal distribution

In case of n independent random variables with any distributions, the distribution of the sum of these random variables moves under certain, but very general conditions asymptotically towards a normal distribution if n approaches infinity. Relations between the normal distribution and some distributions mentioned above are represented by Figure 1.43 and Figure 1.44.

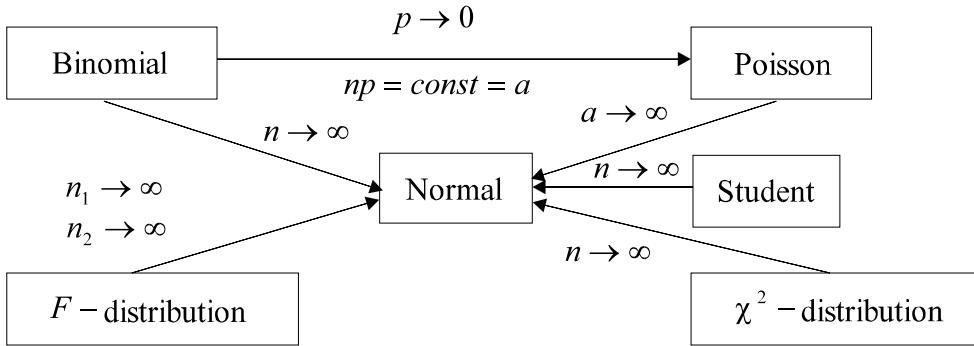


Figure 1.43: A scheme of relations between the normal distribution and some others distributions from the point of view of the limit values of the parameters of the distributions.

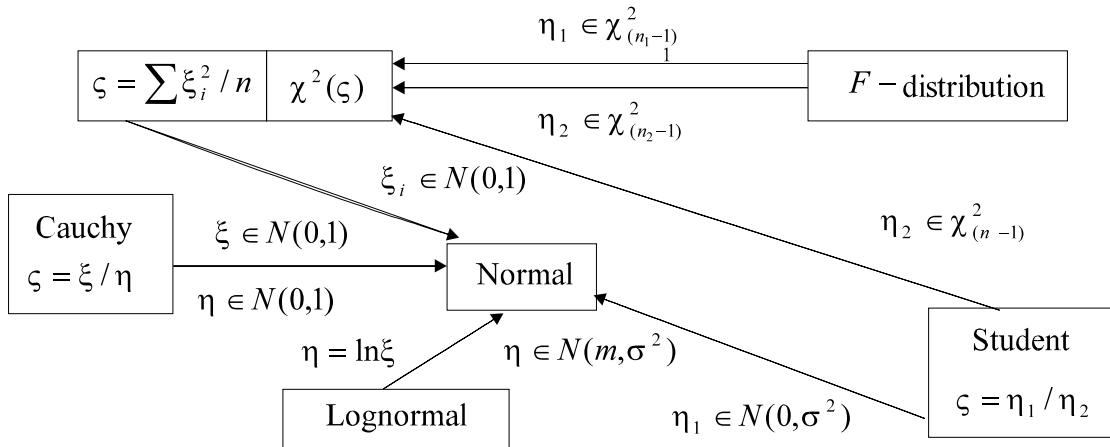


Figure 1.44: A scheme of relations between the normal distribution and some others distributions from the point of view of the types of the random variables which create these distributions.

1.7.12 Confidence interval

Let $\tilde{\rho}$ be the estimate of unknown parameter ρ , then the confidence interval I_β is given by

$$P(|\tilde{\rho} - \rho| < \varepsilon) = \beta, \quad P(\tilde{\rho} - \varepsilon < \rho < \tilde{\rho} + \varepsilon) = \beta, \quad I_\beta = (\tilde{\rho} - \varepsilon, \tilde{\rho} + \varepsilon),$$

where β is the confidence probability.

Confidence interval for the mathematical expectation

The estimation of the mathematical expectation is a sample mean

$$\tilde{\rho} = \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i.$$

1. The variance σ_ξ is known

$$\begin{aligned} \xi_1, \xi_2, \dots, \xi_n; \quad m_\xi, \quad \sigma_m = \sigma_\xi / \sqrt{n}, \\ P\{-l \leq (\bar{\xi} - m_\xi) \sqrt{n} / \sigma_\xi \leq l\} = 2\Phi(l) - 1, \\ P\{\bar{\xi} - l\sigma_\xi / \sqrt{n} \leq m_\xi \leq \bar{\xi} + l\sigma_\xi / \sqrt{n}\} = 2\Phi(l) - 1. \end{aligned}$$

2. The variance σ_ξ is unknown.

$$\begin{aligned} \xi_1, \xi_2, \dots, \xi_n; \quad s^2 = 1/(n-1) \sum_{i=1}^n (\xi_i - \bar{\xi})^2, \\ \eta = \sqrt{n}(\bar{\xi} - m_\xi)/s, \quad P\{|\eta| < t_\beta\} = \beta. \end{aligned}$$

The quantity $\eta = \sqrt{n}(\bar{\xi} - m_\xi)/s$ belongs to the Student distribution (see (1.7.5)) and we can find the value t_β as the quantity β is known (see Figure 1.45).

$$P\{\bar{\xi} - t_\beta s / \sqrt{n} \leq m_\xi \leq \bar{\xi} + t_\beta s / \sqrt{n}\} = \beta.$$

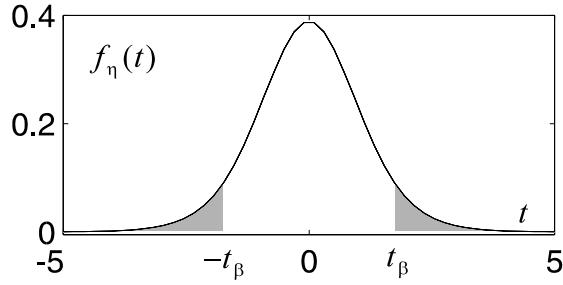


Figure 1.45: Student density function ($n=10$) and its 5.5-percentage density points.

Confidence interval for the variance

Let $\xi_1, \xi_2, \dots, \xi_n$ be random variable

$$\xi_i \in N(m_\xi, \sigma_\xi^2), \quad \frac{(n-1)s^2}{\sigma_\xi^2} \in \chi_{n-1}^2.$$

$$\begin{aligned} 1 - P(\chi^2 > \chi_1^2) = P(\chi^2 < \chi_1^2) = \alpha/2, \\ P(\chi^2 > \chi_2^2) = \alpha/2. \end{aligned}$$

So we can find the quantities χ_1^2 and χ_2^2 (see Figure 1.46)

$$P(\chi_1^2 < (n-1)s^2 / \sigma^2 < \chi_2^2) = 1 - P(\chi^2 < \chi_1^2) - P(\chi^2 > \chi_2^2) = 1 - \alpha,$$

$$P\left(\frac{(n-1)s^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_1^2}\right) = 1 - \alpha.$$

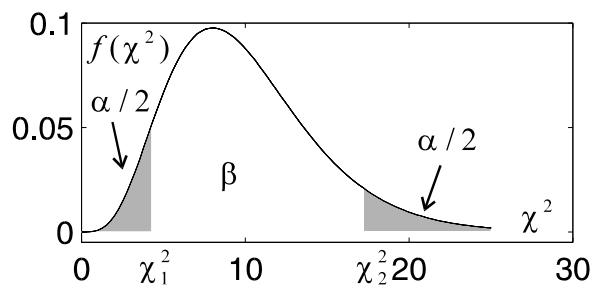


Figure 1.46: χ^2 -square density function. $\alpha=0.13$

2

Information and entropy

Entropy is a measure of the state of the indetermination of the physical system.

2.1 Entropy of the set of discrete states of system

Let x_1, x_2, \dots, x_n be possible sets of system with the probabilities p_1, p_2, \dots, p_n , then the entropy is given by

$$H = - \sum_{i=1}^n p_i \log p_i.$$

In the case of the equiprobable $p_i = 1/n$, we have

$$H = -n \frac{1}{n} \log \frac{1}{n} = \log n.$$

We can define the entropy of the system as

$$H_\xi = M[-\log p_\xi].$$

2.2 Entropy of the complex system

Let the random variables ξ and η have accordingly the realizations x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_m (their probability function can be presented in a table form – see Table 2.1). The joint entropy is given by

$$\begin{aligned} H_{\xi\eta} &= - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log p_{ij}, \\ H_{\xi\eta} &= M[-\log p(\xi, \eta)]. \end{aligned}$$

If the random variables ξ and η are independent, then the joint probability is

$$p(\xi, \eta) = p(\xi)p(\eta)$$

$y_j \setminus x_i$	x_1	x_2	\dots	x_{n-1}	x_n
y_1	p_{11}	p_{12}	\dots	$p_{1\,n-1}$	$p_{1\,n}$
y_2	p_{21}	p_{22}	\dots	$p_{2\,n-1}$	$p_{2\,n}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
y_m	p_{m1}	p_{m2}	\dots	$p_{m\,n-1}$	$p_{m\,n}$

Table 2.1: A table presentation of the joint probability function for two discrete random variables x_i and y_j .

and the entropy is

$$H_{\xi\eta} = H_\xi + H_\eta.$$

In the case of n random variables, the entropy is

$$H_{\xi_1, \xi_2, \dots, \xi_n} = H_{\xi_1} + H_{\xi_2} + \dots + H_{\xi_n}.$$

Let r.v. ξ and η are independent, then the conditional entropy of the system ξ provided that the η is in the state y_j :

$$H_{\xi/y_j} = - \sum_{i=1}^n p(x_i/y_j) \log p(x_i/y_j)$$

or

$$H_{\xi/y_j} = M[-\log p(\xi/y_j)].$$

The mean or perfect entropy of the system is

$$H_{\xi/\eta} = \sum_{j=1}^m p_j H_{\xi/y_j} = \sum_{i=1}^n \sum_{j=1}^m p_j p(x_i/y_j) \log p(x_i/y_j)$$

or taking into account that is

$$p_j p(x_i/y_j) = p_{ij},$$

we have

$$H_{\xi/\eta} = - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log p(x_i/y_j).$$

The entropy of the joint system (ξ, η) is determined as

$$H_{\xi\eta} = H_\xi + H_{\eta/\xi}$$

or

$$H_{\xi\eta} = H_\eta + H_{\xi/\eta}.$$

It is easily provided, starting from the entropy definition and the theorem of the probability product

$$p(\xi, \eta) = p(\xi)p(\eta/\xi) = p(\eta)p(\xi/\eta).$$

2.3 Shannon information (discrete case)

Amount of information is defined by the decrease of the indetermination of the system.

Let ξ be a random variable, describing some physical system. If the entropy before the observation (a priori) is denoted by $H_\xi^{(apr)}$ and the entropy after observation (a posteriori) is denoted by $H_\xi^{(apost)}$, then the Shannon information is determined as

$$I_\xi^{(s)} = H_\xi^{(apr)} - H_\xi^{(apost)}.$$

In the special case, if as a result of the observation ξ the state of the system is completely determined, then

$$H_\xi^{(apost)} = 0$$

and the information equals

$$I_\xi^{(s)} = H_\xi^{(apr)} = - \sum_{i=1}^n p_i \log p_i$$

or

$$I_\xi^{(s)} = M[-\log p(\xi)].$$

The amount of information about the system ξ by the observation the system η is determined as the decrease of the entropy of the system by the observation of the system η

$$I_{\eta \rightarrow \xi}^{(s)} = H_\xi - H_{\xi/\eta}.$$

The information $I_{\eta \rightarrow \xi}^{(s)}$ is called the mean or complete information contained in the system η about the system ξ . It is easily to show that

$$I_{\xi \rightarrow \eta}^{(s)} = H_\eta - H_{\eta/\xi},$$

and taking into account, that

$$H_{\xi\eta} = H_\xi + H_{\eta/\xi} = H_\eta + H_{\xi/\eta},$$

we obtain

$$I_{\xi \rightarrow \eta}^{(s)} = I_{\eta \rightarrow \xi}^{(s)} = I_{\eta \leftrightarrow \xi}^{(s)},$$

$I_{\eta \leftrightarrow \xi}^{(s)}$ is called by *complete mutual information*.

If the random variables ξ and η are independent, then

$$I_{\eta \leftrightarrow \xi} = 0, \quad \text{as} \quad H_{\eta/\xi} = H_\eta \quad \text{and} \quad H_{\xi/\eta} = H_\xi.$$

It is possible to represent $I_{\eta \leftrightarrow \xi}$ through the entropy of the joined system

$$I_{\eta \leftrightarrow \xi} = H_\xi + H_\eta - H_{\xi\eta}$$

and using the operator of the mathematical expectation

$$I_{\eta \leftrightarrow \xi} = M[-\log p(\xi) - \log p(\eta) + \log p(\xi, \eta)],$$

$$I_{\eta \leftrightarrow \xi} = M\left[\log \frac{p(\xi, \eta)}{p(\xi)p(\eta)}\right]$$

or in the following form

$$I_{\eta \leftrightarrow \xi} = \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log \frac{p_{ij}}{p_{\xi i} p_{\eta j}}.$$

The partial information about the system ξ , contained in the concrete state y_i is denoted as $I_{y_i \rightarrow \xi}$. The complete information is a mathematical expectation of the partial information

$$I_{\eta \leftrightarrow \xi} = \sum_{j=1}^m p_{\eta j} I_{y_j \rightarrow \xi},$$

where

$$I_{y_j \rightarrow \xi} = \sum_{i=1}^n p(x_i/y_j) \log \frac{p(x_i/y_j)}{p_{\xi i}}$$

or

$$I_{y_j \rightarrow \xi} = M_{y_i} \left[\log \frac{p(\xi/y_j)}{p(\xi)} \right],$$

where M_{y_i} is an operator of the conditional mathematical expectation. The partial information about the event x_i , obtained as a result of the event y_j , can be represented as

$$I_{y_j \rightarrow x_i} = \log \frac{p(x_i/y_j)}{p_{\xi i}}.$$

2.4 Entropy and information (continuous case)

Entropy of the system, described by the continuous random variable ξ with the density $f(x)$, is determined by

$$H_\xi = - \int_{-\infty}^{\infty} f(x) \log f(x) dx$$

or

$$H_\xi = M[-\log f(x)].$$

In the case of two continuous systems ξ and η with density $f(x, y)$, it is introduced *particular conditional entropy*, i.e. the entropy of the system η provided that the system ξ takes a specific state x :

$$H(\eta/x) = - \int_{-\infty}^{\infty} f(y/x) \log f(y/x) dy.$$

The *complete or average conditional entropy* is defined as

$$H(\eta/\xi) = - \iint_{-\infty}^{\infty} f_\xi(x) f(y/x) \log f(y/x) dxdy$$

or, taking into account

$$f_{\xi\eta}(x, y) = f_\xi(x)f_{\eta/\xi}(y/x),$$

we can write

$$H(\eta/\xi) = - \iint_{-\infty}^{\infty} f_{\xi\eta}(x, y) \log f_{\eta/\xi}(y/x) dx dy.$$

By analogy with discrete case we can show that the entropy of the system (ξ, η) equals

$$H(\xi, \eta) = H(\xi) + H(\eta/\xi).$$

For independent random variables ξ and η , we have

$$H(\xi, \eta) = H(\xi) + H(\eta).$$

Example: To find a density $f_\xi(x)$ of random variable ξ in the class of continuous distribution, which provides the maximum entropy H_ξ ($\max H_\xi$) by given the mathematical expectation m_ξ and variance σ_ξ^2 . Let

$$\int_{-\infty}^{\infty} f_\xi(x) dx = 1, \quad \int_{-\infty}^{\infty} x f_\xi(x) dx = m_\xi, \quad \int_{-\infty}^{\infty} (x - m_\xi)^2 f_\xi(x) dx = \sigma_\xi^2$$

be given, then we will find

$$\hat{f}_\xi(x) \Rightarrow \max \left[- \int_{-\infty}^{\infty} f_\xi(x) \log f_\xi(x) dx \right].$$

It is a typical variation problem

$$\hat{f}_\xi(x) \Rightarrow \max \left[\int_{-\infty}^{\infty} \Phi(f_\xi(x), x) dx \right]$$

under additional conditions

$$\int_{-\infty}^{\infty} \varphi_k(f_\xi(x), x) dx = c_k,$$

where

$$\begin{aligned} \Phi(f_\xi(x), x) &= -f_\xi(x) \log f_\xi(x), \\ \varphi_1 &= f_\xi(x), \quad c_1 = 1, \\ \varphi_2 &= x f_\xi(x), \quad c_2 = m_\xi, \\ \varphi_3 &= (x - m_\xi)^2 f_\xi(x), \quad c_3 = \sigma_\xi^2. \end{aligned}$$

It will be used the method of Lagrange multipliers. We shall write the Euler equation

$$\partial\Phi_1/\partial f = 0,$$

where

$$\Phi_1(f, x) = -f \log f + \lambda_1 f + \lambda_2 x f + \lambda_3 (x - m_\xi)^2 f.$$

If we differentiate on the f_1 we obtain

$$-1 - \log f + \lambda_1 + \lambda_2 x + \lambda_3 (x - m_\xi)^2 = 0.$$

We express the function f from the last equation

$$f_\xi(x) = \exp(\lambda_1 + \lambda_2 x + \lambda_3 (x - m_\xi)^2 - 1).$$

We find from the additional conditions

$$\begin{aligned} \int_{-\infty}^{\infty} \exp(\lambda_1 + \lambda_2 x + \lambda_3 (x - m_\xi)^2 - 1) dx &= 1, \\ \int_{-\infty}^{\infty} x \exp(\lambda_1 + \lambda_2 x + \lambda_3 (x - m_\xi)^2 - 1) dx &= m_\xi, \\ \int_{-\infty}^{\infty} (x - m_\xi)^2 \exp(\lambda_1 + \lambda_2 x + \lambda_3 (x - m_\xi)^2 - 1) dx &= \sigma_\xi^2. \end{aligned}$$

It is easily to show that

$$\lambda_1 = 1 - \ln(\sqrt{2\pi}\sigma_\xi), \quad \lambda_2 = 0, \quad \lambda_3 = -(2\sigma_\xi^2)^{-1}$$

and finally we obtain

$$\hat{f}_\xi(x) = \frac{1}{\sigma_\xi \sqrt{2\pi}} \exp\left\{-\frac{(x - m_\xi)^2}{2\sigma_\xi^2}\right\}$$

the normal distribution, which provides the maximum entropy under the fixed first two moments m_ξ and σ_ξ^2 .

The *complete mutual information* contained in continuous systems ξ and η , is defined as

$$I_{\eta \leftrightarrow \xi} = \iint_{-\infty}^{\infty} f_{\xi\eta}(x, y) \log \frac{f_{\xi\eta}(x, y)}{f_\xi(x)f_\eta(y)} dx dy$$

or

$$I_{\eta \leftrightarrow \xi} = M[\log(f_{\xi\eta}/f_\xi f_\eta)].$$

If the r.v. ξ and η are independent, then the complete mutual information equals zero:

$$I_{\eta \leftrightarrow \xi} = 0.$$

Example: We consider a simple model

$$\eta = \xi + \zeta,$$

where η is a measured signal, ξ is a transmitted signal, with the normal density $N(m_\xi = 0, \sigma_\xi^2)$, ζ is an additive noise with the normal density $(m_\zeta, \sigma_\zeta^2)$, ξ and ζ are independent. It is required to find the amount of information contained in the measured signal η about the transmitted signal ξ . By definition:

$$I_{\eta \leftrightarrow \xi} = M \left[\log \frac{f_{\xi \eta}}{f_\xi f_\eta} \right] = M \left[\log \frac{f_{\eta/\xi}}{f_\eta} \right] = M \varkappa,$$

where $\varkappa = \log(f_{\eta/\xi}/f_\eta)$. We can write $\sigma_\eta^2 = \sigma_\xi^2 + \sigma_\zeta^2$ and

$$f_\eta(y) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_\xi^2 + \sigma_\zeta^2}} \exp \left\{ -\frac{1}{2} \frac{y^2}{(\sigma_\xi^2 + \sigma_\zeta^2)} \right\},$$

$$f_{\eta/\xi}(y/x) = \frac{1}{\sqrt{2\pi} \sigma_\eta \sqrt{1-r^2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[\frac{y}{\sigma_\eta} - r \frac{x}{\sigma_\xi} \right]^2 \right\}.$$

We shall find the correlation coefficient r

$$r = \frac{1}{\sigma_\eta \sigma_\xi} M(\eta \cdot \xi) = \frac{1}{\sigma_\eta \sigma_\xi} [M(\xi^2) + M(\xi \zeta)] = \frac{\sigma_\xi^2}{\sigma_\eta^2}.$$

Substituting r to $f_{\eta/\xi}(y/x)$, we obtain

$$f_{\eta/\xi}(y/x) = \frac{1}{\sqrt{2\pi} \sigma_\zeta} \exp \left\{ -\frac{(y-x)^2}{2\sigma_\zeta^2} \right\} = \frac{1}{\sqrt{2\pi} \sigma_\zeta} \exp \left\{ -\frac{z^2}{2\sigma_\zeta^2} \right\}.$$

We shall find the random value \varkappa

$$\varkappa = \log \frac{\sqrt{\sigma_\xi^2 + \sigma_\zeta^2}}{\sigma_\zeta} + \frac{1}{\ln 2} \left[\frac{\zeta^2}{2\sigma_\zeta^2} - \frac{\eta^2}{2(\sigma_\xi^2 + \sigma_\zeta^2)} \right]$$

and taking the mathematical expectation of \varkappa , we obtain

$$I_{\eta \leftrightarrow \xi} = M \varkappa = \log \frac{\sqrt{\sigma_\xi^2 + \sigma_\zeta^2}}{\sigma_\zeta} + \frac{1}{\ln 2} \left[\frac{M\zeta^2}{2\sigma_\zeta^2} - \frac{M\eta^2}{2(\sigma_\xi^2 + \sigma_\zeta^2)} \right].$$

Taking into account that $M\zeta^2 = \sigma_\zeta^2$, $M\eta^2 = \sigma_\xi^2 + \sigma_\zeta^2$ we obtain finally

$$I_{\eta \leftrightarrow \xi} = \log \frac{\sqrt{\sigma_\xi^2 + \sigma_\zeta^2}}{\sigma_\zeta}.$$

2.5 Fisher information

Let the random variable ξ be distributed according to $f_\xi(x, \theta)$, where θ is a parameter of the density, and the $f_\xi(x, \theta)$ be differentiated on the θ . Then the *Fisher information* about the unknown parameter θ , contained in the r.v. ξ is defined as

$$I_\xi^{(F)}(\theta) = M \left[\left(\frac{\partial \log f_\xi(x, \theta)}{\partial \theta} \right)^2 \right].$$

For the random vector $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$, and the parameter vector $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_s)$ it is introduced the *Fisher matrix* with elements

$$I_{ss'}^{(F)}(\vec{\theta}) = M \left[\frac{\partial \log f_{\vec{\xi}}(x_1, x_2, \dots, x_n, \vec{\theta})}{\partial \theta_s} \frac{\partial \log f_{\vec{\xi}}(x_1, x_2, \dots, x_n, \vec{\theta})}{\partial \theta_{s'}} \right].$$

Properties of the Fisher information.

1. Let the random variables ξ_1 and ξ_2 be independent random variables with the Fisher information about a parameter θ accordingly $I_{\xi_1}^{(F)}(\theta)$ and $I_{\xi_2}^{(F)}(\theta)$. Then the Fisher information about the parameter θ contained in the pair ξ_1, ξ_2 equals

$$I_{\xi_1 \xi_2}^{(F)}(\theta) = I_{\xi_1}^{(F)}(\theta) + I_{\xi_2}^{(F)}(\theta).$$

Proof:

$$\begin{aligned} I_{\xi_1 \xi_2}^{(F)}(\theta) &= M\{\partial[\log f_{\xi_1}(x, \theta) \cdot f_{\xi_2}(x, \theta)]/\partial \theta\}^2 = \\ &= M[(\partial \log f_{\xi_1}(x, \theta)/\partial \theta)^2] + M[(\partial \log f_{\xi_2}(x, \theta)/\partial \theta)^2] + \\ &\quad + 2M[\partial \log f_{\xi_1}(x, \theta)/\partial \theta \cdot \partial \log f_{\xi_2}(x, \theta)/\partial \theta] = \\ &= I_{\xi_1}^{(F)}(\theta) + I_{\xi_2}^{(F)}(\theta). \end{aligned}$$

Taking into account that

$$\begin{aligned} M[\partial \log f_{\xi_1}(x, \theta)/\partial \theta \cdot \partial \log f_{\xi_2}(x, \theta)/\partial \theta] &=, \\ M[\partial \log f_{\xi_1}(x, \theta)/\partial \theta] \cdot M[\partial \log f_{\xi_2}(x, \theta)/\partial \theta] &= 0 \end{aligned}$$

as

$$\int_{-\infty}^{\infty} f_\xi(x, \theta) \frac{1}{f_\xi(x, \theta)} \frac{\partial f_\xi(x, \theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f_\xi(x, \theta) dx = 0.$$

2. Let the random variables $\xi_1, \xi_2, \dots, \xi_n$ be independent and equally distributed with the Fisher information $I_\xi^{(F)}(\theta)$. Then the Fisher information $I_{\vec{\xi}}^{(F)}(\theta)$ equals

$$I_{\vec{\xi}}^{(F)} = n I_\xi^{(F)}(\theta)$$

- a) For two r.v. ξ_1 and ξ_2 and vector parameter $\vec{\theta}$, we obtain the Fisher information matrix

$$I_{\xi_1 \xi_2}^{(F)}(\vec{\theta}) = I_{\xi_1}^{(F)}(\vec{\theta}) + I_{\xi_2}^{(F)}(\vec{\theta}).$$

- b) For the random vector $\vec{\xi} = [\xi_1, \xi_2, \dots, \xi_n]$ and the vector parameter $\vec{\theta}$, we obtain the Fisher information matrix

$$I_{\vec{\xi}}^{(F)}(\vec{\theta}) = n I_{\xi}^{(F)}(\vec{\theta}).$$

3. Let $T = T(\xi)$ be given the function of the random variable ξ , then the following inequality is satisfied

$$I_{\xi}^{(F)}(\vec{\theta}) - I_{T(\xi)}^{(F)}(\vec{\theta}) \geq 0.$$

In the case of one parameter

$$I_{\xi}^{(F)} \geq I_{T(\xi)}^{(F)},$$

amount of the Fisher information contained in the random variables ξ always more or equal the amount of information contained in the function of the random variable $T(\xi)$.

4. We can represent the Fisher information in the form:

$$I_{\xi}^{(F)}(\theta) = M[-\partial^2 \log f_{\xi}(x, \theta) / \partial \theta^2],$$

since it is easily to show

$$\frac{\partial^2 \log f}{\partial \theta^2} = -\frac{1}{f^2} \left(\frac{\partial f}{\partial \theta} \right)^2 + \frac{1}{f} \frac{\partial^2 f}{\partial \theta^2}, \quad \left(\frac{1}{f} \frac{\partial f}{\partial \theta} \right)^2 = \left(\frac{\partial \log f}{\partial \theta} \right)^2$$

and

$$\begin{aligned} M \left(\frac{1}{f} \frac{\partial^2 f}{\partial \theta^2} \right) &= \int_{-\infty}^{\infty} f_{\xi}(x, \theta) \frac{1}{f_{\xi}(x, \theta)} \frac{\partial^2 f_{\xi}(x, \theta)}{\partial \theta^2} dx = \\ &= \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{\infty} f_{\xi}(x, \theta) dx = 0. \end{aligned}$$

In the case of the vector parameter $\vec{\theta}$ we obtain

$$I_{ss'}^{(F)}(\vec{\theta}) = M[-\partial^2 \log f(x, \vec{\theta}) / \partial \theta_s \partial \theta_{s'}].$$

Example: Let the random variable ξ be normally distributed $\xi \in N(m_{\xi}, \sigma_{\xi}^2)$. Let the mean value m_{ξ} be unknown and the variance σ_{ξ}^2 is known. Then the Fisher information about the mean value equals

$$I_{\xi}^{(F)}(m_{\xi}) = - \int_{-\infty}^{\infty} \frac{\partial^2}{\partial m_{\xi}^2} \left[-\frac{(x - m_{\xi})^2}{2\sigma_{\xi}^2} - \ln \sigma_{\xi} \sqrt{2\pi} \right] \times$$

$$\begin{aligned}
& \times \frac{1}{\sqrt{2\pi}\sigma_\xi} \exp\left[-\frac{(x-m_\xi)^2}{2\sigma_\xi^2}\right] dx = \\
& = \int_{-\infty}^{\infty} \frac{1}{\sigma_\xi^2\sqrt{2\pi}\sigma_\xi} \exp\left[-\frac{(x-m_\xi)^2}{2\sigma_\xi^2}\right] dx = \frac{1}{\sigma_\xi^2}.
\end{aligned}$$

3

Random functions and its properties

Definitions: Let t be a nonrandom value, then the random function $\xi(t)$ is called *stochastic process* if the value t is continuous and $\xi(t_i)$ is called *time series* if the value t is discrete.

We assume that we have the n independent experiments. In each experiment we obtain the n realizations of the random function $\xi(t)$ (see Figure 3.1):

$$x_1(t), x_2(t), \dots, x_n(t).$$

The random function $\xi(t)$ at the moment $t = t_1$ is the random variable $\xi_1 = \xi(t_1)$,

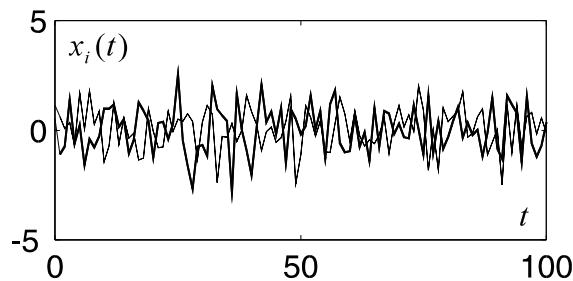


Figure 3.1: Three realizations of the stochastic process

described by the density $f(x_1/t_1)$.

For two points t_1 and t_2 , we have r.v. $\xi_1 = \xi(t_1)$ and $\xi_2 = \xi(t_2)$, which are described by the density $f(x_1, x_2/t_1, t_2)$.

The random function $\xi(t)$ is completely determined by the density

$$f(x_1, x_2, \dots, x_n/t_1, t_2, \dots, t_n).$$

The first two moments play very important role by the practical applications.

3.1 The expected value and variance

The expected value or mean is defined by

$$m_{\xi_1} = M[\xi(t_1)] = \langle \xi(t_1) \rangle$$

or for an arbitrary t

$$m_\xi(t) = \int_{-\infty}^{\infty} xf(x/t)dx.$$

The variance or dispersion is defined by

$$D[\xi(t)] = M[(\xi(t) - m_\xi(t))^2] = \langle (\xi(t) - m_\xi(t))^2 \rangle.$$

The correlation function (autocorrelation) function is defined by

$$\begin{aligned} R(t_1, t_2) &= \text{cov}(\xi(t_1), \xi(t_2)) = M[(\xi(t_1) - m_\xi(t_1))(\xi(t_2) - m_\xi(t_2))] = \\ &= \langle (\xi(t_1) - m_\xi(t_1))(\xi(t_2) - m_\xi(t_2)) \rangle \end{aligned}$$

or using the definition of the operator of mathematical expectation:

$$\begin{aligned} D[\xi(t)] &= \int_{-\infty}^{\infty} (x(t) - m_\xi(t))^2 f(x/t) dx, \\ R(t_1, t_2) &= \iint_{-\infty}^{\infty} (x(t_1) - m_\xi(t_1))(x(t_2) - m_\xi(t_2)) f(x_1, x_2/t_1, t_2) dx_1 dx_2. \end{aligned}$$

The average on time (on parameter t) of the random function $\xi(t)$ is defined by

$$\langle \xi(t) \rangle_T = \lim_{T \rightarrow \infty} (1/2T) \int_{-T}^T \xi(t) dt.$$

3.2 Properties of random functions

3.2.1 Stationarity

We shall give two definitions of stationarity.

A. Strict sense. We say that a random function $\xi(t)$ is *stationary (in the strict sense)* if its density are not affected by a shift in the time origin

$$f(x_1, \dots, x_n/t_1, \dots, t_n) = f(x_1, \dots, x_n/t_1 + t_0, \dots, t_n + t_0)$$

for any t_0 .

In the particular cases $n = 1$ and $n = 2$ we have accordingly

$$f(x_1/t_1) = f(x_1/0) = f(x_1)$$

and

$$f(x_1, x_2/t_1, t_2) = f(x_1, x_2/0, \tau) \quad \tau = t_2 - t_1,$$

i.e. one-dimensional density does not depend on time, and two-dimensional density depends only on time difference.

B. Wide sense. We say that a random function $\xi(t)$ is *stationary in the wide sense*, if its expected value and variance are constants and its autocorrelation depends only on $\tau = t_2 - t_1$

$$m_\xi(t) = \int_{-\infty}^{\infty} x f(x) dx = \text{const},$$

$$D[\xi(t)] = \int_{-\infty}^{\infty} (x(t) - m_\xi(t))^2 f(x) dx = \text{const},$$

$$R(t_1, t_2) = \int_{-\infty}^{\infty} (x(t_1) - m_\xi(t_1))(x(t_2) - m_\xi(t_2)) f(x_1, x_2/t_1, t_2) dx_1 dx_2 = R(t_2 - t_1) = R(\tau),$$

$$m_\xi = \text{const}, \quad D[\xi(t)] = \text{const}, \quad R(t_1, t_2) = R(t_2 - t_1) = R(\tau).$$

Examples of nonstationary processes are represented in the Figure 3.2.

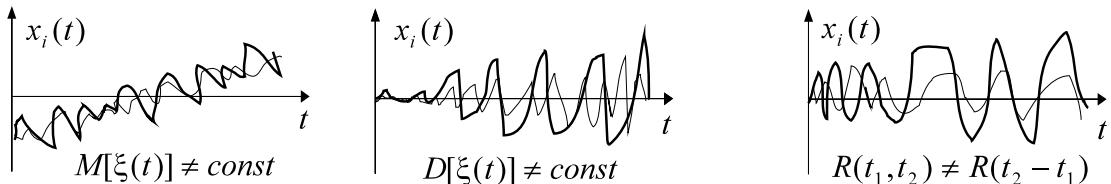


Figure 3.2: Examples of nonstationary processes.

3.2.2 Ergodicity

Ergodicity deals with the problem of determining the statistics of a process $\xi(t)$ from a signal observation. Stochastic process $\xi(t)$ is ergodic in the most general form if (with probability 1) all its statistics can be determined from a single function of the process $\xi(t)$. Stochastic process $\xi(t)$ is ergodic if time averages equal ensemble averages (i.e. expected values)

$$P [\langle \xi(t) \rangle_T = M[\xi(t)]] = 1.$$

(i.e. we can calculate $M[\xi(t)]$ at the fixed point t_i using the set of realizations of the stochastic process (see Figure 3.3a) or calculate this quantity by averaging on time domain using only one realization (see Figure 3.3b) of the stochastic process.)

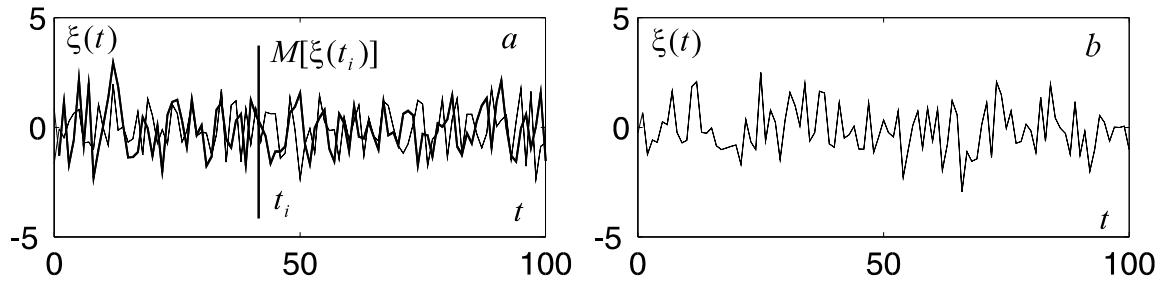


Figure 3.3: An illustration of the ergodicity.

3.3 Properties of autocorrelation functions

1. The autocorrelation function is an even function

$$R(\tau) = R(-\tau).$$

It follows from the definition.

2. The value of the real autocorrelation function is equal or less than the variance, i.e.

$$D[\xi(t)] = R(0) \geq R(\tau).$$

3. The product $[\xi(t) - m_\xi(t)]$ on the arbitrary nonrandom function $z(t)$ leads to the realization of the following inequality

$$\int_a^b \int_a^b z(t_1)z(t_2)R(t_2 - t_1)dt_1 dt_2 \geq 0.$$

Examples of the approximation of autocorrelation function – see Figure 3.4.

4. Let the $\varphi(t)$ be the nonrandom function and $\xi(t)$ be the random function. We consider the function $\eta(t) = \xi(t) + \varphi(t)$. Taking into account that the mean value of the η is

$$m_\eta(t) = m_\xi(t) + \varphi(t),$$

we obtain the autocorrelation function $R_\eta(t_1, t_2)$

$$\begin{aligned} R_\eta(t_1, t_2) &= M[(\eta(t_1) - m_\eta(t_1))(\eta(t_2) - m_\eta(t_2))] = \\ &= M[(\xi(t_1) - m_\xi(t_1))(\xi(t_2) - m_\xi(t_2))] = R_\xi(t_1, t_2). \end{aligned}$$

3.4 Action of a linear operator on a random function

Consider a linear system with impulse response the real function $h(t)$ and input $\xi(t)$. Its output $\eta(t)$ (see Figure 3.5) is given by

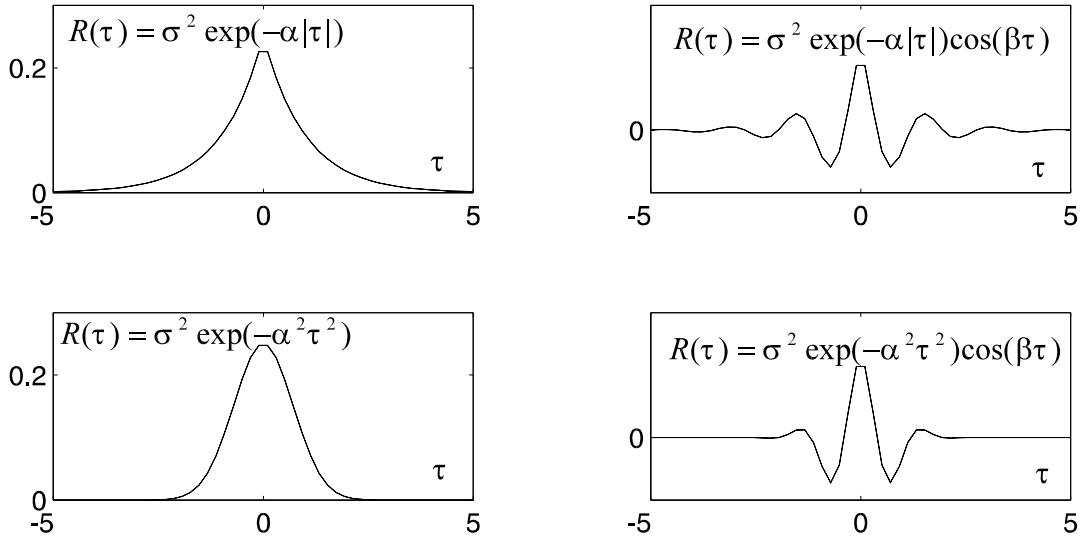


Figure 3.4: Examples of the approximation of autocorrelation function ($\sigma = 0.5$, $\alpha = 1.0$, $\beta = 4.0$).

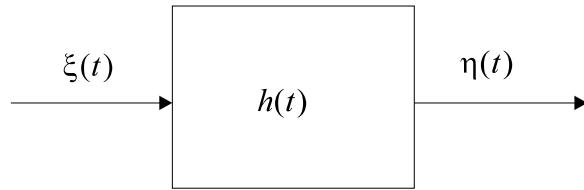


Figure 3.5: Scheme of an action of the operator H .

$$\eta(t) = H[\xi(t)] = \int_{-\infty}^{\infty} \xi(t - \tau) h(\tau) d\tau.$$

Properties of the linear operator H .

1. $H[c\xi(t)] = cH[\xi(t)]$.
2. $H[\xi_1(t) + \xi_2(t)] = H\xi_1(t) + H\xi_2(t)$.
3. $\eta(t) = H[\xi(t)]$, $m_\eta(t) = M[H[\xi(t)]] = H[M[\xi(t)]] = Hm_\xi(t)$,
 $R_\eta(t_1, t_2) = M[(\eta(t_1) - m_\eta(t_1))(\eta(t_2) - m_\eta(t_2))] =$
 $= M\{(H[\xi(t_1)] - H[m_\xi(t_1)])(H[\xi(t_2)] - H[m_\xi(t_2)])\} =$
 $= H_{t_1}H_{t_2}M[(\xi(t_1) - m_\xi(t_1))(\xi(t_2) - m_\xi(t_2))] = H_{t_1}H_{t_2}R_\xi(t_1, t_2)$.

3.5 Cross-correlation function

We consider the system of a two-dimensional process $\xi(t)$ and $\eta(t)$.

The *cross-correlation function* is defined by

$$R_{\xi\eta}(t_1, t_2) = M[(\xi(t_1) - m_\xi(t_1))(\eta(t_2) - m_\eta(t_2))].$$

For the real random functions, we shall have

$$R_{\xi\eta}(t_1, t_2) = R_{\eta\xi}(t_2, t_1)$$

and for stationary $\xi(t)$ and $\eta(t)$

$$R_{\xi\eta}(\tau) = R_{\eta\xi}(-\tau).$$

According to the definition of the cross-correlation function we can write

$$R_{\xi\eta}(t_1, t_2) = \iint_{-\infty}^{\infty} (x(t_1) - m_\xi(t_1))(y(t_2) - m_\eta(t_2)) f_{\xi\eta}(x, y/t_1, t_2) dx dy$$

and using Cauchy-Bunaykovskii inequality, we shall obtain

$$\begin{aligned} & [\iint_{-\infty}^{\infty} (x(t_1) - m_\xi(t_1))(y(t_2) - m_\eta(t_2)) f(x, y/t_1, t_2) dx dy]^2 \leq \\ & \leq \iint_{-\infty}^{\infty} (x(t_1) - m_\xi(t_1))^2 f(x, y/t_1, t_2) dx dy \times \\ & \quad \times \iint_{-\infty}^{\infty} (y(t_2) - m_\eta(t_2))^2 f(x, y/t_1, t_2) dx dy \Rightarrow \\ & [R_{\xi\eta}(t_1, t_2)]^2 \leq \sqrt{R_\xi(t_1, t_1) R_\eta(t_2, t_2)}. \end{aligned}$$

After the normalization we shall obtain the normalized cross-correlation function

$$r_{\xi\eta}(t_1, t_2) = \frac{R_{\xi\eta}(t_1, t_2)}{\sqrt{R_\xi(t_1, t_1)} \sqrt{R_\eta(t_2, t_2)}}, \quad -1 \leq r_{\xi\eta}(t_1, t_2) \leq 1$$

and

$$R_{\xi\eta}(\tau) \leq \sqrt{D_\xi D_\eta}.$$

3.6 Wiener-Khinchin theorem and power spectrum

1. Let the time average of the modulus

$$\langle |\xi| \rangle_T = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |x(t)| dt$$

and of squared value

$$\langle |\xi|^2 \rangle_T = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |x(t)|^2 dt$$

exist.

2. Let $\xi(t)$ be the random function with the bounded variation. The correlation function

$$R_\xi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \xi(t)\xi(t+\tau)dt$$

exists always, if $\langle |\xi|^2 \rangle_T$ exists. The cross-correlation function

$$R_{\xi\eta}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \xi(t)\eta(t+\tau)dt$$

exists always, if $\langle |\xi|^2 \rangle_T$ and $\langle |\eta|^2 \rangle_T$ exist.

The spectral density $R_\xi(\omega)$ for the function $\xi(t)$ and mutual spectral density $R_{\xi\eta}(\omega)$ for the pair of functions $\xi(t)$ and $\eta(t)$ is defined by the Khinchin-Wiener relation

$$\begin{aligned} R_\xi(\omega) &= \int_{-\infty}^{\infty} R_\xi(\tau) \exp\{-i\omega\tau\} d\tau = 2 \int_0^{\infty} R_\xi(\tau) \cos \omega\tau d\tau, \\ R_\xi(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} R_\xi(\omega) \exp\{i\omega\tau\} d\omega = \frac{1}{\pi} \int_0^{\infty} R_\xi(\omega) \cos \omega\tau d\omega, \\ R_{\xi\eta}(\omega) &= \int_{-\infty}^{\infty} R_{\xi\eta}(\tau) \exp\{-i\omega\tau\} d\tau, \\ R_{\xi\eta}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} R_{\xi\eta}(\omega) \exp\{i\omega\tau\} d\omega. \end{aligned}$$

For the real random function $\xi(t)$ the spectral density is even:

$$R_\xi(\omega) = R_\xi(-\omega).$$

3.7 Examples of correlation functions and power spectrum

Example 1: Graphic presentation of the correlation function

$$R_\xi(\tau) = \sigma_\xi^2 \exp\{-\alpha|\tau|\} \quad (3.1)$$

and its spectrum

$$\begin{aligned}
 R_\xi(\omega) &= \sigma_\xi^2 \int_{-\infty}^{\infty} \exp\{-i\omega\tau - \alpha|\tau|\} d\tau = \sigma_\xi^2 \left[\int_0^{\infty} \exp\{-(i\omega + \alpha)\tau\} d\tau + \right. \\
 &\quad \left. + \int_{-\infty}^0 \exp\{(\alpha - i\omega)\tau\} d\tau \right] = \sigma_\xi^2 \left[\frac{1}{\alpha + i\omega} + \frac{1}{\alpha - i\omega} \right] = 2\sigma_\xi^2 \frac{\alpha}{\alpha^2 + \omega^2} \quad (3.2)
 \end{aligned}$$

– see Figures 3.6a and 3.6b correspondingly.

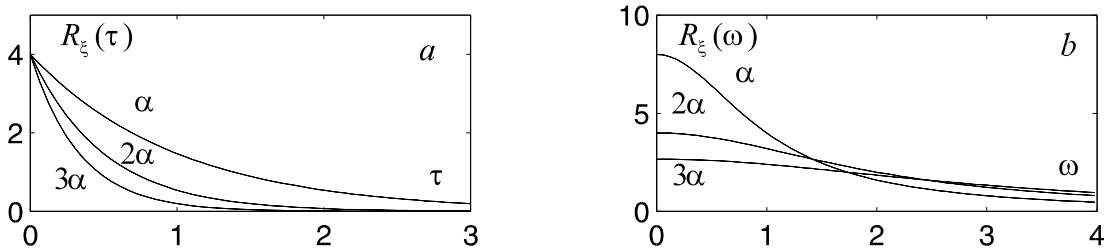


Figure 3.6: Graphic presentation of correlation function (3.1) (a) and its spectrum (3.2)(b) ($\sigma = 2$, $\alpha = 1$).

Example 2: $R_\xi(\tau) = \sigma_\xi^2 \exp\{-\alpha|\tau|\} \cos \omega_0 \tau$.

Using the representation of $\cos \omega_0 \tau$

$$\cos \omega_0 \tau = (\exp\{i\omega_0 \tau\} + \exp\{-i\omega_0 \tau\})/2$$

and taking the Fourier transform, we shall obtain.

$$\begin{aligned}
 R_\xi(\omega) &= \frac{\sigma_\xi^2}{2} \left[\int_{-\infty}^{\infty} \exp\{-i\omega\tau - \alpha|\tau| + i\omega_0\tau\} d\tau + \right. \\
 &\quad \left. + \int_{-\infty}^{\infty} \exp\{-i\omega\tau - \alpha|\tau| - i\omega_0\tau\} d\tau \right] = \\
 &= \sigma_\xi^2 \left[\frac{\alpha}{(\omega - \omega_0)^2 + \alpha^2} + \frac{\alpha}{(\omega + \omega_0)^2 + \alpha^2} \right].
 \end{aligned}$$

Example 3: $R_\xi(\tau) = \sigma_\xi^2 \exp\{-\alpha^2 \tau^2\} \cos \omega_0 \tau$.

Using the representation of $\cos \omega_0 \tau$

$$\cos \omega_0 \tau = (\exp\{i\omega_0 \tau\} + \exp\{-i\omega_0 \tau\})/2$$

and taking the Fourier transform, we shall obtain

$$\begin{aligned} R_\xi(\omega) &= \frac{\sigma_\xi^2}{2} \left[\int_{-\infty}^{\infty} \exp\{-\alpha^2\tau^2 + i\omega_0\tau - i\omega\tau\} d\tau + \right. \\ &\quad \left. + \int_{-\infty}^{\infty} \exp\{-\alpha^2\tau^2 - i\omega_0\tau - i\omega\tau\} d\tau \right]. \end{aligned}$$

Each integral might be reduced to the Poisson integral, transforming the exponent to the form

$$\left[\left(\alpha\tau + i\frac{\omega}{2\alpha} \mp i\frac{\omega_0}{2\alpha} \right)^2 + \frac{1}{4\alpha^2}(\omega \pm \omega_0)^2 \right].$$

Calculating these integrals, we shall obtain

$$R_\xi(\omega) = \frac{\sigma_\xi^2 \sqrt{\pi}}{2\alpha} \left[\exp\left\{-\frac{(\omega + \omega_0)^2}{4\alpha^2}\right\} + \exp\left\{-\frac{(\omega - \omega_0)^2}{4\alpha^2}\right\} \right].$$

3.8 Estimation of numerical characteristics of random function

Let the realization of the random function be registered with the digitization step Δt :

$$x_i(t_j), \quad j = 1, \dots, m; \quad j = 1, \dots, n.$$

The estimation of the mean value is:

$$\hat{m}_\xi(t_j) = \frac{1}{n} \sum_{i=1}^n x_i(t_j).$$

The estimations of the correlation function can be presented as

$$\hat{R}_\xi(t_j, t_k) = \frac{1}{n-1} \sum_{i=1}^n [(x_i(t_j) - \hat{m}_\xi(t_j))(x_i(t_k) - \hat{m}_\xi(t_k))]$$

or

$$\hat{R}_\xi(t_j, t_k) = \frac{n}{n-1} \sum_{i=1}^n x_i(t_j)x_i(t_k) - \frac{n}{n-1} \hat{m}_\xi(t_j)\hat{m}_\xi(t_k).$$

For the stationary and ergodic processes we shall obtain the estimations of the mean value

$$\hat{m}_\xi = (1/m) \sum_{j=1}^m x(t_j)$$

(for discrete case)

$$\hat{m}_\xi = \frac{1}{T} \int_0^T x(t) dt,$$

(for continuous case) and the estimation of the correlation function

$$\hat{R}_\xi(\tau) = \frac{1}{m-l-1} \sum_{j=1}^{m-l} (x(t_j) - \hat{m}_\xi)(x(t_j + \tau) - \hat{m}_\xi)$$

(for discrete case),

$$\hat{R}_\xi(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} [x(t) - \hat{m}_\xi][x(t+\tau) - \hat{m}_\xi] dt$$

(for continuous case).

4

Elements of mathematical statistics

4.1 Sampling. Estimation.

Let $\vec{\xi} = (x_1, x_2, \dots, x_n)$, be a random vector defined by digital registrations of the outcomes of a random experiment. The values which this random vector takes on or its realizations are called *observations*, *measurements*, or *data*.

Let the function $g(x_1, \dots, x_n)$ be an estimate.

Example 1: The estimate of the average value

$$\bar{x} = (1/n) \sum_{i=1}^n x_i.$$

Example 2: The estimate of the variance

$$s^2 = (1/n) \sum_{i=1}^n (x_i - \bar{x})^2.$$

Example 3: The estimate of the coefficient of variance

$$v = \frac{s}{\bar{x}}.$$

Let the random vector $\vec{\xi}$ belong to the density

$$\vec{\xi} \in f(\vec{x}, \vec{\theta}), \quad \theta \in \Omega,$$

where $\vec{\theta}$ is an unknown parameter vector. The principal elements which determine the solution of mathematical statistic problems, are the following:

1. Class of distributions

$$\mathcal{F} = \{f(\vec{x}, \vec{\theta}), \vec{\theta} \in \Omega\}$$

for the random vector $\vec{\xi}$.

2. Structure of the space D of decisions $d, d \in D$.
3. Space of observations (data) $\vec{x} \in X$.
4. Space of unknown parameters $\vec{\theta} \in \Omega$

The interaction between X , D and Ω is shown on the Figure 4.1. The unknown

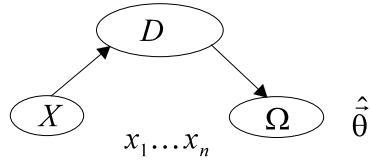


Figure 4.1: The interaction between space of decisions, space of observations and space of parameters.

parameter contained in $\vec{\theta}$ will be estimated by means of the function $g(x_1, \dots, x_n)$ of the observations \vec{x} . This is called a *point estimation*. The function $g(\vec{x})$ is called an *estimator*

$$\hat{\theta}_n = g(x_1, \dots, x_n).$$

4.2 Consistent estimator

An estimator is called *consistent*, if with the probability of one the sequence of estimators for an unlimited increasing sample size comes arbitrary close to the parameter to be estimated

$$P(|\hat{\theta}_n - \theta_0| \leq \varepsilon) \geq 1 - \eta,$$

where $\varepsilon > 0$, $\eta > 0$ (see Figure 4.2).

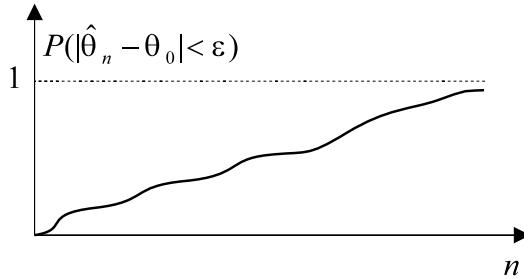


Figure 4.2: An illustration of the convergence on probability.

4.3 Unbiased estimator

Let $\hat{\theta}_n$ be the estimation of the parameter and θ_0 be the true value of the parameter. The expected value $M[\hat{\theta}_n - \theta_0]$ of the error of the estimation is called *bias*. It should be equal to zero. We obtain an *unbiased estimator* if

$$b_n(\hat{\theta}) = M(\hat{\theta}_n) - \theta_0 = 0, \quad M[\hat{\theta}] = \theta_0$$

is satisfied. Asymptotic unbiased estimation:

$$M(\hat{\theta}_n) \xrightarrow{n \rightarrow \infty} \theta_0$$

(examples for an illustration – see Figure 4.3).

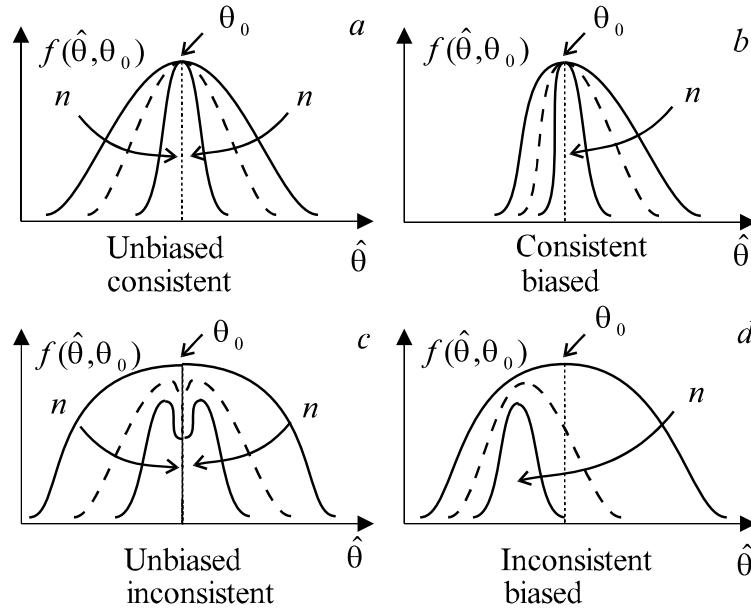


Figure 4.3: Examples of the types of estimators.

Example: Let (x_1, x_2, \dots, x_n) be the repeated sampling, normally distributed $x_i \in N(m_\xi, \sigma_\xi^2)$. It will be necessary to find an estimation \hat{S}^2 of the variance σ^2 .

We calculate the sample variance

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2.$$

The expected value is

$$M(s^2) = \frac{1}{n} \sum_{i=1}^n M(x_i^2) - M(\bar{x}^2) = \left(1 - \frac{1}{n}\right) \sigma^2.$$

Hence, the unbiased estimation of the variance will be in a form

$$\hat{S}^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad M(\hat{S}^2) = \frac{n}{n-1} M(s^2) = \sigma^2.$$

4.4 Rao-Kramer inequality. Efficient estimator

The lower bound for the variance of the parameter estimation is determined by Rao-Kramer inequality.

Let x_1, x_2, \dots, x_n to be repeated sampling with the density

$$L(\vec{x}, \theta) = \prod_{i=1}^n f(x_i, \theta).$$

Let $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$ be the estimation. The density $L(\vec{x}, \theta)$ is normalized

$$\int \dots \int L(x_1, x_2, \dots, x_n, \theta) dx_1, \dots, dx_n = 1.$$

If we shall differentiate on parameter θ we shall obtain

$$\int \dots \int \frac{\partial L}{\partial \theta} dx_1, \dots, dx_n = 0. \quad (4.1)$$

By virtue of the unbiased estimation $\hat{\theta}$

$$M\hat{\theta} = \theta$$

or

$$\int \dots \int \hat{\theta} L dx_1, \dots, dx_n = \theta, \quad \int \dots \int \hat{\theta}(x_1, \dots, x_n) \frac{\partial L}{\partial \theta} dx_1, \dots, dx_n = 1. \quad (4.2)$$

Taking into account (4.1) and (4.2) we can write

$$\int \dots \int [\hat{\theta} - \theta] \frac{\partial L(x_1, \dots, x_n, \theta)}{\partial \theta} dx_1, \dots, dx_n = 1$$

or

$$\int \dots \int [\hat{\theta} - \theta] \left[\frac{1}{L} \frac{\partial L}{\partial \theta} \right] L dx_1, \dots, dx_n = 1.$$

We can write the last relation in the form

$$M \left[(\hat{\theta} - \theta) \left(\frac{1}{L} \frac{\partial L}{\partial \theta} \right) \right] = 1.$$

Denoting by

$$\hat{\theta} - \theta = V \quad \text{and} \quad \frac{1}{L} \frac{\partial L}{\partial \theta} = W$$

and using the analog of Cauchy-Bunyakovskii inequality we shall have

$$M(V^2) \cdot M(W^2) \geq [M(V \cdot W)]^2$$

or

$$M(V^2) \cdot M(W^2) \geq 1, \quad M(\hat{\theta} - \theta)^2 \cdot M \left[\frac{1}{L} \frac{\partial L}{\partial \theta} \right]^2 \geq 1. \quad (4.3)$$

Assuming that $L \neq 0$, we can write

$$\begin{aligned}\frac{\partial \ln L}{\partial \theta} &= \frac{1}{L} \frac{\partial L}{\partial \theta}, \\ \frac{\partial^2 \ln L}{\partial \theta^2} &= -\frac{1}{L^2} \left[\frac{\partial L}{\partial \theta} \right]^2 + \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2}.\end{aligned}\quad (4.4)$$

Then we multiply (4.4) on L and obtain expected value

$$M \left[\frac{1}{L} \frac{\partial^2 L}{\partial \theta^2} \right] = \int \dots \int \frac{\partial^2 L}{\partial \theta^2} dx_1, \dots, dx_n = 0$$

and

$$M \left[\frac{1}{L} \frac{\partial L}{\partial \theta} \right]^2 = -M \left[\frac{\partial^2 \ln L}{\partial \theta^2} \right]. \quad (4.5)$$

Taking into account (4.5) we shall write the inequality (4.3) in the form

$$D(\hat{\theta}) = M(\hat{\theta} - \theta)^2 \geq \frac{1}{M \left[-\frac{\partial^2 \ln L}{\partial \theta^2} \right]}. \quad (4.6)$$

The inequality (4.6) is called the *Rao-Kramer inequality*.

The expression

$$M \left[-\frac{\partial^2 \ln L}{\partial \theta^2} \right] = M \left[\frac{\partial \ln L}{\partial \theta} \right]^2 = I^{(F)}(\theta)$$

is called the Fisher information. The inverse Fisher information is a lower bound for the variance of the estimation $\hat{\theta}$

$$D(\hat{\theta}) \geq (I^{(F)}(\theta))^{-1}.$$

In the case of the vector parameter $\vec{\theta} = (\theta_1, \dots, \theta_s)$, the Rao-Kramer inequality is written as

$$D(\hat{\vec{\theta}}) \geq (I^{(F)}(\vec{\theta}))^{-1},$$

where

$$D_{ss'} = M[(\hat{\theta}_s - \theta_s)(\hat{\theta}_{s'} - \theta_{s'})]$$

is the covariance matrix of the estimation vector parameter and

$$I_{ss'}^{(F)} = M \left[-\frac{\partial^2 \ln L(\vec{x}, \vec{\theta})}{\partial \theta_s \partial \theta_{s'}} \right] = M \left[\frac{\partial \ln L(\vec{x}, \vec{\theta})}{\partial \theta_s} \frac{\partial \ln L(\vec{x}, \vec{\theta})}{\partial \theta_{s'}} \right]$$

is the *Fisher information matrix*. The estimator $\hat{\theta}$ is called *efficient* if the Rao-Kramer inequality passes to the equality

$$D(\hat{\theta}) = [I^{(F)}(\theta)]^{-1}.$$

4.5 Sufficient estimate

If the estimator $\hat{\theta}(x_1, \dots, x_n)$ contains all information on θ , which is available in the observations (x_1, \dots, x_n) , one speaks of a *sufficient estimator*.

The density function $L(\vec{x}, \theta)$ of repeated sampling is represented, in the case of sufficient estimation, in the form:

$$L(\vec{x}, \theta) = g(\hat{\theta}, \theta)h(\vec{x}).$$

4.6 Robust estimator

An estimator is said to be *robust* if its distribution is insensitive to small changes in the distribution of the sampling or more generally, if it is insensitive to small deviations from the assumptions.

5

Models of measurement data

The first step of the geophysical interpretation is a construction of the model of the measurement data. The model is the functional relationship between the observations and the unknown parameters of the geophysical objects or medium. The type of the function follows from the physical law and from the experiment itself. The observation for the estimation of the parameter represents the result of random experiment. The random noise and its properties are very important part of the model.

5.1 Additive models

Let the observations of the geophysical field u be conducted in the discrete point x_k , y_l , z_m at the time t_i :

$$u(x_k, y_l, z_m, t_i), \quad x_k = k\Delta x, \quad y_l = l\Delta y, \quad z_m = m\Delta z, \quad t_i = i\Delta t.$$

Let the unknown vector of parameters of the geophysical field be

$$\vec{\theta} = (\theta_1, \dots, \theta_s),$$

where s is a number of parameters. Let the functional relationship between the observed data u and the unknown parameters $\vec{\theta}$ be described by $f(\vec{\theta}, x_k, y_l, z_m, t_i)$. Let the random noise ε be an additive component of the model, then we can write the additive model in the form:

$$u(x_k, y_l, z_m, t_i) = f(x_k, y_l, z_m, t_i, \vec{\theta}) + \varepsilon(x_k, y_l, z_m, t_i).$$

This model is a basis for the construction of the statistical interpretation.

5.2 Models of the quantitative interpretation

The additive model, introduced above, concerns to the type of models of the *quantitative interpretation*. The problem consist of the estimation of the unknown parameters $\vec{\theta}$ by the mathematical statistic methods.

Example 1: The model of the seismic trace:

$$u(t_i) = \sum_{\mu=1}^m A_\mu \varphi(t_i - \tau_\mu) + \varepsilon(t_i),$$

where $\varphi(t)$ is a shape of the seismic signal, A_μ is an amplitude and τ_μ is an arrival time of the seismic signal with the number μ . The parameter vector has in this case the form

$$\vec{\theta} = \|A_\mu, \tau_\mu\|_{\mu=1}^M.$$

Example of the seismic trace is presented in Figure 5.1a. The trace is a result of numerical simulation for the 1-D wave equation for piecewise-homogeneous (Figure 5.1b) half-space with the free surface. There are three waves in the trace and each wave can be described by the amplitude A_μ and arrival time τ_μ . The continuous on time model can be rewritten in the frequency domain

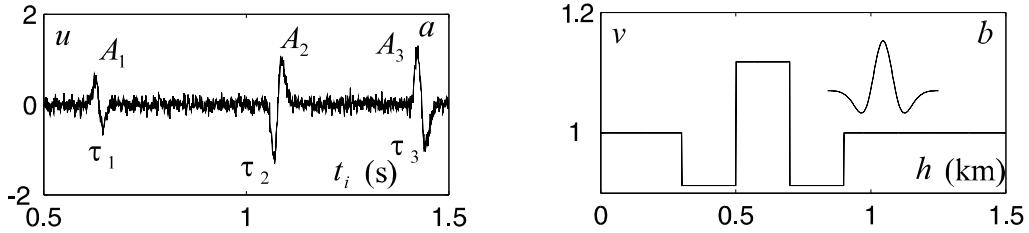


Figure 5.1: Example of seismic trace. (a) – seismic trace with the Gaussian noise ($N(0, 0.1)$); (b) – velocity as a function of depth and time dependence of the source. The source and receiver are located near free surface.

$$u(\omega) = \sum_{\mu=1}^m A_\mu \Phi(\omega) \exp\{-i\omega\tau_\mu\} + E(\omega)$$

using the Fourier transform:

$$u(\omega) = \int_{-\infty}^{\infty} u(t) \exp\{-i\omega t\} dt, \quad \Phi(\omega) = \int_{-\infty}^{\infty} \varphi(t) \exp\{-i\omega t\} dt,$$

$$E(\omega) = \int_{-\infty}^{\infty} \varepsilon(t) \exp\{-i\omega t\} dt.$$

For the discrete model in time domain, we can use Z -transform ($Z = e^{-i\omega\Delta t}$):

$$u(Z) = \sum_{\mu=1}^m A_\mu \Phi(Z) Z^{\tau_\mu / \Delta t} + E(Z).$$

Example 2: Let the vertical component of the magnetic field be registered along the linear profile and it is caused by a source of the type a vertical magnetic dipole, then the model can be written as

$$u(x_k) = \frac{M(2h^2 - x_k^2)}{(x_k^2 + h^2)^{5/2}} + \varepsilon(x_k), \quad (5.1)$$

where M is a magnetic moment of the dipole, h is a depth of the dipole top. In Figure 5.2 the magnetic field is satisfied to a formula (5.1) ($h = 1$ km, $M = 1$).

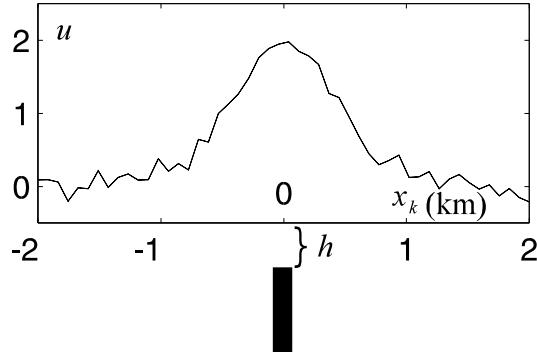


Figure 5.2: The magnetic field together with Gaussian noise ($N(0, \sigma)$) along the linear profile. The field is produced by a formula (5.1) ($h = 1$ km, $M = 1$, $\sigma = 0.1$).

5.3 Regression model

For the fitting of the geophysical observations, it is very often used a *regression* model or a linear model. The linear relation between the parameters and the observations follows generally after a linearization from physical or mathematical laws

$$u(x_k, y_l, z_m, t_i) = \sum_{j=0}^S \theta_j \psi_j(x_k, y_l, z_m, t_i) + \varepsilon(x_k, y_l, z_m, t_i)$$

in the general case.

Example 1: By processing of the two-dimensional magnetic field u , we shall write the model in the form:

$$u(x_k, z_m) = \sum_{j=0}^S \theta_j \psi_j(x_k, z_m) + \varepsilon(x_k, z_m),$$

where θ_j is a magnetic moment of the anomaly with the number j .

Example 2: The model of a seismogram:

$$u(x_k, t_i) = \sum_{j=0}^S \theta_j \psi_j(x_k, t_i) + \varepsilon(x_k, t_i),$$

where θ_j is an amplitude of seismic wave, $\psi_j(x_k, t_i)$ is a shape of seismic wave with unknown time delay.

In practice, it is used very often one-dimensional model for the approximation of geophysical data.

Example 3: The observation of a magnetic field along a linear profile:

$$u(x_k) = \sum_{j=0}^S \theta_j \psi_j(x_k) + \varepsilon(x_k).$$

Example 4: The seismic trace, registered by one geophone:

$$u(t_i) = \sum_{j=0}^S \theta_j \psi_j(t_i) + \varepsilon(t_i).$$

We can rewrite mentioned above linear models in a compact form by vectors and matrices:

$$\vec{u} = \psi \vec{\theta} + \vec{\varepsilon},$$

where

$$\vec{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix}, \quad \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{pmatrix}, \quad \psi = \begin{pmatrix} \psi_{11} & \dots & \psi_{1S} \\ \dots & \dots & \dots \\ \psi_{K1} & \dots & \psi_{KS} \end{pmatrix}$$

and ψ is a plan matrix.

5.4 Models of qualitative interpretation

If after the interpretation of the data it is possible to obtain the information about the qualitative state of the object, then the relevant model of the measurement data is called the *model of the qualitative interpretation*

$$u(x_k, y_l, z_m, t_i) = \begin{cases} \text{either } f_1(x_k, y_l, z_m, t_i) + \varepsilon(x_k, y_l, z_m, t_i), \\ \text{or } f_2(x_k, y_l, z_m, t_i) + \varepsilon(x_k, y_l, z_m, t_i)', \\ \dots & \dots \\ \text{or } f_N(x_k, y_l, z_m, t_i) + \varepsilon(x_k, y_l, z_m, t_i)' \end{cases}$$

where f_ν is a field of the object in the state with the number $\nu = 1, 2, \dots, N$.

Example 1: Detecting of a seismic signal on the background noise:

$$u(t_i) = \begin{cases} \text{either } A\varphi(t_i - \tau) + \varepsilon(t_i,) \\ \text{or } \varepsilon(t_i), \end{cases}$$

where A is a known amplitude, τ is a known arrival time of the seismic signal (see Figure 5.3). The problem consist to detect seismic signal or to say it is only noise.

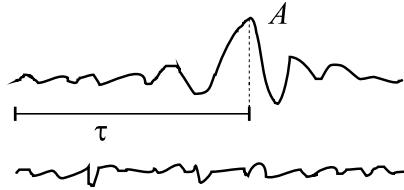


Figure 5.3: Illustration of qualitative interpretation (“noise and signal or noise”).

Example 2: Choice between two seismic signals

$$u(t_i) = \begin{cases} \text{either } A_1\varphi(t_i - \tau_1) + \varepsilon(t_i) \\ \text{or } A_2\varphi(t_i - \tau_2) + \varepsilon(t_i) \end{cases}$$

where A_1 and τ_1 are the amplitude and arrival time of the 1-st wave, A_2 and τ_2 are the parameters of the 2-nd wave. The problem consist to choice between two signals by processing a seismic trace (see Figure 5.4).

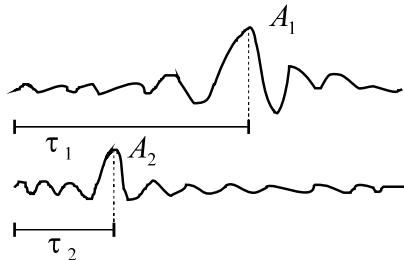


Figure 5.4: Illustration of qualitative interpretation (“signal (number 1) and noise or signal (number 2) and noise”).

Example 3: Resolution of seismic signals in time domain:

$$u(t_i) = \begin{cases} \text{either } A_0\varphi_0(t_i - \tau_0) + \varepsilon(t_i) \\ \text{or } A_1\varphi_1(t_i - \tau_1) + A_2\varphi_2(t_i - \tau_2) + \varepsilon(t_i). \end{cases}$$

The problem consists to decide: we have on the interval of the seismic trace only one wave with parameters (A_0, τ_0) or two waves with parameters (A_1, τ_1) and (A_2, τ_2) (see Figure 5.5).

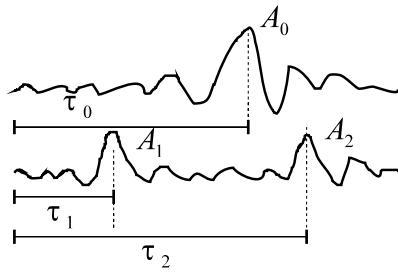


Figure 5.5: Illustration of qualitative interpretation (“signal and noise or two signals and noise”).

Example 4: Resolution of seismic signals in frequency domain (see Figure 5.6):

$$u(\omega) = \begin{cases} \text{either} & A_0\Phi_0(\omega - \omega_0) + E(\omega) \\ \text{or} & A_1\Phi_1(\omega - \omega_1) + A_2\Phi_2(\omega - \omega_2) + E(\omega). \end{cases}$$

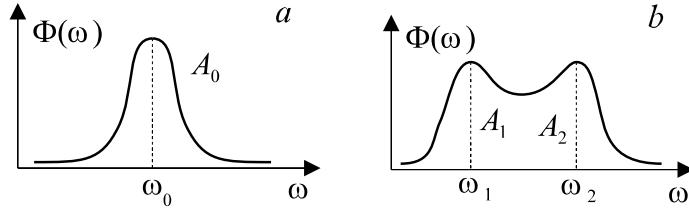


Figure 5.6: Illustration of qualitative interpretation in frequency domain (“signal and noise or two signals and noise”).

5.5 Models of qualitative-quantitative interpretation

By the solution of complicated interpretation problems, the geophysicist deals with the models of the type:

$$u(x_k, y_l, z_m, t_i) = \begin{cases} \text{either} & f_1(x_k, y_l, z_m, t_i, \vec{\theta}_1) + \varepsilon(x_k, y_l, z_m, t_i), \\ \text{or} & f_2(x_k, y_l, z_m, t_i, \vec{\theta}_2) + \varepsilon(x_k, y_l, z_m, t_i), \\ \dots & \dots \\ \text{or} & f_N(x_k, y_l, z_m, t_i, \vec{\theta}_N) + \varepsilon(x_k, y_l, z_m, t_i), \end{cases} \quad \nu = 1, 2, \dots, N$$

where the state of the object is characterized by N different deterministic components or signals f_ν , each component depends on an own vector $\vec{\theta}_\nu$ of unknown parameters. The problem consists of searching of an optimal procedure of the choice between the possible states and in the estimation of unknown parameters. We note that by the model construction it is necessary to give the properties of the random component.

5.6 Random component models and its properties

The representation of the random component properties determines the structure of the estimation algorithm. Therefore it is necessary to analyze carefully the properties of the random component.

5.6.1 The normal distribution

Taking into account the physical nature of random noise and using the central limit theorem, we can for many practical interesting cases make an assumption about *normal distribution* of the random component

$$\vec{\varepsilon} \in N(\vec{\varepsilon}_0, R).$$

The density of the normally distributed $\vec{\varepsilon}$ is

$$p(\vec{\varepsilon}) = (2\pi)^{-n/2} |R|^{-1/2} \exp\left\{-\frac{1}{2}(\vec{\varepsilon} - \vec{\varepsilon}_0)^T R^{-1} (\vec{\varepsilon} - \vec{\varepsilon}_0)\right\},$$

where the elements of the covariance matrix are

$$R_{ii'} = M[(\varepsilon_i - \varepsilon_{0i})(\varepsilon_{i'} - \varepsilon_{0i'})]$$

and elements of the correlation matrix are

$$r_{ii'} = \frac{R_{ii'}}{\sqrt{R_{ii}}\sqrt{R_{i'i'}}}.$$

5.6.2 The random components is a stationary process

The correlation matrix will be written in the form:

$$r = \begin{bmatrix} 1 & r_1 & r_2 & \dots & r_n \\ \dots & \dots & \dots & \dots & \dots \\ r_n & r_{n-1} & r_{n-2} & \dots & 1 \end{bmatrix}.$$

5.6.3 The random component is a non-correlated process

The elements of the covariance matrix are written in the form:

$$R_{ii'} = \sigma_i^2 \delta_{ii'}, \quad \delta_{ii'} = \begin{cases} 1 & \text{by } i = i', \\ 0 & \text{by } i \neq i', \end{cases}$$

$$R = \sigma^2 E, \quad R^{-1} = \frac{1}{\sigma^2} E,$$

where E is an identity matrix. The normal density has the following presentation:

$$p(\vec{\varepsilon}) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\vec{\varepsilon} - \vec{\varepsilon}_0)^2\right\}. \quad (5.2)$$

5.6.4 The random component is Markov process

The elements of the covariance matrix in the case of the Markov process will be written as

$$R_{ii'} = \sigma^2 \exp\{-c|i - i'|\} \quad (5.3)$$

(see Figure 5.7). Introducing a notation

$$\gamma = \exp(-c)$$

we can write the covariance matrix

$$R = \sigma^2 \begin{bmatrix} 1 & \gamma & \gamma^2 & \dots & \gamma^n \\ \gamma & 1 & \gamma & \dots & \gamma^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ \gamma^n & \gamma^{n-1} & \gamma^{n-2} & \dots & 1 \end{bmatrix}$$

The determinant R is equal to

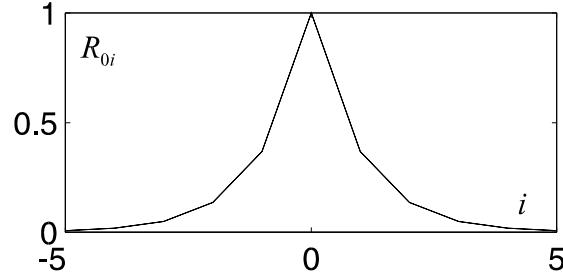


Figure 5.7: The elements of the covariance matrix in the case of Markov process (5.3) ($\sigma = 1$, $c = 1$).

$$|R| = \sigma^{2n} (1 - \gamma^2)^{n-1}.$$

The inverse matrix is represented in the form

$$R^{-1} = \frac{1}{\sigma^2(1 - \gamma^2)} \begin{bmatrix} 1 & -\gamma & 0 & \dots & 0 \\ -\gamma & 1 + \gamma^2 & -\gamma & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

The normal density for the Markov process is

$$\begin{aligned} p(\vec{\varepsilon}) &= (2\pi\sigma^2)^{-n/2} (1 - \gamma^2)^{(n-1)/2} \exp\left\{-\frac{1}{2\sigma^2(1 - \gamma^2)} \times \right. \\ &\times \left. [(1 + \gamma^2) \sum_{i=1}^n \varepsilon_i^2 - 2\gamma \sum_{i=1}^n \varepsilon_i \varepsilon_{i+1} - \gamma^2 (\varepsilon_1^2 + \varepsilon_n^2)]\right\}. \end{aligned}$$

5.6.5 The Laplace distribution

The *Laplace distribution*

$$p(\varepsilon) = \frac{1}{2\alpha} \exp \left\{ -\frac{|\varepsilon - \varepsilon_0|}{\alpha} \right\}$$

is more effective than the normal distribution for the description of the random component in the case of the presence of outliers. The “tails” of the Laplace distribution falls down slower than the “tails” of the normal distribution (5.2) (see Figure 5.8).

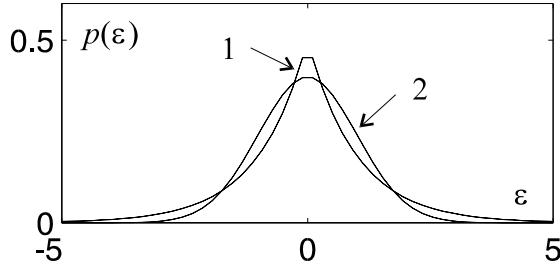


Figure 5.8: Laplace and normal distributions. 1 – Laplace distribution ($\varepsilon_0 = 0$, $\alpha = 1$); 2 – normal distribution ($\varepsilon_0 = 0$, $\sigma = 1$).

5.7 Model with random parameter

Let the parameters θ be a random variables, then the model can be written in the form

$$u(x_k, t_i) = \sum_{\mu=1}^M A_\mu \varphi(t_i - \tau_{\mu k}) + \varepsilon(x_k, t_i),$$

where A_μ and τ_μ are random amplitudes and random arrival times respectively with the number μ . The shape of the signal $\varphi(t_i)$ is deterministic. By the solution of the practical geophysical problems it is often assumed that the random parameters have the normal distribution:

$$A \in N(0, \sigma_A^2), \quad \tau \in N(0, \sigma_\tau^2), \quad \varepsilon \in N(0, \sigma_\varepsilon^2).$$

5.8 A priori information about field of parameters

By a priori information (or prior information) we mean information which is obtained independently of the results of measurements.

Generally, a priori information can be represented by an algebraic system

$$\vec{\Phi}(\vec{\theta}) = \vec{c}$$

or an inequality

$$\vec{c}_1 \leq \vec{\Phi}(\vec{\theta}) \leq \vec{c}_2,$$

where $\vec{\theta}$ is a vector of unknown parameters, $\vec{\Phi}(\vec{\theta})$ is a given vector-function.

For the practical application it is possible to be restricted to a linear system

$$A\vec{\theta} = \vec{V}$$

or linear system of the inequality system

$$\vec{V}_1 \leq A\vec{\theta} \leq \vec{V}_2,$$

where \vec{V}_1 and \vec{V}_2 are the given lower and upper bounds.

If we do not know anything about an unknown parameter θ , it may take values from $-\infty$ to $+\infty$. Hence, the appropriate prior probability density function is

$$p(\theta) = \text{const} \quad \text{with} \quad -\infty < \theta < \infty.$$

This is an improper density function, because $\int_{-\infty}^{\infty} p(\theta)d\theta \neq 1$.

It was mentioned that entropy is a measure of uncertainty. By applying the principle of maximum entropy may select distributions which contain the largest amount of uncertainty compatible with given constraints. Since prior information is generally incomplete, beyond the given information the prior information should be as uncertain or non-informative as possible.

We use now the principle of maximum entropy to derive distributions, which contain maximum uncertainty compatible with the constraints:

1. The density function of a random vector $\vec{\theta} = \{\theta_1, \dots, \theta_s\}$, which components are defined on the intervals $(\theta'_s - \theta''_s)$ and which maximizes the entropy is the density of the *uniform distribution*

$$p(\vec{\theta}) = \begin{cases} \prod_{s=1}^S 1/(\theta'_s - \theta''_s) & \theta''_s \leq \theta_s \leq \theta'_s, \\ 0 & \theta_s. \end{cases}$$

2. The density function of a random vector $\vec{\theta} = \{\theta_1, \dots, \theta_s\}$ with the expected vector $\langle \vec{\theta} \rangle$ and the variances $\sigma_{\theta_s}^2$, which components are defined on the interval $(-\infty, \infty)$ and which maximizes the entropy, is the density of the *normal distribution*

$$p(\vec{\theta}) = \prod_{s=1}^S (2\pi\sigma_{\theta_s}^2)^{-1/2} \exp\left\{-\frac{1}{2} \sum_{s=1}^S \frac{(\theta_s - \langle \theta_s \rangle)^2}{\sigma_{\theta_s}^2}\right\}.$$

Example: We consider the representation of the estimate of a seismic signal parameters

$$f(A, \tau) = A\varphi(t_i - \tau).$$

We assume that the amplitude A has a normal distribution with the expected value $\langle A \rangle$ and the variance σ_A^2 , and the arrival time τ has the uniform distribution. The prior information is given by

$$p(A, \tau) = \begin{cases} [\sqrt{2\pi}\sigma_A(\tau' - \tau'')]^{-1} \times & \text{by } \tau'' \leq \tau \leq \tau', \\ \times \exp\{-(A - \langle A \rangle)^2/2\sigma_A^2\} & \\ 0 & \text{for others } \tau. \end{cases}$$

6

Solution of the inverse geophysical problems: case of the linear model

Let the geophysical observation, which contain the information on the unknown parameters $\vec{\theta}$, be collected in the vector \vec{u} . The unknown parameters contained in $\vec{\theta}$ will be estimated by means of the estimate function $\lambda(\vec{u}, \vec{\theta})$. We consider now the case of the linear model of the measurement data

$$\vec{u} = \psi\vec{\theta} + \vec{\varepsilon}.$$

We begin from the least squared method.

6.1 Method of least squares (MLS)

The *method of least squares* of estimating unknown parameters $\vec{\theta}$ is given by minimizing the sum of weighted squares

$$\begin{aligned}\lambda(\vec{\theta}) &= (\vec{u} - \psi\vec{\theta})^T W (\vec{u} - \psi\vec{\theta}), \\ \hat{\vec{\theta}} &= \arg \min \lambda(\vec{\theta}),\end{aligned}$$

where W is a diagonal weighted function $W = \text{diag}(w_1, \dots, w_n)$. To find the minimum, λ is differentiated with respect to each of the parameters and each resulting equation is set equal to zero

$$\frac{\partial \lambda}{\partial \theta_s} = 0 \quad \Rightarrow \quad (\psi^T W \psi) \vec{\theta} = \psi^T W \vec{u}.$$

We can interpret the weight function as an inverse covariance matrix of the random component $\vec{\varepsilon}$

$$W = R_\varepsilon^{-1} \quad \text{or} \quad W = \sigma_\varepsilon^{-2} I,$$

where σ_ε^{-2} is the variance of the random component. In this case the estimate follows by

$$\hat{\vec{\theta}} = (\psi^T \psi)^{-1} \psi^T \vec{u}.$$

The covariance matrix of the estimation $\hat{\vec{\theta}}$ can be represented in the form

$$\begin{aligned} R_{\theta} &= M[(\hat{\vec{\theta}} - \vec{\theta})(\hat{\vec{\theta}} - \vec{\theta})^T] = M[(\psi^T \psi)^{-1} \psi^T \vec{u} - \vec{\theta}] [\vec{u}^T \psi (\psi^T \psi)^{-1} - \vec{\theta}^T] = \\ &= M[(\psi^T \psi)^{-1} (\psi^T \psi) \vec{\theta} - \vec{\theta} + (\psi^T \psi)^{-1} \psi^T \vec{\varepsilon}] [(\psi^T \psi) (\psi^T \psi)^{-1} \vec{\theta}^T - \vec{\theta}^T + \vec{\varepsilon}^T \psi (\psi^T \psi)^{-1}] \\ &= (\psi^T \psi)^{-1} \psi^T M[\vec{\varepsilon} \vec{\varepsilon}^T] \psi (\psi^T \psi)^{-1} = \sigma_{\varepsilon}^2 (\psi^T \psi)^{-1}. \end{aligned}$$

6.1.1 Properties of the MLS-estimation

1. Unbiased estimation $\hat{\theta}_{MLS}$:

$$\begin{aligned} M[\hat{\vec{\theta}}] &= (\psi^T \psi)^{-1} \psi^T M[\vec{u}] = \\ &= (\psi^T \psi)^{-1} \psi^T M[\psi \vec{\theta} + \vec{\varepsilon}] = \vec{\theta}. \end{aligned}$$

It is due to $M[\vec{\varepsilon}] = 0$.

2. The efficient estimation $\hat{\theta}_{MLS}$:

$$R_{ss}(\hat{\vec{\theta}}) \leq R_{ss}(\vec{\theta}^*) \quad s = 1, \dots, S,$$

where $\vec{\theta}^*$ is an arbitrary estimation. The MLS-estimation has the minimum variance of the estimation in the class of the linear estimation.

3. The mean squared error of the observation vector \vec{u} and the estimated $\hat{\vec{u}} = \psi \hat{\vec{\theta}}$ is determined by

$$\begin{aligned} \hat{\sigma}_{\varepsilon}^2 &\Rightarrow M[(\vec{u} - \psi \hat{\vec{\theta}})^T (\vec{u} - \psi \hat{\vec{\theta}})] = \sigma_{\varepsilon}^2 (n - S) \Rightarrow \\ &\Rightarrow \hat{\sigma}_{\varepsilon}^2 = (\vec{u} - \psi \hat{\vec{\theta}})^T (\vec{u} - \psi \hat{\vec{\theta}}) / (n - S) = \hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}} / (n - S). \end{aligned}$$

Example: To find amplitudes of seismic waves with known arrival times τ_s .

The model of the seismic trace is

$$\vec{u} = \psi \vec{A} + \vec{\varepsilon},$$

where $\psi = ||\varphi(t_i - \tau_s)||_{n \times S}$ is a matrix, $\vec{\varepsilon}$ is a normal distributed uncorrelated random noise. The amplitudes MLS-estimation is determined by

$$\hat{\vec{A}} = (\psi^T \psi)^{-1} \psi^T \vec{u}$$

and the covariance matrix of the estimation $\hat{\vec{A}}$ is

$$R_{\hat{A}} = \hat{\sigma}_{\varepsilon}^2 (\psi^T \psi)^{-1},$$

where

$$\hat{\sigma}_{\varepsilon}^2 = \frac{(\vec{u} - \psi \hat{\vec{A}})^T (\vec{u} - \psi \hat{\vec{A}})}{n - S}$$

is the mean squared error.

6.2 Method of least squares: non-linear case

In general, there will be no linear relations between the parameters $\vec{\theta}$ and the observation \vec{u}

$$\vec{u} = \vec{\varphi}(\vec{\theta}) + \vec{\varepsilon},$$

where $\varphi_i(\vec{\theta})$ denote differentiable functions of the unknown parameters.

We can linearize functions $\varphi_i(\vec{\theta})$ by the Taylor expansion. We get with the reference parameters $\vec{\theta}_0$:

$$\vec{\varphi}(\vec{\theta}) = \vec{\varphi}(\vec{\theta}_0) + \psi \Delta \vec{\theta} + \dots$$

We can rewrite the model in the form

$$\tilde{\vec{u}} = \psi \Delta \vec{\theta} + \vec{\varepsilon},$$

where $\tilde{\vec{u}} = \vec{u} - \vec{\varphi}(\vec{\theta}_0)$, ψ is a matrix

$$\|\psi_{js}\| = \partial \varphi_j / \partial \theta_s |_{\vec{\theta}=\vec{\theta}_0}$$

and $\Delta \vec{\theta} = \vec{\theta} - \vec{\theta}_0$.

The estimates of the parameters are used to compute iteratively new approximate values.

Let the estimate of the first iteration will be

$$\Delta \hat{\vec{\theta}}^{(1)} = (\psi_0^T W_0 \psi_0)^{-1} \psi_0^T W_0 \tilde{\vec{u}}_0, \quad \hat{\vec{\theta}}^{(1)} = \vec{\theta}_0 + \Delta \hat{\vec{\theta}}^{(1)},$$

where W_0 is a diagonal weight matrix

$$W_0 = (\hat{\sigma}_{\varepsilon 0}^2)^{-1} I$$

and the mean squared error is

$$\hat{\sigma}_{\varepsilon 0}^2 = \frac{(\vec{u} - \psi_0 \vec{\theta}_0)^T (\vec{u} - \psi_0 \vec{\theta}_0)}{n - S}.$$

The estimation of the second iteration will be

$$\Delta \hat{\vec{\theta}}^{(2)} = (\psi_1^T W_1 \psi_1)^{-1} \psi_1^T W_1 \tilde{\vec{u}}_1,$$

where $\tilde{\vec{u}}_1 = \vec{u} - \psi_1 \hat{\vec{\theta}}^{(1)}$ is a new data, the matrix ψ_1 is

$$\|\psi_{js}^{(1)}\| = \left. \frac{\partial \varphi_j}{\partial \theta_s} \right|_{\vec{\theta}=\hat{\vec{\theta}}_1}$$

and the weight matrix of the second iteration is

$$W_1 = (\hat{\sigma}_{\varepsilon 1}^2)^{-1} I = \left[\frac{(\vec{u} - \psi_0 \vec{\theta}_1)^T (\vec{u} - \psi_0 \vec{\theta}_1)}{n - S} \right]^{-1} I.$$

Then we test the threshold. If the inequality is satisfied

$$\left| \frac{\Delta \hat{\theta}_s^{(2)}}{\hat{\theta}_s^{(2)}} \right| \leq \beta,$$

the iteration procedure is finished, in the opposite case we continue the iteration procedure.

For example, we can use this procedure to estimate the travel times τ_s of the signals in the model

$$\vec{u} = \psi \vec{A} + \vec{\varepsilon},$$

(see 6.1) where $\psi = \|\varphi(t_i - \tau_s)\|$ is a matrix $(n \times S)$.

6.3 Method of least squares: orthogonal polynomials

The polynomial model often serves the purpose to fit a curve or a surface. For example, in the case of a linear model

$$\vec{u} = \psi \vec{\theta} + \vec{\varepsilon}$$

the elements of matrix ψ can be $\psi_{is} = x_i^s$. It is a very important special case of the orthogonal polynomials.

We can write the i -th row of the matrix ψ as

$$\psi_i = \|\varphi_0(x_i), \varphi_1(x_i), \dots, \varphi_{S-1}(x_i)\|.$$

The orthogonality condition is written by

$$\sum_{i=1}^n \varphi_s(x_i) \varphi_{s'}(x_i) = \begin{cases} \sum_{i=1}^n \varphi_s^2(x_i) & \text{for } s = s' \\ 0 & \text{for } s \neq s' \end{cases}$$

and $\psi^T \psi$ follows by

$$\psi^T \psi = \begin{vmatrix} \sum_{i=1}^n \varphi_0^2(x_i) & 0 & \dots & 0 \\ 0 & \sum_{i=1}^n \varphi_1^2(x_i) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sum_{i=1}^n \varphi_{S-1}^2(x_i) \end{vmatrix}.$$

In the special case of the Chebyshev polynomial, we get

$$\varphi_0(x_i) = 1, \quad \varphi_1(x_i) = x_i + b_0 \varphi_0(x_i).$$

To find the coefficient b_0 we multiply on the $\varphi_0(x_i)$ and sum up

$$b_0 \Rightarrow \sum_{i=1}^n \varphi_1(x_i) \varphi_0(x_i) = \sum_{i=1}^n x_i \varphi_0(x_i) + b_0 \sum_{i=1}^n \varphi_0^2(x_i).$$

Taking into account that

$$\sum_{i=1}^n \varphi_1(x_i) \varphi_0(x_i) = 0,$$

we get

$$b_0 = -\frac{\sum_{i=1}^n x_i \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)},$$

and we can write $\varphi_1(x_i)$ in the explicit form

$$\varphi_1(x_i) = x_i - \frac{\sum_{i=1}^n x_i \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)} \varphi_0(x_i).$$

For the function $\varphi_2(x_i)$ we get

$$\varphi_2(x_i) = x_i^2 + b_1 \varphi_1(x_i) + b_0 \varphi_0(x_i).$$

The coefficients b_0 and b_1 we find using the orthogonality conditions

$$\begin{aligned} \sum_{i=1}^n \varphi_2(x_i) \varphi_0(x_i) &= 0, & b_0 &= -\frac{\sum_{i=1}^n x_i^2 \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)}, \\ \sum_{i=1}^n \varphi_2(x_i) \varphi_1(x_i) &= 0, & b_1 &= -\frac{\sum_{i=1}^n x_i^2 \varphi_1(x_i)}{\sum_{i=1}^n \varphi_1^2(x_i)}. \end{aligned}$$

Substituting the expression for b_0 and b_1 in the function $\varphi_2(x_i)$, we get

$$\varphi_2(x_i) = x_i^2 - \frac{\sum_{i=1}^n x_i^2 \varphi_1(x_i)}{\sum_{i=1}^n \varphi_1^2(x_i)} \varphi_1(x_i) - \frac{\sum_{i=1}^n x_i^2 \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)} \varphi_0(x_i).$$

For the function $\varphi_s(x_i)$ we get

$$\varphi_s(x_i) = x_i^s - \frac{\sum_{i=1}^n x_i^s \varphi_{s-1}(x_i)}{\sum_{i=1}^n \varphi_{s-1}^2(x_i)} \varphi_{s-1}(x_i) - \dots - \frac{\sum_{i=1}^n x_i^s \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)} \varphi_0(x_i),$$

$$\varphi_s(x_i) = x_i^s + b_{s-1}\varphi_{s-1}(x_i) + \dots + b_0\varphi_0(x_i).$$

The coefficient b_q ($q < s$) is determined by taking into account the orthogonality conditions

$$\sum_{i=1}^n x_i^s \varphi_q(x_i) + b_q \sum_{i=1}^n \varphi_q^2(x_i) = 0, \quad b_q = -\frac{\sum_{i=1}^n x_i^s \varphi_q(x_i)}{\sum_{i=1}^n \varphi_q^2(x_i)}.$$

The estimate of the parameter $\hat{\theta}_s$ is determined by

$$\begin{aligned} \hat{\theta}_s &= \sum_{j=1}^n \varphi_{sj} u_j, \quad s = 0, 1, 2, \dots, S-1, \\ \hat{\vec{\theta}} &= \Phi \vec{u}, \end{aligned}$$

where the matrix Φ is written by

$$\Phi = \left\| \begin{array}{cccc} \varphi_0(x_1) & \varphi_0(x_2) & \dots & \varphi_0(x_n) \\ \varphi_1(x_1) & \varphi_1(x_2) & \dots & \varphi_1(x_n) \\ \dots & \dots & \dots & \dots \\ \varphi_{S-1}(x_1) & \varphi_{S-1}(x_2) & \dots & \varphi_{S-1}(x_n) \end{array} \right\|$$

under condition of the orthonormalization

$$\sum_{i=1}^n \varphi_{si}^2 = 1.$$

Let the covariance matrix of the random component be $R_\varepsilon = \sigma_\varepsilon^2 I$, then we get

$$R_\theta = \sigma_\varepsilon^2 I.$$

The estimate of the mean squared error is

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-S} (\vec{u} - \psi \hat{\vec{\theta}})^T (\vec{u} - \psi \hat{\vec{\theta}}).$$

6.4 Method of least squares with constraints

The unknown parameters of the linear model

$$\vec{u} = \psi \vec{\theta} + \vec{\varepsilon}$$

are now a subject to linear constraints

$$A \vec{\theta} = \vec{V},$$

where A denotes a $[K \times S]$ matrix of known coefficients, and V a known $[K \times 1]$ vector.

The best unbiased estimation $\tilde{\vec{\theta}}$ of the vector $\vec{\theta}$ of unknown parameters will be determined by the method of the Lagrange multipliers.

We shall construct the function

$$\Phi(\vec{\theta}) = (\vec{u} - \psi\vec{\theta})^T(\vec{u} - \psi\vec{\theta}) - 2\vec{\lambda}^T(A\vec{\theta} - \vec{V}),$$

where $\vec{\lambda}$ is a $(K \times 1)$ vector of the Lagrange multipliers. We minimize this function differentiation on the parameters θ_s

$$\begin{aligned}\partial\Phi/\partial\theta_s &= 0, \quad s = 1, \dots, S, \\ -2\vec{u}^T\psi + 2\vec{\theta}^T\psi^T\psi - 2\vec{\lambda}^TA &= 0 \mid (\psi^T\psi)^{-1}.\end{aligned}$$

We multiply by the matrix $(\psi^T\psi)^{-1}$ and using the MLS-estimation $\hat{\vec{\theta}} = (\psi^T\psi)^{-1}\psi^T\vec{u}$, we get

$$\tilde{\vec{\theta}}^T = \hat{\vec{\theta}}^T + \vec{\lambda}^TA(\psi^T\psi)^{-1} \mid A^T.$$

We multiply by the matrix A^T and we get

$$\tilde{\vec{\theta}}^T A^T = \hat{\vec{\theta}}^T A^T + \vec{\lambda}^T A(\psi^T\psi)^{-1} A^T.$$

Taking into account that the estimate $\tilde{\vec{\theta}}$ has to satisfy the linear constraints

$$\tilde{\vec{\theta}}^T A^T = V^T,$$

we shall find the vector of the multipliers

$$\vec{\lambda}^T = (\vec{V}^T - \hat{\vec{\theta}}^T A^T)(A(\psi^T\psi)^{-1} A^T)^{-1}$$

and we get finally the estimate

$$\boxed{\tilde{\vec{\theta}}^T = \hat{\vec{\theta}}^T + (\vec{V}^T - \hat{\vec{\theta}}^T A^T)(A(\psi^T\psi)^{-1} A^T)^{-1} A(\psi^T\psi)^{-1}}.$$

As we shall show below the covariance matrix $R_{\tilde{\vec{\theta}}}$ of the estimate $\tilde{\vec{\theta}}$ can be written

$$\boxed{R_{\tilde{\vec{\theta}}} = R_{\hat{\vec{\theta}}} - \sigma_\varepsilon^2(\psi^T\psi)^{-1} A^T (A(\psi^T\psi)^{-1} A^T)^{-1} A(\psi^T\psi)^{-1}},$$

where

$$R_{\hat{\vec{\theta}}} = \hat{\sigma}_\varepsilon^2(\psi^T\psi)^{-1}, \quad \hat{\sigma}_\varepsilon^2 = \frac{1}{n - S - K}(\vec{u} - \hat{\vec{\theta}})^T(\vec{u} - \hat{\vec{\theta}}).$$

We write the definition of the covariance matrix $R_{\tilde{\vec{\theta}}}$

$$\begin{aligned}R_{\tilde{\vec{\theta}}} &= M[(\tilde{\vec{\theta}} - \vec{\theta})(\tilde{\vec{\theta}} - \vec{\theta})^T] = M[(\hat{\vec{\theta}} - \vec{\theta} + (\psi^T\psi)^{-1} A^T \times \\ &\times (A(\psi^T\psi)^{-1} A^T)^{-1}(\vec{V} - A\hat{\vec{\theta}}))(\hat{\vec{\theta}}^T - \vec{\theta}^T + (\vec{V}^T - \hat{\vec{\theta}}^T A^T) \times \\ &\times (A(\psi^T\psi)^{-1} A^T)^{-1} A(\psi^T\psi)^{-1})]\end{aligned}$$

or introducing notations $B = (\psi^T \psi)^{-1}$ and $C = (A(\psi^T \psi)^{-1} A^T)^{-1}$ we get

$$R_{\tilde{\theta}} = [(\hat{\vec{\theta}} - \vec{\theta} + BA^T C(\vec{V} - A\hat{\vec{\theta}})) \times (\hat{\vec{\theta}}^T - \vec{\theta}^T + (\vec{V}^T - \hat{\vec{\theta}}^T A^T) CAB)]$$

Taking into account that

$$\begin{aligned} R_{\tilde{\theta}} &= M[(\hat{\vec{\theta}} - \vec{\theta})(\hat{\vec{\theta}} - \vec{\theta})^T] = M[(B\psi^T(\psi\vec{\theta} + \vec{\varepsilon}) - \vec{\theta}) \times \\ &\quad \times (B\psi^T(\psi\vec{\theta} + \vec{\varepsilon}) - \vec{\theta})^T] = BM[\vec{\varepsilon}\vec{\varepsilon}^T] = \sigma_\varepsilon^2 B \end{aligned}$$

and

$$\begin{aligned} M[(\vec{V} - AB\psi^T(\psi\vec{\theta} + \vec{\varepsilon}))(B\psi^T(\psi\vec{\theta} + \vec{\varepsilon}) - \vec{\theta})^T] &= \\ &= M[(\vec{V} - A\vec{\theta} - AB\psi^T\vec{\varepsilon})(\vec{\theta} + B\psi^T\vec{\varepsilon} - \vec{\theta})^T] = \\ &= AB\psi^T M[\vec{\varepsilon}\vec{\varepsilon}^T] \psi B = \sigma_\varepsilon^2 AB \end{aligned}$$

and

$$\begin{aligned} M[(\vec{V} - AB\psi^T(\psi\vec{\theta} + \vec{\varepsilon}))(\vec{V}^T - (\vec{\varepsilon}^T + \vec{\theta}^T \psi^T) \psi B A^T)] &= \\ &= AB\psi^T M[\vec{\varepsilon}\vec{\varepsilon}^T] \psi B A^T = \sigma_\varepsilon^2 A B A^T. \end{aligned}$$

Substituting these expressions into $R_{\tilde{\theta}}$ we get

$$R_{\tilde{\theta}} = R_{\hat{\vec{\theta}}} - 2\sigma_\varepsilon^2 B A^T C A B + \sigma_\varepsilon^2 B A^T \underbrace{C A B A^T}_{=I} C A B = R_{\hat{\vec{\theta}}} - \sigma_\varepsilon^2 B A^T C A B,$$

as to be showed.

This method has very important application for solution of geophysical problems. For example, taking into account the borehole observations of the p -wave velocity by acoustic logging (\vec{V}) as linear constraints, we approximate the p -wave velocity obtained by surface seismic observations (\vec{u}).

6.5 Maximum likelihood method (MLM)

The method of least squares do not need any statements on the kind of distribution of the measurement vector \vec{u} .

When estimating parameters by the *maximum-likelihood method*, however, the density function of the observations has to be specified.

We identify the probability of the data given the parameters, as the *likelihood* of the parameters given the data.

Let the random vector \vec{u} of the geophysical observations have the density function $f(\vec{x}, \vec{\theta})$, depending on the unknown and fixed parameters $\vec{\theta}$ then the *likelihood function* $L(\vec{u}, \vec{\theta})$ is defined by

$$L(\vec{u}, \vec{\theta}) = f(\vec{x}, \vec{\theta}).$$

The estimate of the fixed parameters $\vec{\theta}$, which maximizes $L(\vec{u}, \vec{\theta})$ or $\ln L(\vec{u}, \vec{\theta})$ is called *maximum likelihood estimate*

$$\hat{\vec{\theta}}_{MLM} = \arg \max \ln L(u_1, \dots, u_n; \vec{\theta}).$$

To find an extremum we differentiate $\ln L$ on parameters θ_s

$$\frac{\partial \ln L(u_1, \dots, u_n; \vec{\theta})}{\partial \theta_s} = 0.$$

This expression is called maximum likelihood equation.

Properties of the maximum likelihood estimation.

1. $\hat{\vec{\theta}}_{MLM}$ is a *consistent estimate*.
2. $\hat{\vec{\theta}}_{MLM}$ is an *asymptotically unbiased estimate* $M[\hat{\vec{\theta}}_n] \xrightarrow{n \rightarrow \infty} \vec{\theta}$.
3. $\hat{\vec{\theta}}_{MLM}$ is an *asymptotically normal estimate* $\sqrt{n}(\hat{\vec{\theta}}_n - \vec{\theta}) \xrightarrow{n \rightarrow \infty} N(0, I^{(F)}(\vec{\theta}))$, where $I^{(F)}(\vec{\theta})$ is the Fisher information matrix.
4. $\hat{\vec{\theta}}_{MLM}$ is an *asymptotically efficient estimate*

$$\sigma_{\hat{\vec{\theta}}_n}^2 \xrightarrow{n \rightarrow \infty} [I^{(F)}(\vec{\theta}_n)]^{-1}$$

and in the case of the vector parameter

$$R(\hat{\vec{\theta}}) \xrightarrow{n \rightarrow \infty} [I^{(F)}(\hat{\vec{\theta}}_n)]^{-1}.$$

Example 1: In the case of the *arbitrary additive model* of geophysical data with a *normal distribution* of a random component, we shall get

$$\vec{u} = \vec{f}(\vec{\theta}) + \vec{\varepsilon}, \quad \vec{u} = (u_1, \dots, u_n), \quad \vec{\varepsilon} \in N(0, R_\varepsilon).$$

We write down the logarithm of the likelihood function

$$l(\vec{u}, \vec{\theta}) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln R_\varepsilon - \frac{1}{2} (\vec{u} - \vec{f}(\vec{\theta}))^T R_\varepsilon^{-1} (\vec{u} - \vec{f}(\vec{\theta})),$$

omitting terms independent of parameters we get

$$l_1(\vec{u}, \vec{\theta}) = -\frac{1}{2} (\vec{u} - \vec{f}(\vec{\theta}))^T R_\varepsilon^{-1} (\vec{u} - \vec{f}(\vec{\theta})).$$

In the special case, if

$$R_\varepsilon = \sigma_\varepsilon^2 I,$$

we obtain

$$l_1(\vec{u}, \vec{\theta}) = -\frac{1}{2\sigma_\varepsilon^2} (\vec{u} - \vec{f}(\vec{\theta}))^T (\vec{u} - \vec{f}(\vec{\theta})).$$

We see that the maximization of the likelihood function is equivalent to the minimization of the squared form

$$(\vec{u} - \vec{f}(\vec{\theta}))^T (\vec{u} - \vec{f}(\vec{\theta})),$$

i.e. in the case of the normal distribution of random component ε the maximum-likelihood passes to the least squares method. The maximum-likelihood equation follows

$$(\vec{u} - \vec{f}(\vec{\theta}))^T R_\varepsilon^{-1} \frac{\partial \vec{f}}{\partial \theta_s} = 0, \quad s = 1, 2, \dots, S.$$

Example 2: In the case of the *linear model* with the *normal distribution*

$$\vec{u} = \psi \vec{\theta} + \vec{\varepsilon}, \quad \vec{\varepsilon} \in N(0, R_\varepsilon)$$

we can write the likelihood function as

$$l_1(\vec{u}, \vec{\theta}) = -\frac{1}{2}(\vec{u} - \psi \vec{\theta})^T R_\varepsilon^{-1} (\vec{u} - \psi \vec{\theta})$$

and maximum likelihood equations as

$$(\vec{u} - \psi \vec{\theta})^T R_\varepsilon^{-1} \psi = 0.$$

The MLM-estimate is written in the explicit form

$$\hat{\vec{\theta}}_{MLM} = (\psi^T R_\varepsilon^{-1} \psi)^{-1} \psi^T R_\varepsilon^{-1} \vec{u}$$

and the covariance matrix of estimate asymptotically is

$$R_{\hat{\vec{\theta}}} = \sigma^2 (\psi^T \psi)^{-1}.$$

In this case MLM-estimate is equal to LSM-estimate.

Example 3: In the case of *linear model* with the Laplace distribution

$$\vec{u} = \psi \vec{\theta} + \vec{\varepsilon}, \quad \vec{\varepsilon} \in L(0, \sigma_\varepsilon^2)$$

we can write the likelihood function as

$$l_1(\vec{u}, \vec{\theta}) = -\frac{2}{\sigma_\varepsilon^2} |\vec{u} - \psi \vec{\theta}|.$$

We see that the maximization of this likelihood function is equivalent to the method of least modulus.

6.6 Method of least modulus

We consider now the linear model

$$\vec{u} = \psi\vec{\theta} + \vec{\varepsilon}$$

of the observations, which contains outliers. The robust estimate is given by the minimization of the norm L_1 or by the minimization of the modulus of deviation

$$\lambda(\vec{\theta}) = \sum_{i=1}^n v_i |u_i - \sum_{s=1}^S \psi_{is} \theta_s|,$$

where v_i are weight coefficients. The estimate can be written as

$$\hat{\vec{\theta}} = \arg \min_{\vec{\theta}} \sum_{i=1}^n v_i |u_i - \sum_{s=1}^S \psi_{is} \theta_s|.$$

To find the estimate we introduce a function of vector $\vec{\theta}$ and $\vec{\rho}$

$$\lambda_1(\vec{\theta}, \vec{\rho}) = \sum_{i=1}^n \frac{v_i^2 |u_i - \sum_{s=1}^S \psi_{is} \theta_s|^2}{v_i |u_i - \sum_{s=1}^S \psi_{is} \rho_s|}$$

and we note that

$$\lambda_1(\vec{\theta}, \vec{\theta}) = \lambda(\vec{\theta}).$$

The function $\lambda_1(\vec{\theta}, \vec{\rho})$ makes possible to construct an iterative procedure.

Let be given an initial vector $\vec{\theta}^{(0)}$. We substitute the fixed value $\vec{\theta}^{(0)}$ into λ_1 in place of $\vec{\rho}$, then

$$\lambda_1(\vec{\theta}, \vec{\theta}^{(0)}) = \sum_{i=1}^n w_i^0 |u_i - \sum_{s=1}^S \psi_{is} \theta_s|^2,$$

where

$$w_i^0 = \frac{v_i}{|u_i - \sum_{s=1}^S \psi_{is} \theta_s^{(0)}|}$$

are the elements of diagonal matrix

$$W^{(0)} = \text{diag}(w_1^0, w_2^0, \dots, w_n^0).$$

We can use the weighted least squares method for finding of the first iteration of the estimate $\vec{\theta}$

$$\begin{aligned} \hat{\vec{\theta}} &= (\psi^T W^{(0)} \psi)^{-1} \psi^T W^{(0)} \vec{u}, \\ \hat{\vec{\theta}}^{(1)} &= \hat{\vec{\theta}}. \end{aligned}$$

We use the estimate $\hat{\vec{\theta}}^{(1)}$ to calculate the new weight coefficients $w_i^{(1)}$ in the function $\lambda_1(\vec{\theta}, \vec{\theta}^{(1)})$

$$\lambda_1(\vec{\theta}, \vec{\theta}^{(1)}) = \sum_{i=1}^n w_i^{(1)} (u_i - \sum_{s=1}^S \psi_{is} \theta_s)^2,$$

where

$$w_i^{(1)} = v_i / |u_i - \sum_{s=1}^S \psi_{is} \theta_s^{(1)}|.$$

The second iteration of the estimate is given by

$$\begin{aligned}\hat{\vec{\theta}} &= (\psi^T W^{(1)} \psi)^{-1} \psi^T W^{(1)} \vec{u}, \\ \vec{\theta}^{(2)} &= \hat{\vec{\theta}}.\end{aligned}$$

Then we substitute $\vec{\theta}^{(2)}$ instead of vector $\vec{\rho}$ to function $\lambda_1(\vec{\theta}, \vec{\rho})$ and find the third iteration $\vec{\theta}^{(3)}$.

The iteration procedure will be continued until the threshold will be satisfied

$$\frac{|\theta_s^{(k)} - \theta_s^{(k-1)}|}{|\theta_s^{(k)}|} < \delta \sim 10^{-2} \div 10^{-3}, \quad s = 1, \dots, S.$$

The basic element of the considered procedure is the weighted least squares method.

6.7 Huber robust method

The next robust method was suggested by Huber, this method is a combination of the least squares method for relatively small random values ε_i and the least modulus method for big values ε_i . We consider the linear model

$$u_i = \vec{\psi}_i^T \vec{\theta} + \varepsilon_i.$$

The estimate is determined by

$$\hat{\vec{\theta}} = \arg \min_{\vec{\theta}} \sum_{i=1}^n H(u_i - \vec{\psi}_i^T \vec{\theta})$$

where

$$H(\varepsilon) = \begin{cases} \varepsilon^2/2 & |\varepsilon| < c, \\ c|\varepsilon| - c^2/2 & |\varepsilon| \geq c, \end{cases}$$

(see Figure 6.1)

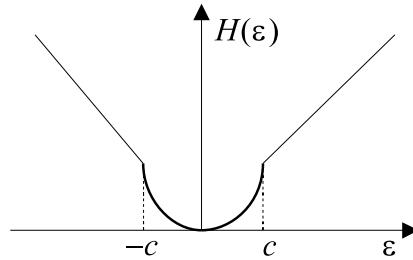


Figure 6.1: The Huber function for a divergence between measured data and the model values.

6.8 Andrews robust method

We consider now robust algorithm suggested by Andrews. The basis element of this algorithm is again the weighted least squares method. The estimate will be given by the minimization of the non-linear function

$$\hat{\vec{\theta}} = \arg \min_{\vec{\theta}} \sum_{i=1}^n H\left(\frac{u_i - \vec{\psi}_i^T \vec{\theta}}{\sigma}\right),$$

where

$$H(\varepsilon) = \begin{cases} 1 - \cos(\varepsilon/c) & |\varepsilon| < c\pi, \\ 0 & |\varepsilon| \geq c\pi, \end{cases}$$

$$\varepsilon_i = \frac{u_i - \vec{\psi}_i^T \vec{\theta}}{\sigma}$$

(see Figure 6.2). As an estimate for the standard σ we shall take the median of the

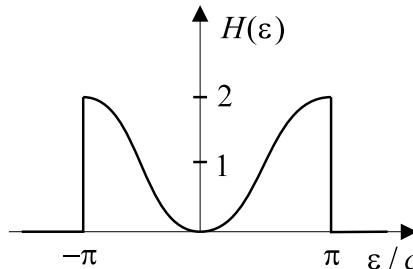


Figure 6.2: The Andrews function for a divergence between measured data and the model values.

modulus

$$\hat{\sigma} = \text{med} |u_i - \vec{\psi}_i^T \vec{\theta}|.$$

The problem is reduced to the solution of a set of equations

$$\sum_{i=1}^n h(\varepsilon_i) \psi_{is} = 0, \quad s = 1, \dots, S$$

where

$$h(\varepsilon_i) = H'(\varepsilon_i) = \begin{cases} \sin(\varepsilon_i/c)/c & |\varepsilon_i| < c\pi, \\ 0 & |\varepsilon_i| \geq c\pi. \end{cases}$$

For the solution of the non-linear set of equations the following iteration procedure is suggested:

1. We give an initial vector $\vec{\theta}_0$ or we can use the least square method estimate

$$\hat{\vec{\theta}}^{(0)} = (\psi^T \psi)^{-1} \psi^T \vec{u}.$$

We consider an algorithm for the estimation $\hat{\vec{\theta}}_0^{(k+1)}$ by the obtained estimate $\hat{\vec{\theta}}_0^{(k)}$.

2. We calculate the deviation

$$\hat{\varepsilon}_i^{(k)} = u_i - \vec{\psi}_i^T \hat{\vec{\theta}}^{(k)}.$$

3. We find an estimation for the standard σ

$$\hat{\sigma}^{(k)} = \text{med } \hat{\varepsilon}_i^{(k)}.$$

4. We calculate weight coefficients

$$w_i^{(k)} = \begin{cases} [\sin(\hat{\varepsilon}_i^{(k)}/c)/c][\hat{\varepsilon}_i^{(k)}]^{-1} & |\varepsilon_i|/c < \pi, \\ 0 & |\varepsilon_i|/c \geq \pi. \end{cases}$$

5. We solve a set of linear equations

$$\sum_{i=1}^n w_i^k \psi_{is} \hat{\varepsilon}_i^{(k+1)} = 0$$

or

$$\sum_{i=1}^n w_i^k \psi_{is} (u_i - \vec{\psi}_i^T \hat{\vec{\theta}}^{(k+1)}) = 0$$

and finally the $(k+1)$ -th iteration of the estimate will be given by

$$\hat{\vec{\theta}}^{(k+1)} = (\psi^T W^{(k)} \psi)^{-1} \psi^T W^{(k)} \vec{u}.$$

6. We check the threshold condition

$$\frac{|\hat{\theta}_s^{(k+1)} - \hat{\theta}_s^{(k)}|}{|\hat{\theta}_s^{(k)}|} \leq \delta_1 \sim 10^{-2} \div 10^{-3}.$$

If this condition is satisfied, then we stop the procedure. If it is not satisfied we shall pass to the next $(k+2)$ -th iteration.

6.9 Bayes method and statistical regularization

The Bayesian approach can be considered as a generalization of the maximum likelihood method, where a priori information about the parameters enters into the optimization scheme. The basic Bayes concept is, that the probability density function $p(\vec{\theta}/\vec{u})$ of the parameter $\vec{\theta}$ given the observation \vec{u} (the posterior density) is obtained from the density $p(\vec{\theta})$ of the parameters $\vec{\theta}$ (the prior density), the density $p(\vec{\theta}/\vec{u})$ of the observations \vec{u} given the parameters $\vec{\theta}$ (likelihood function) and the density $p(\vec{u})$ of the observations \vec{u} by

$$p(\vec{\theta}/\vec{u}) = \frac{p(\vec{\theta})p(\vec{u}/\vec{\theta})}{p(\vec{u})}.$$

Based on the posterior distribution we shall estimate the unknown parameters $\vec{\theta}$.

The Bayes point estimate is defined by the expected value for $\vec{\theta}$ computed with the posterior density

$$\hat{\vec{\theta}}_B = \int \vec{\theta} p(\vec{\theta}/\vec{u}) d\vec{\theta}.$$

In analogy to the standard statistical techniques the *generalized maximum likelihood estimate* $\hat{\vec{\theta}}_{MAP}$ of the parameter vector $\vec{\theta}$ is the *mode* of the posterior density $p(\vec{\theta}/\vec{u})$, i.e. the value for $\vec{\theta}$ which maximizes $p(\vec{\theta}/\vec{u})$, hence

$$\hat{\vec{\theta}}_{MAP} = \arg \max_{\vec{\theta}} p(\vec{\theta}/\vec{u})$$

or taking into account that the logarithm is the monotone function

$$\hat{\vec{\theta}}_{MAP} = \arg \max_{\vec{\theta}} \ln p(\vec{\theta}/\vec{u}).$$

If the maximum lies inside the admissible domain for parameters $\vec{\theta}$ and the function $\ln p(\vec{\theta}/\vec{u})$ has the first derivative. The maximum a posteriori estimate is found from the equation

$$\frac{\partial \ln p(\vec{\theta}/\vec{u})}{\partial \theta_s} = 0, \quad s = 1, \dots, S$$

or taking into account the Bayes formula for the $\ln p(\vec{\theta}/\vec{u})$

$$\ln p(\vec{\theta}/\vec{u}) = \ln p(\vec{\theta}) + \ln p(\vec{u}/\vec{\theta}) - \ln p(\vec{u})$$

we can write

$$\frac{\partial \ln p(\vec{\theta})}{\partial \theta_s} + \frac{\partial \ln p(\vec{u}/\vec{\theta})}{\partial \theta_s} = 0, \quad s = 1, 2, \dots, S.$$

The first term takes into account the prior information, the second one corresponds to the observations.

If the function $\ln p(\vec{\theta})$ changes weakly in the admissible domain of parameters $\vec{\theta}$, we can neglect by the first term. In general, if the prior density $p(\vec{\theta})$ is tended to

constant $p(\vec{\theta}) = \text{const}$, the maximum a posteriori estimate passes to the maximum likelihood estimate. If the prior density is not constant the maximum a posteriori estimate is stable. The using of the prior information lies on the basis of the *statistical regularization*.

Example: We find the maximum a posteriori estimate in the particular case of the linear model

$$\vec{u} = \psi\vec{\theta} + \vec{\varepsilon}$$

and normal distributed $\vec{\varepsilon}$ and $\vec{\theta}$

$$\vec{\varepsilon} \in N(0, R_\varepsilon), \quad \vec{\theta} \in N(\langle \vec{\theta} \rangle, R_\theta).$$

In this case we can write

$$\begin{aligned} \ln p(\vec{\theta}/\vec{u}) &= -\frac{1}{2}(\vec{u} - \psi\vec{\theta})^T R_\varepsilon^{-1}(\vec{u} - \psi\vec{\theta}) - \frac{1}{2}(\vec{\theta} - \langle \vec{\theta} \rangle)^T R_\theta^{-1} \times \\ &\quad \times (\vec{\theta} - \langle \vec{\theta} \rangle) - \frac{n+S}{2} \ln(2\pi) - \frac{1}{2}(\ln |R_\theta| + \ln |R_\varepsilon|) - \ln p(\vec{u}). \end{aligned}$$

The maximum a posteriori equation follows with

$$\frac{\partial \ln p(\vec{\theta}/\vec{u})}{\partial \theta_s} = 0, \quad s = 1, 2, \dots, S$$

and substituting the expression for $\ln p(\vec{\theta}/\vec{u})$ we have

$$(\psi^T R_\varepsilon^{-1} \psi + R_\theta^{-1})\vec{\theta} = \psi^T R_\varepsilon^{-1} \vec{u} + R_\theta^{-1} \langle \vec{\theta} \rangle$$

and the *maximum a posteriori estimate* written as

$$\tilde{\vec{\theta}}_{MAP} = (\psi^T R_\varepsilon^{-1} \psi + R_\theta^{-1})^{-1}(\psi^T R_\varepsilon^{-1} \vec{u} + R_\theta^{-1} \langle \vec{\theta} \rangle)$$

and the covariance matrix of the estimate is determined by

$$R_{\tilde{\vec{\theta}}_{MAP}} = (\psi^T R_\varepsilon^{-1} \psi + R_\theta^{-1})^{-1}.$$

We can rewrite the estimate in other form, by using two identities

$$\begin{aligned} (\psi^T R_\varepsilon^{-1} \psi + R_\theta^{-1})^{-1} &= R_\theta - R_\theta \psi^T (\psi R_\theta \psi^T + R_\varepsilon)^{-1} \psi R_\theta, \\ (\psi^T R_\varepsilon^{-1} \psi + R_\theta^{-1})^{-1} \psi^T R_\varepsilon^{-1} &= R_\theta \psi^T (\psi R_\theta \psi^T + R_\varepsilon)^{-1}, \end{aligned}$$

in the following form

$$\tilde{\vec{\theta}}_{MAP} = \langle \vec{\theta} \rangle + R_\theta \psi^T (\psi R_\theta \psi^T + R_\varepsilon^{-1})^{-1}(\vec{u} - \psi \langle \vec{\theta} \rangle)$$

and the covariance matrix

$$R_{\tilde{\vec{\theta}}_{MAP}} = R_\theta - R_\theta \psi^T (\psi R_\theta \psi^T + R_\varepsilon^{-1})^{-1} \psi R_\theta.$$

In the particular case of the uncorrelated random component $\vec{\varepsilon}$

$$R_\varepsilon = \sigma_\varepsilon^2 I_n,$$

and the random vector parameters $\vec{\theta}$

$$R_\theta = \sigma_\theta^2 I_S$$

we can write down the both of the considered estimates

$$\tilde{\vec{\theta}}_{MAP} = (\psi^T \psi + \alpha I_S)^{-1} (\psi^T \vec{u} + \alpha \langle \vec{\theta} \rangle)$$

and

$$\tilde{\vec{\theta}}_{MAP} = \langle \vec{\theta} \rangle + \psi^T (\psi \psi^T + \alpha I_n)^{-1} (\vec{u} - \psi \langle \vec{\theta} \rangle),$$

where $\alpha = \sigma_\varepsilon^2 / \sigma_\theta^2$.

The corresponding covariance matrices are written as

$$R_{\tilde{\vec{\theta}}_{MAP}} = \sigma_\varepsilon^2 (\psi^T \psi + \alpha I_S)^{-1}$$

and

$$R_{\tilde{\vec{\theta}}_{MAP}} = \sigma_\theta^2 (I_S - \psi^T (\psi \psi^T + \alpha I_n)^{-1} \psi).$$

We note that, we can interpret the parameter α as a parameter of the *Tikhonov regularization*. By the statistical regularization the parameter α has very clear physical sense — it is an energetic relation “noise/signal”. By tending α to zero, the maximum a posteriori estimate $\tilde{\vec{\theta}}_{MAP}$ passes to the maximum likelihood estimate $\hat{\vec{\theta}}_{MLS}$.

6.9.1 The recursion algorithm of the maximum a posteriori probability method

Using the presentation of the maximum a posteriori estimate in the form

$$\tilde{\vec{\theta}}_{MAP} = \langle \vec{\theta} \rangle + R_\theta \psi^T (\psi R_\theta \psi^T + R_\varepsilon)^{-1} (\vec{u} - \psi \langle \vec{\theta} \rangle)$$

and the corresponding covariance matrix

$$R_{\tilde{\vec{\theta}}_{MAP}} = R_\theta - R_\theta \psi^T (\psi R_\theta \psi^T + R_\varepsilon)^{-1} \psi R_\theta$$

we can construct the recursion procedure.

As an initial parameter vector we shall take the a priori expected vector $\langle \vec{\theta} \rangle$ and as an initial covariance matrix, we shall take the a priori covariance matrix R_θ , i.e.

$$\vec{\theta}^{(0)} = \langle \vec{\theta} \rangle \quad \text{and} \quad R_\theta^{(0)} = R_\theta.$$

We assume that the data are independent i.e.

$$R_\varepsilon = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2).$$

The first recursion for the estimate is computed by the formula

$$\vec{\theta}^{(1)} = \vec{\theta}^{(0)} + \gamma^{(1)} \Delta \vec{\theta}^{(1)},$$

where

$$\begin{aligned}\Delta \vec{\theta}^{(1)} &= R_\theta^{(0)} \vec{\psi}_1^T, \quad \gamma^{(1)} = N_1 / \delta_1^2, \\ N_1 &= u_1 - \vec{\psi}_1 \vec{\theta}^{(0)}, \quad \delta_1^2 = \sigma_{\varepsilon 1}^2 + \vec{\psi}_1 \Delta \vec{\theta}^{(1)}, \quad \vec{\psi}_1 = \|\psi_{11}, \psi_{12}, \dots, \psi_{1S}\|.\end{aligned}$$

The first recursion for the covariance matrix is given by

$$R_\theta^{(1)} = R_\theta^{(0)} - \frac{\Delta \vec{\theta}^{(1)} (\Delta \vec{\theta}^{(1)})^T}{\delta_1^2}.$$

Let be obtained the estimate and the covariance matrix for the $(i-1)$ -th recursion

$$\vec{\theta}^{(i-1)}, \quad R_\theta^{(i-1)},$$

then the i -th estimate and covariance matrix can be written as

$$\vec{\theta}^{(i)} = \vec{\theta}^{(i-1)} + \gamma^i \Delta \vec{\theta}^{(i)},$$

where

$$\begin{aligned}\Delta \vec{\theta}^{(i)} &= R_\theta^{(i-1)} \vec{\psi}_i^T, \quad \gamma^{(i)} = N_i / \delta_i^2, \\ N_i &= u_i - \vec{\psi}_i \vec{\theta}^{(i-1)}, \quad \delta_i^2 = \sigma_{\varepsilon i}^2 + \vec{\psi}_i \Delta \vec{\theta}^{(i)}\end{aligned}$$

and

$$R_\theta^{(i)} = R_\theta^{(i-1)} - \Delta \vec{\theta}^{(i)} (\Delta \vec{\theta}^{(i)})^T / \delta_i^2.$$

It should be noted that this procedure is stable and can be easily realized for arbitrary number of the data. This procedure can be used for the construction of dynamic or Kalman filter.

6.9.2 Kalman filter

Analysis of seismic signals shows that they change its shape by the propagation through the medium. We can take into account this changing using the following models

$$\begin{aligned}u_i &= \vec{\varphi}_i \vec{\theta}_i + \varepsilon_i, \quad \varepsilon_i \in N(0, \sigma_{\varepsilon i}^2), \\ \vec{\theta}_i &= F_{ii-1} \vec{\theta}_{i-1} + \vec{G} \delta_{i-1}, \quad \delta_i \in N(0, \sigma_{\delta i}^2),\end{aligned}$$

where vectors of amplitudes are

$$\vec{\theta}_i = \begin{bmatrix} \theta_i^{(1)} \\ \vdots \\ \theta_i^{(M)} \end{bmatrix}, \quad \vec{\theta}_{i-1} = \begin{bmatrix} \theta_{i-1}^{(1)} \\ \vdots \\ \theta_{i-1}^{(M)} \end{bmatrix}, \quad \vec{G} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The components of vectors θ_i and θ_{i-1} are connected by

$$\begin{bmatrix} \theta_i^{(2)} = \theta_{i-1}^{(1)} \\ \dots \\ \theta_i^{(M)} = \theta_{i-1}^{(M-1)} \end{bmatrix}, \quad \vec{\varphi}_i = [\varphi_i^{(1)}, \dots, \varphi_i^{(M)}]$$

is a vector of the signal shape (see Figure 6.3). The transition matrix from vector $\vec{\theta}_{i-1}$ to the vector $\vec{\theta}_i$ is given by

$$F_{ii-1} = \begin{bmatrix} r_{i-1}(1) & r_{i-1}(2) & \dots & r_{i-1}(M-1) & r_{i-1}(M) \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix},$$

where $r_{i-1}(\mu)$ are the elements of the a priori autocorrelation function.

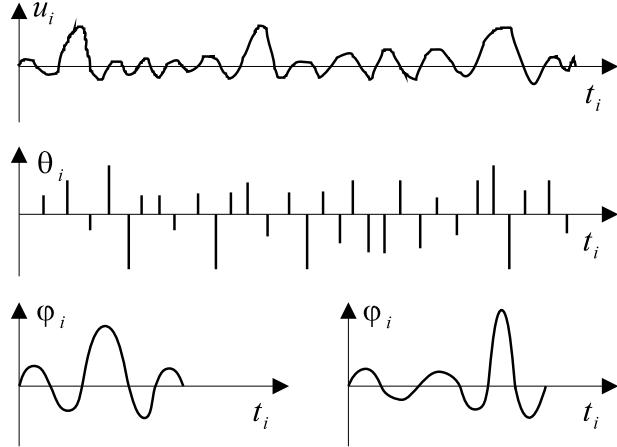


Figure 6.3: Illustration of the Kalman filtering procedure.

The estimate $\vec{\theta}_i$ is found by the maximization of the posterior density

$$\ln p(\vec{\theta}_i / \vec{u})$$

or, taking into account the normal distribution of the random components ε_i and δ_i , we shall obtain the function

$$l(\vec{\theta}_i) = \frac{-1}{2\sigma_{\varepsilon_i}^2} (u_i - \vec{\varphi}_i \vec{\theta}_i)^2 - \frac{1}{2} (\vec{\theta}_i - \hat{\vec{\theta}}_i')^T R_{\vec{\theta}_i'}^{-1} (\vec{\theta}_i - \hat{\vec{\theta}}_i'),$$

where $\hat{\vec{\theta}}_i' = F_{ii-1} \vec{\theta}_i$ and the covariance matrix

$$R_{\vec{\theta}_i'}^{-1} = \langle \vec{\theta}_i \vec{\theta}_i^T \rangle = F_{ii-1} R_{\theta_{i-1}} F_{ii-1}^T + \vec{G} \sigma_{\delta_{i-1}}^2 \vec{G}^T.$$

The estimate can be written as a result of the maximization of the function $l(\vec{\theta}_i)$

$$\tilde{\vec{\theta}}_i = \left(\frac{\vec{\varphi}_i^T \vec{\varphi}_i}{\sigma_{\varepsilon i}^2} + R_{\theta'_i}^{-1} \right)^{-1} \left(\frac{\vec{\varphi}_i^T u_i}{\sigma_{\varepsilon i}^2} + R_{\theta'_i}^{-1} \hat{\vec{\theta}}'_i \right)$$

or using identities, introduced above, we obtain the recursion formula for the estimate

$$\tilde{\vec{\theta}}_i = \hat{\vec{\theta}}'_i + R_{\theta'_i} \vec{\varphi}_i^T (\vec{\varphi}_i R_{\theta'_i} \vec{\varphi}_i^T + \sigma_{\varepsilon i}^2)^{-1} (u_i - \vec{\varphi}_i \hat{\vec{\theta}}'_i)$$

and the corresponding covariance matrix can be written as

$$R_{\theta_i} = R_{\theta'_i} - R_{\theta'_i} \vec{\varphi}_i^T (\vec{\varphi}_i R_{\theta'_i} \vec{\varphi}_i^T + \sigma_{\varepsilon i}^2)^{-1} \vec{\varphi}_i R_{\theta'_i}.$$

As an initial vector and covariance matrix we can take for the simplicity the following presentation

$$\vec{\theta}_0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad R_{\theta_0} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

6.10 Singular analysis and method of least squares

We consider again the linear model

$$\vec{u} = \psi \vec{\theta} + \vec{\varepsilon}, \quad \vec{\varepsilon} \in N(0, \sigma_{\varepsilon}^2).$$

By the practical application of the method of least squares, we can not obtain satisfactory results in the case of a bad condition of a matrix $\psi^T \psi$, which is subjected to inversion. As a measure of condition, it is usually used a *condition number*

$$\beta = G/g,$$

where

$$G = \max_{\theta} \frac{\|\psi^T \psi \vec{\theta}\|_1}{\|\vec{\theta}\|_1}, \quad g = \min_{\theta} \frac{\|\psi^T \psi \vec{\theta}\|_1}{\|\vec{\theta}\|_1}, \quad \|\vec{\theta}\|_1 = \sum_{s=1}^S |\theta_s|.$$

If the matrix $\psi^T \psi$ is singular then the value g will be equal to zero.

The properties of the condition number.

1. The condition number β is always more or equal to one

$$\beta \geq 1 \quad \text{as} \quad G \geq g.$$

2. If the matrix $\psi^T \psi$ is a diagonal matrix as in the case of orthogonal polynomials, then

$$\beta = \frac{\max |(\psi^T \psi)_{ii}|}{\min |(\psi^T \psi)_{ii}|}.$$

3. If an inverse matrix exists, then

$$\beta = \|\psi^T \psi\| \|(\psi^T \psi)^{-1}\|.$$

4. If the matrix $\psi^T \psi$ is a singular one, then

$$\beta = \infty.$$

The big values of condition number demonstrate the presence of almost linearly dependent rows of the matrix.

The arbitrary real matrix ψ can be represented by the factorization or singular decomposition

$$\psi = Q\Sigma P^T,$$

where $Q_{m \times n}$ and $P_{S \times S}$ are orthogonal matrices, i.e.

$$QQ^T = I_{n \times n}, \quad \text{and} \quad PP^T = I_{S \times S},$$

and the matrix $\Sigma_{n \times S}$ is a diagonal one with elements

$$\Sigma_{ij} = \begin{cases} \sigma_i & i = j, \quad i = 1, \dots, n, \\ 0 & i \neq j, \quad j = 1, \dots, S, \end{cases} \quad \Sigma = \underbrace{\begin{bmatrix} \sigma_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_S \\ 0 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 0 \end{bmatrix}}_S \Bigg\} n. \quad (6.1)$$

The matrix Q consists of the orthonormal eigenvectors of the matrix $\psi\psi^T$, and the matrix P consists of the orthogonal eigenvectors of the matrix $\psi^T\psi$. The diagonal elements of the matrix Σ are nonnegative values of square roots from the eigenvalues of the matrix $\psi^T\psi$, they are called *singular values*. We assume that

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_S \geq 0.$$

For simplicity, we consider uncorrelated random component ε , then we shall obtain the least squares estimate

$$\hat{\vec{\theta}} = \arg \min \lambda(\vec{\theta}),$$

where

$$\begin{aligned} \lambda(\vec{\theta}) &= (\vec{u} - \psi\vec{\theta})^T(\vec{u} - \psi\vec{\theta}) = (\vec{u} - Q\Sigma P^T\vec{\theta})^T QQ^T \times \\ &\times (\vec{u} - Q\Sigma P^T\vec{\theta}) = (\vec{y} - \Sigma\vec{b})^2, \end{aligned}$$

where $\vec{y} = Q^T\vec{u}$ and $\vec{b} = P^T\vec{\theta}$.

The minimization of the function λ gives the estimate

$$\hat{\vec{b}} = \Sigma^+ \vec{y}$$

and

$$\hat{\vec{\theta}} = P\Sigma^+Q^T\vec{u}.$$

The matrix Σ^+ is called pseudo-inverse matrix and can be written as

$$\Sigma^+ = \underbrace{\left[\begin{array}{cccccc} \sigma_1^{-1} & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \sigma_S^{-1} & 0 & \dots & 0 \end{array} \right]}_{n} \underbrace{\dots}_{n-S}.$$

The properties of the pseudo-inverse matrix.

1. $\Sigma\Sigma^+\Sigma = \Sigma$.
2. $\Sigma^+\Sigma\Sigma^+ = \Sigma^+$.
3. $\Sigma\Sigma^+$ is a symmetric matrix.
4. $\Sigma^+\Sigma$ is a symmetric matrix.

It is possible to show that the matrix Σ^+ always exists and it is unique. If the diagonal elements $\sigma_{ii} \neq 0$, then the estimate is easily found

$$b_i = \frac{y_i}{\sigma_{ii}}, \quad i = 1, 2, \dots, i_0.$$

But for small value σ_{ii} this procedure is undesirable. We must choice the threshold α starting from a priori accuracy of the data. Then we compare σ_{ii} with this value α .

If $\sigma_{ii} \geq \alpha$ then $b_i = y_i/\sigma_{ii}$ for $i = 1, 2, \dots, i_0$.

If $\sigma_{ii} < \alpha$ then $b_i = 0$ for $i = i_0 + 1, i_0 + 2, \dots, S$.

This procedure is called *singular analysis*.

Example: We consider the singular analysis estimate in the case of the random vector parameters $\vec{\theta}$ with

$$\langle \vec{\theta} \rangle = 0 \quad \text{and} \quad R_\theta = \sigma_\theta^2 I.$$

The random vector $\vec{\varepsilon}$ is uncorrelated

$$R_\varepsilon = \sigma_\varepsilon^2 I$$

and the random vector $\vec{\varepsilon}$ and $\vec{\theta}$ are uncorrelated.

Using the singular decomposition of the matrix ψ we can write the model as

$$\vec{y} = \Sigma \vec{b} + \vec{e},$$

where $\vec{e} = Q^T \vec{\varepsilon}$. The solution will be looked for as a linear form

$$\hat{\vec{b}} = \Gamma \vec{y}, \tag{6.2}$$

where the matrix Γ is a diagonal matrix $[S \times n]$

$$\Gamma = \underbrace{\begin{bmatrix} \gamma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \gamma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \gamma_S & 0 & \dots & 0 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} \gamma_1 \\ \gamma_1 \\ \vdots \\ \gamma_S \end{bmatrix}}_{n-S} - ?.$$

To determine the estimate $\hat{\vec{b}}$ it is sufficiently to determine the elements $\gamma_1, \dots, \gamma_S$.

We use the least squares criterion. The error of estimate and according covariance matrix have the form

$$\vec{\xi} = \hat{\vec{b}} - \vec{b} = (\Gamma\Sigma - I)\vec{b} + \Gamma\vec{e}$$

and

$$\begin{aligned} \langle \vec{\xi}\vec{\xi}^T \rangle &= (\Gamma\Sigma - I)\langle \vec{b}\vec{b}^T \rangle(\Gamma\Sigma - I)^T + \\ &+ \Gamma\langle \vec{e}\vec{e}^T \rangle\Gamma^T = (\Gamma\Sigma - I)R_b(\Gamma\Sigma - I)^T + \Gamma R_e \Gamma^T. \end{aligned}$$

The mean squared error is determined by

$$\sigma_\xi^2 = \text{Sp}(\langle \vec{\xi}\vec{\xi}^T \rangle) = \sum_{j=1}^S [(\gamma_j\sigma_j - 1)^2\sigma_{b_j}^2 + \gamma_j\sigma_{e_j}^2].$$

Taking into account that

$$R_b = P^T \langle \vec{\theta}\vec{\theta}^T \rangle P = \sigma_\theta^2 I$$

and

$$R_e = Q^T \langle \vec{\xi}\vec{\xi}^T \rangle Q = \sigma_\varepsilon^2 I,$$

we can write

$$\sigma_\xi^2 = \sum_{j=1}^S [(\gamma_j\sigma_j - 1)^2\sigma_\theta^2 + \gamma_j^2\sigma_\varepsilon^2].$$

Differentiating of the σ_ξ^2 with respect to γ_j

$$\begin{aligned} \frac{\partial \sigma_\xi^2}{\partial \gamma_j} &= 0, \quad j = 1, \dots, S, \\ 2\sigma_j(\gamma_j\sigma_j - 1)\sigma_\theta^2 + 2\gamma_j\sigma_\varepsilon^2 &= 0 \end{aligned}$$

we shall find the estimate $\hat{\gamma}_j$ with *statistical regularization*

$$\hat{\gamma}_j = \frac{\sigma_j}{(\sigma_j^2 + \sigma_\varepsilon^2/\sigma_\theta^2)}.$$

In the particular case $\sigma_\theta \rightarrow \infty$, we obtain the *least squares* estimate

$$\hat{\gamma}_j = \frac{1}{\sigma_j}.$$

In the case of singular analysis the mean square error is written in the form

$$\lambda(i_0) = \langle \|\xi\|^2 \rangle = \sigma_\theta^2(S - i_0) + \sigma_\varepsilon^2 \sum_{j=1}^{i_0} 1/\sigma_j^2.$$

The maximization of the $\lambda(i_0)$ with respect to i_0 gives

$$\sigma_{i_0} \geq \sigma_\varepsilon^2 / \sigma_\theta^2 = \alpha \quad \text{and} \quad \sigma_{i_0+1} < \sigma_\varepsilon^2 / \sigma_\theta^2 = \alpha,$$

$$\gamma_j = \begin{cases} 1/\sigma_j & j = 1, 2, \dots, i_0 \quad \sigma_{i_0} \geq \alpha, \\ 0 & j = i_0 + 1, \dots, S \quad \sigma_{i_0+1} < \alpha \end{cases}$$

(see Figure 6.4). We note, that the singular analysis is one of the regularization

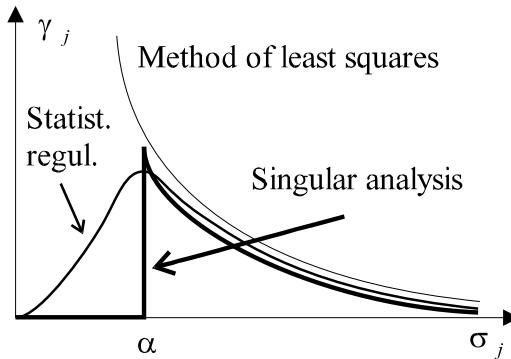


Figure 6.4: The dependence of the elements of the matrix Γ (6.2) on eigenvalues σ_j (6.1) for the cases of the least squares estimation, statistical regularization and singular analysis.

approaches.

6.10.1 Resolution matrix

We consider the model

$$\vec{u} = \psi \vec{\theta} + \vec{\varepsilon}.$$

The estimate can be written as

$$\hat{\vec{\theta}} = L \vec{u}.$$

We introduce the resolution matrix

$$R = L\psi \Rightarrow I_S.$$

In this case

$$\hat{\vec{\theta}} \xrightarrow{\varepsilon \rightarrow 0} R\vec{\theta}.$$

For the least squares estimate the operator L is written by:

$$L = (\psi^T \psi)^{-1} \psi^T, \quad R = L\psi = I_S.$$

We consider the singular decomposition

$$\psi = Q\Sigma P^T.$$

The resolution matrix can be written as

$$R = LQ\Sigma P^T = CP^T.$$

We shall find \vec{c}_j by minimizing the square form

$$\lambda(\vec{c}_j) = (P\vec{c}_j - \vec{\delta}_j)^T (P\vec{c}_j - \vec{\delta}_j),$$

where

$$\vec{\delta}_j = \|\delta_{ij}\|_{i,j=1}^S, \quad \delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j, \end{cases}$$

$$\vec{c}_j = (P^T P)^{-1} P^T \vec{\delta}_j = P^T \vec{\delta}_j,$$

$$C = P, \quad R = PP^T.$$

6.11 The method of Backus and Gilbert

Assume that unknown parameter function is described by a function $\theta(t)$, and that we consider a finite amount of discrete data u_1, u_2, \dots, u_n , which are linear functionals of θ through kernels $\varphi_i(t)$:

$$u_i = \int_0^T \varphi_i(t) \theta(t) dt + \varepsilon_i, \quad \varepsilon_i \in N(0, R_\varepsilon), \quad i = 1, 2, \dots, n,$$

where ε_i is a normal distributed random values. The problem consist of the recovery of the function $\theta(t)$.

The true (unknown) parameter function is denoted by $\theta(t_0)$:

$$\langle \theta(t_0) \rangle = \int_0^T C(t_0, t) \theta(t) dt,$$

where the function $C(t_0, t)$ should be close to the δ -function

$$C(t_0, t) = \delta(t - t_0) \Rightarrow \langle \theta(t_0) \rangle = \theta(t_0).$$

We assume that $C(t_0, t)$ is a normalized function

$$\int_0^T C(t_0, t) dt = 1.$$

Let $C(t_0, t)$ be a linear combination of the functions $\varphi_i(t)$

$$C(t_0, t) = \sum_{i=1}^n \varphi_i(t) \rho_i(t_0),$$

where $\rho_i(t_0)$ are unknown coefficients. Substituting this expression to the estimate $\langle \theta(t_0) \rangle$, we shall obtain

$$\langle \theta(t_0) \rangle = \sum_{i=1}^n \rho_i(t_0) f_i,$$

where

$$f_i = \int_0^T \varphi_i(t) \theta(t) dt.$$

We can write the model as

$$u_i = f_i + \varepsilon_i.$$

We shall seek the estimate as a linear combination of the observed data

$$\langle \hat{\theta}(t_0) \rangle = \sum_{i=1}^n \rho_i(t_0) u_i.$$

The problem now is to obtain the best constant $\rho_i(t_0)$. We introduce the *resolution length*

$$s(t_0) = 12 \int_{t_0-L/2}^{t_0+L/2} (t - t_0)^2 C^2(t_0, t) dt,$$

$$C(t_0, t) = \begin{cases} 1/L & \text{for } t \in [t_0 - L/2 \leq t \leq t_0 + L/2] \\ 0 & \text{for } t \notin [t_0 - L/2 \leq t \leq t_0 + L/2]. \end{cases}$$

The function $C(t_0, t)$ is presented at Figure 6.5. In this case we have

$$s(t_0) = (12/3L^2)((L/2)^3 - (-L/2)^3) = L.$$

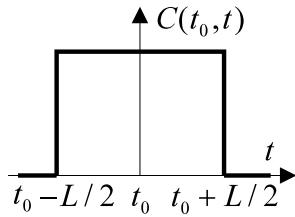
We can write $s(t_0)$ as a squared form

$$s(t_0) = \vec{\rho}^T(t_0) S(t_0) \vec{\rho}(t_0),$$

where the vector $\vec{\rho}(t_0)$ and matrix $S(t_0) = \|S_{\mu\nu}(t_0)\|_{\mu,\nu}^n$ can be written as

$$\vec{\rho}(t_0) = [\rho_1(t_0), \rho_2(t_0), \dots, \rho_n(t_0)],$$

$$S_{\mu\nu}(t_0) = 12 \int_{t_0-L/2}^{t_0+L/2} \varphi_\mu(t) \varphi_\nu(t) (t - t_0)^2 dt.$$

Figure 6.5: Graphic presentation of the $C(t_0, t)$ function.

The retrieval error is presented by

$$\begin{aligned}\sigma_\rho^2(t_0) &= \langle (\langle \theta(t_0) \rangle - \langle \hat{\theta}(t_0) \rangle)^2 \rangle = \langle \left(\sum_{i=1}^n \rho_i(t_0)(u_i - f_i) \right)^2 \rangle = \\ &= \vec{\rho}^T(t_0) R_\varepsilon \vec{\rho}(t_0).\end{aligned}$$

Finally, it is proposed to minimize the sum of weighted squared forms of $s(t_0)$ and $\sigma_\rho^2(t_0)$

$$g(\vec{\rho}(t_0)) = s(t_0) \cos \theta + \sigma_\rho^2(t_0) \beta \sin \theta = \vec{\rho}^T(t_0) G \vec{\rho}(t_0), \quad 0 \leq \theta \leq \pi/2,$$

where $G = S \cos \theta + \beta R_\varepsilon \sin \theta$. We shall find an estimate of $\vec{\rho}$

$$\hat{\vec{\rho}} = \arg \min g(\vec{\rho}(t_0))$$

under the condition

$$\vec{B}^T \vec{\rho}(t_0) = 1, \quad \vec{B} = \|B_\nu\|_{\nu=1}^n, \quad B_\nu = \int_0^T \varphi_\nu(t) dt.$$

This is a problem on a conditional extremum

$$\begin{aligned}\Phi(\vec{\rho}) &= \vec{\rho}^T G \vec{\rho} - 2\lambda(\vec{\rho}^T \vec{B} - 1), \\ \frac{\partial \Phi}{\partial \rho_s} &= 0, \quad s = 1, 2, \dots, S.\end{aligned}$$

As a result we get the system of linear equations

$$G \vec{\rho} - \lambda \vec{B} = 0$$

and can write the estimate of $\vec{\rho}$ in the next form

$$\hat{\vec{\rho}} = \lambda G^{-1} \vec{B}.$$

Multiplying the last relation on the vector \vec{B}^T we obtain

$$\vec{B}^T \hat{\vec{\rho}} = \lambda \vec{B}^T G^{-1} \vec{B}$$

and using the condition

$$\vec{B}^T \hat{\vec{\rho}} = 1,$$

we shall obtain

$$\lambda = 1/(\vec{B}^T G^{-1} \vec{B}).$$

Finally, the estimate of $\vec{\rho}$ is written down as

$$\hat{\vec{\rho}} = G^{-1} \vec{B} / (\vec{B}^T G^{-1} \vec{B}).$$

7

Solution of the inverse geophysical problems: case of the nonlinear model

7.1 Newton-Le Came method

Let us consider the model of observation data

$$\vec{u} = \vec{f}(\vec{\theta}) + \vec{\varepsilon}$$

$\varepsilon \in N(0, R_\varepsilon)$, $f(\theta)$ is the nonlinear function of the parameter vector $\vec{\theta}$. The maximum likelihood method leads in this case to the minimization of the quadratic form

$$l_1(\vec{u}, \vec{\theta}) = \frac{1}{2}(\vec{u} - \vec{f}(\vec{\theta}))^T R_\varepsilon^{-1} (\vec{u} - \vec{f}(\vec{\theta})).$$

We assume that an initial vector $\vec{\theta}^{(0)}$ is known. We expand the function $l_1(\vec{\theta})$ in the Taylor series in the vicinity of the value $\vec{\theta}^{(0)}$, being restricted to a squared $\vec{\theta}^{(0)}$

$$l_1(\vec{\theta}) \approx l_1(\vec{\theta}^{(0)}) + \Delta\vec{\theta}^T \vec{d} - (1/2)\Delta\vec{\theta}^T C\Delta\vec{\theta},$$

where

$$\Delta\vec{\theta} = \vec{\theta} - \vec{\theta}_0, \quad d_s = \frac{\partial l_1(\vec{\theta})}{\partial \theta_s} \Big|_{\vec{\theta}=\vec{\theta}^{(0)}}, \quad c_{ss'} = -\frac{\partial^2 l_1(\vec{\theta})}{\partial \theta_s \partial \theta_{s'}} \Big|_{\vec{\theta}=\vec{\theta}^{(0)}}.$$

We seek for the solution on the first iteration step from the equation

$$\frac{\partial l_1(\vec{\theta})}{\partial \theta_s} = 0 \Rightarrow C\Delta\vec{\theta} = \vec{d}, \quad s = 1, 2, \dots, S.$$

Le-Came suggested to take the mean value of matrix elements $c_{ss'}$

$$\tilde{c}_{ss'}^{(0)} = \langle c_{ss'}^{(0)} \rangle = \left\langle -\frac{\partial^2 l_1(\vec{\theta})}{\partial \theta_s \partial \theta_{s'}} \Big|_{\vec{\theta}=\vec{\theta}^{(0)}} \right\rangle.$$

We obtain the first iteration

$$\Delta \hat{\vec{\theta}}^{(1)} = [\tilde{C}^{(0)}]^{-1} \vec{d}^{(0)}, \quad \hat{\vec{\theta}}^{(1)} = \vec{\theta}^{(0)} + \Delta \hat{\vec{\theta}}^{(1)}.$$

The second iteration we find by the solution of the linear equation:

$$\Delta \hat{\vec{\theta}}^{(2)} = [\tilde{C}^{(1)}]^{-1} \vec{d}^{(1)}, \quad \hat{\vec{\theta}}^{(2)} = \hat{\vec{\theta}}^{(1)} + \Delta \hat{\vec{\theta}}^{(2)},$$

where

$$d_s^{(1)} = \frac{\partial l_1(\vec{\theta})}{\partial \theta_s} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(1)}}, \quad \tilde{c}_{ss'}^{(1)} = \left\langle -\frac{\partial^2 l_1(\vec{\theta})}{\partial \theta_s \partial \theta_{s'}} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(1)}} \right\rangle.$$

The n -th iteration will be determined by

$$\Delta \hat{\vec{\theta}}^{(n)} = [\tilde{C}^{(n-1)}]^{-1} \vec{d}^{(n-1)}, \quad \hat{\vec{\theta}}^{(n)} = \hat{\vec{\theta}}^{(n-1)} + \Delta \hat{\vec{\theta}}^{(n)},$$

where

$$d_s^{(n-1)} = \frac{\partial l_1(\vec{\theta})}{\partial \theta_s} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(n-1)}}, \quad \tilde{c}_{ss'}^{(n-1)} = \left\langle -\frac{\partial^2 l_1(\vec{\theta})}{\partial \theta_s \partial \theta_{s'}} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(n-1)}} \right\rangle.$$

The iteration procedure is finished if the threshold condition is satisfied

$$|\Delta \hat{\theta}_s^{(n)} / \hat{\theta}_s^{(n)}| \leq \beta \sim 10^{-2} \div 10^{-3}$$

for all $s = 1, 2, \dots, S$.

For the additive model

$$\vec{u} = \vec{f}(\vec{\theta}) + \vec{\varepsilon}$$

we shall obtain an explicit form of elements d_s and $\tilde{c}_{ss'}$:

$$\begin{aligned} d_s^{(n)} &= (\vec{u} - \vec{f}(\vec{\theta}))^T R_\varepsilon^{-1} \frac{\partial \vec{f}}{\partial \theta_s} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(n)}}, \\ \tilde{c}_{ss'}^{(n)} &= \left\langle -(\vec{u} - \vec{f}(\vec{\theta}))^T R_\varepsilon^{-1} \frac{\partial^2 \vec{f}}{\partial \theta_s \partial \theta_{s'}} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(n)}} + \frac{\partial \vec{f}^T}{\partial \theta_s} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(n)}} R_\varepsilon^{-1} \frac{\partial \vec{f}}{\partial \theta_{s'}} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(n)}} \right\rangle = \\ &= -\underbrace{\langle \vec{\varepsilon}^T \rangle}_{=0} R_\varepsilon^{-1} \frac{\partial^2 \vec{f}}{\partial \theta_s \partial \theta_{s'}} + \frac{\partial \vec{f}^T}{\partial \theta_s} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(n)}} R_\varepsilon^{-1} \frac{\partial \vec{f}}{\partial \theta_{s'}} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(n)}} = \\ &= \frac{\partial \vec{f}^T}{\partial \theta_s} R_\varepsilon^{-1} \frac{\partial \vec{f}}{\partial \theta_{s'}} \Big|_{\vec{\theta}=\hat{\vec{\theta}}^{(n)}}. \end{aligned}$$

As an examples we consider the estimation of the parameters of magnetic objects.

Example 1: The thread of magnetic poles (magnetization is vertical)

$$f_k(\vec{\theta}) = \frac{2Mh}{h^2 + (x_k - \xi)^2}, \quad \vec{\theta} = \{M, h, \xi\},$$

where M is the magnetic moment of the thread of poles, h is the depth of the thread of poles bedding, ξ is a coordinate of the function $f_k(\vec{\theta})$ maximum.

Example 2: The vertical magnetic dipole

$$f_k(\vec{\theta}) = \frac{M(2h^2 - x_k^2)}{(x_k^2 + h^2)^{5/2}}, \quad \vec{\theta} = \{M, h\},$$

where M is the magnetic moment of a dipole, h is the depth of bedding.

Example 3: The point magnetic pole

$$f_k(\vec{\theta}) = \frac{Mh}{(x_k^2 + h^2)^{3/2}}, \quad \vec{\theta} = \{M, h\},$$

where M is the magnetic moment of the point pole, h is depth of its bedding.

7.2 Method of least squares: nonlinear case

Let us consider a model of observation data

$$\vec{u} = \vec{\varphi}(\vec{\theta}) + \vec{\varepsilon}.$$

Taking into account the linearization of the function $\vec{\varphi}(\vec{\theta})$ in the vicinity of an initial parameter vector $\vec{\theta}^{(0)}$

$$\vec{\varphi}(\vec{\theta}) \approx \vec{\varphi}(\vec{\theta}^{(0)}) + \psi \Delta \vec{\theta},$$

where

$$\|\psi_{js}^{(0)}\| = \partial \varphi_j / \partial \theta_s |_{\vec{\theta}=\vec{\theta}_0} \quad \text{and} \quad \Delta \vec{\theta}^{(0)} = \vec{\theta} - \vec{\theta}^{(0)}.$$

We shall rewrite the model in the form

$$\tilde{\vec{u}}_0 = \psi_0 \Delta \vec{\theta} + \vec{\varepsilon},$$

where $\tilde{\vec{u}}_0 = \vec{u} - \vec{\varphi}(\vec{\theta}_0)$.

The first iteration can be written in the explicit form

$$\hat{\Delta \vec{\theta}}^{(1)} = (\psi_0^T W_0 \psi_0)^{-1} \psi_0^T W_0 \tilde{\vec{u}}_0,$$

where W_0 is the weight matrix, determined by

$$W_0 = (\hat{\sigma}_{\varepsilon_0}^2)^{-1} I = (\hat{\vec{\varepsilon}}_0^T \hat{\vec{\varepsilon}}_0 / (n - S))^{-1} I_{S \times S}$$

and $\hat{\vec{\varepsilon}}_0 = \vec{u} - \psi_0 \hat{\vec{\theta}}^{(0)}$. The estimation of the parameter vector for the first iteration is found as

$$\hat{\vec{\theta}}^{(1)} = \hat{\Delta \vec{\theta}}^{(1)} + \vec{\theta}^{(0)}.$$

The second iteration will be written as

$$\hat{\vec{\theta}}^{(2)} = \hat{\Delta \vec{\theta}}^{(2)} + \hat{\vec{\theta}}^{(1)},$$

where

$$\begin{aligned}\hat{\Delta\vec{\theta}}^{(2)} &= (\psi_1^T W_1 \psi_1)^{-1} \psi_1^T W_1 \tilde{u}_1, \\ \tilde{u}_1 &= \vec{u} - \varphi(\hat{\vec{\theta}}^{(1)}), \quad \|\psi_{js}^{(1)}\| = \left. \frac{\partial \varphi_j}{\partial \theta_s} \right|_{\vec{\theta}=\hat{\vec{\theta}}^{(1)}}, \\ W_1 &= (\hat{\sigma}_{\varepsilon 1}^2)^{-1} I = \left(\frac{\hat{\vec{\varepsilon}}_1^T \hat{\vec{\varepsilon}}_1}{n-S} \right)^{-1} I_{S \times S}, \\ \hat{\vec{\varepsilon}}_1 &= \vec{u} - \psi_1 \hat{\vec{\theta}}^{(1)}.\end{aligned}$$

The n -th iteration has the following presentation

$$\begin{aligned}\hat{\vec{\theta}}^{(n)} &= \hat{\Delta\vec{\theta}}^{(n)} + \hat{\vec{\theta}}^{(n-1)}, \\ \hat{\Delta\vec{\theta}}^{(n)} &= (\psi_{n-1}^T W_{n-1} \psi_{n-1})^{-1} \psi_{n-1}^T W_{n-1} \tilde{u}_{n-1}, \\ \tilde{u}_{n-1} &= \vec{u} - \varphi(\hat{\vec{\theta}}^{(n-1)}), \quad \|\psi_{js}^{(n-1)}\| = \left. \frac{\partial \varphi_j}{\partial \theta_s} \right|_{\vec{\theta}=\hat{\vec{\theta}}^{(n-1)}}, \\ W_{n-1} &= \left(\frac{\hat{\vec{\varepsilon}}_{n-1}^T \hat{\vec{\varepsilon}}_{n-1}}{n-S} \right)^{-1} I_{S \times S}, \\ \hat{\vec{\varepsilon}}_{n-1} &= \vec{u} - \psi_{n-1} \hat{\vec{\theta}}^{(n-1)}.\end{aligned}$$

The iteration procedure is finished if the threshold condition

$$|\Delta\hat{\theta}_s^{(n)}/\hat{\theta}_s^{(n)}| \leq \beta \sim 10^{-2} \div 10^{-3}$$

is satisfied for all $s = 1, 2, \dots, S$.

It should be noted that we use the method of least squares for linear model on each step of the iteration procedure.

7.3 Method of least squares for non-stationary model

Let us consider the model with a non-stationary noise

$$\vec{u} = \psi \vec{\theta} + \vec{\varepsilon}.$$

The variance σ_i^2 varies on time

$$\sigma_i^2 = \Sigma(f_i, \vec{\rho}) = \rho_0 + \rho_1 f_i^2$$

is proportional to the energy of the signal

$$f_i = \vec{\psi}_i^T \vec{\theta}.$$

Our problem consist to the estimation of the parameter vector $\vec{\theta}$ and parameter vector $\vec{\rho}(\rho_0, \rho_1)$ of the function $\Sigma(f_i, \vec{\rho})$. The iteration process is following.

- As an initial parameter vector $\vec{\theta}^{(0)}$, we can use an ordinary least squares method estimation

$$\hat{\vec{\theta}}^{(0)} = (\psi^T \psi)^{-1} \psi^T \vec{u}.$$

- The estimate of f_i is written as

$$\hat{f}_i = \vec{\psi}_i^T \hat{\vec{\theta}}^{(0)}.$$

- The squared deviation is

$$\hat{\varepsilon}_i^2 = (u_i - \hat{f}_i)^2.$$

We introduce the notation

$$\hat{B} = \begin{Bmatrix} 1 & \hat{f}_1^2 \\ 1 & \hat{f}_2^2 \\ \dots & \dots \\ 1 & \hat{f}_n^2 \end{Bmatrix}, \quad \hat{C} = \begin{Bmatrix} \hat{\varepsilon}_1^2 \\ \hat{\varepsilon}_2^2 \\ \dots \\ \hat{\varepsilon}_n^2 \end{Bmatrix}.$$

- The estimation of the parameter $\vec{\rho}$:

$$\hat{\vec{\rho}} = \arg \min (\hat{C} - \hat{B} \vec{\rho})^T (\hat{C} - \hat{B} \vec{\rho}), \quad \hat{\vec{\rho}} = (\hat{B}^T \hat{B})^{-1} \hat{B}^T \hat{C}.$$

- The estimation of the variance

$$\hat{\sigma}_i^2 = \hat{\rho}_0 + \hat{\rho}_1 \hat{f}_i^2.$$

- For the estimation of the parameter vector $\vec{\theta}$, taking into account the non-stationary $\hat{\sigma}_i^{-2}$ as the weighted function:

$$\tilde{\vec{\theta}} = (\psi^T W \psi)^{-1} \psi^T W \vec{u},$$

where

$$W = \text{diag}(\hat{\sigma}_1^{-2}, \hat{\sigma}_2^{-2}, \dots, \hat{\sigma}_n^{-2}).$$

- The estimation of the squared deviation:

$$\tilde{\varepsilon}_i^2 = (u_i - \tilde{f}_i)^2,$$

where

$$\tilde{f}_i = \vec{\psi}_i^T \tilde{\vec{\theta}}.$$

- The iteration procedure is finished, if the threshold condition is satisfied

$$\tilde{\varepsilon}_i^2 \leq \beta_0.$$

The procedure is continued if $\tilde{\varepsilon}_i^2 \geq \beta_0$, then we put $\hat{\vec{\theta}}^{(0)} = \tilde{\vec{\theta}}$ and go to the point 2.

7.4 Interval estimation method

Let the θ be an unknown parameter. The point estimate of this parameter $\hat{\theta}$ is a function of observation data

$$\hat{\theta} = \hat{\theta}(u_1, u_2, \dots, u_n).$$

We shall introduce a random interval I_β , which is defined as

$$P(|\hat{\theta} - \theta| < \delta) = \beta, \quad I_\beta = [\hat{\theta} - \delta, \hat{\theta} + \delta],$$

where β is a confidence probability and I_β is the confidence interval (see Figure 7.1).

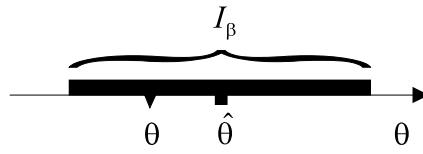


Figure 7.1: The confidence interval I_β and estimated parameter $\hat{\theta}$.

We assume that we have obtained for the model $\vec{u} = \psi\vec{\theta} + \varepsilon$ the MLS-estimate $\hat{\theta}_s$, which is unbiased $\langle \hat{\theta}_s \rangle = \theta_s$ and has the following covariance matrix

$$R_{\theta_s} = \sigma_\varepsilon^2 [(\psi^T W \psi)^{-1}]_{ss}, \quad s = 1, 2, \dots, S,$$

where W is the weight matrix.

After the normalization, we shall obtain the standard normal distribution:

$$\frac{\hat{\theta}_s - \theta_s}{\sigma_\varepsilon [(\psi^T W \psi)^{-1}]_{ss}^{1/2}} \in N(0, 1).$$

For the squared form of the deviation we obtain χ^2 -distribution with $n - S$ degree of freedom:

$$\hat{\varepsilon} = \vec{u} - \psi\hat{\vec{\theta}}, \quad \frac{\hat{\varepsilon}^T W \hat{\varepsilon}}{\sigma_\varepsilon^2} \in \chi_{n-S}^2.$$

It is well-known, that the ratio of the random value with the normal distribution $N(0, 1)$ to the random value with χ^2 -distribution, has the Student or t -distribution:

$$t_{n-S} = \frac{\hat{\theta}_s - \theta_s}{\sigma_\varepsilon [(\psi^T W \psi)^{-1}]^{1/2} (1/\sigma_\varepsilon) [\hat{\varepsilon}^T W \hat{\varepsilon} / (n - S)]^{1/2}} \in St(t_{n-S})$$

The confidence probability will be written as

$$P[|t_{n-S}| \leq \gamma] = \beta.$$

The mean square error $\hat{\sigma}_\varepsilon$ and parameter error $\hat{\sigma}_{\theta_s}$ are determined as

$$\hat{\sigma}_\varepsilon = \frac{\hat{\varepsilon}^T W \hat{\varepsilon}}{n - S}, \quad \hat{\sigma}_{\theta_s} = \hat{\sigma}_\varepsilon [(\psi^T W \psi)^{-1}]_{ss}.$$

The confidence interval for the parameter θ_s can be represented as

$$I_\beta = \left[\hat{\theta}_s \pm \gamma [(\psi^T W \psi)^{-1}]_{ss}^{1/2} \left[\frac{\hat{\varepsilon}^T W \hat{\varepsilon}}{n - S} \right]^{1/2} \right] = [\hat{\theta}_s \pm \gamma \hat{\sigma}_{\theta_s}].$$

For the estimate \hat{f} of the function

$$\hat{f} = \psi \hat{\theta}$$

We shall find the confidence domain, taking into account that

$$\frac{\hat{f}_i - f_i}{\sigma_\varepsilon [\psi (\psi^T W \psi)^{-1} \psi^T]_{ii}^{1/2}} \in N(0, 1)$$

and

$$t_{n-S} = \frac{\hat{f}_i - f_i}{[\psi (\psi^T W \psi)^{-1} \psi^T]_{ii}^{1/2} [\hat{\varepsilon}^T W \hat{\varepsilon} / (n - S)]^{1/2}} \in \text{St}(t_{n-S}).$$

The confidence probability will be written as

$$P(|t_{n-S}| \leq \gamma) = \beta$$

and the confidence domain is represented by

$$I_\beta = \hat{f}_i \pm \gamma \hat{\sigma}_\varepsilon [\psi (\psi^T \psi)^{-1} \psi^T]_{ii}^{1/2}$$

(see Figure 7.2).

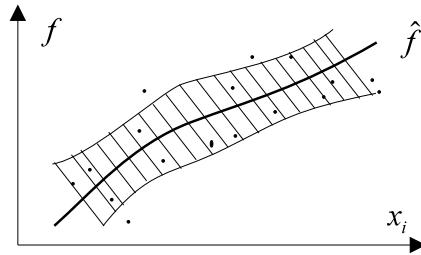


Figure 7.2: Graphic presentation of the confidence domain I_β .

7.5 Genetic algorithms

Genetic algorithms (GA's), first proposed by John Holland (1975), are based on analogies with the processes of biological evaluation. Genetic algorithms are very important methods for the solution of non-linear problems. The basic steps in a GA are the *coding*, *selection*, *crossover* and *mutation*.

7.5.1 Coding

Common to any basic GA is the digitization of the list of model parameters using a binary coding scheme.

Consider, for example, the coding of compressional wave velocity. The low velocity limit of interest may be 1500 m/s and the upper limit 1810 m/s. Assume that the desired resolution is 10 m/s. For this coding five bits is required (see Figure 7.3). The algorithm must now determine the fitness of the individual models. This means

<table border="1"><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	$v_{min} = 1500 \text{ m/s}$
0	0	0	0	0		
<table border="1"><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr></table>	0	0	0	0	1	$v = 1510 \text{ m/s}$
0	0	0	0	1		
<table border="1"><tr><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td></tr></table>	0	0	0	1	0	$v = 1520 \text{ m/s}$
0	0	0	1	0		
<table border="1"><tr><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	0	0	0	1	1	$v = 1530 \text{ m/s}$
0	0	0	1	1		
.....					
<table border="1"><tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr></table>	1	1	1	1	1	$v_{max} = 1810 \text{ m/s}$
1	1	1	1	1		

Figure 7.3: The binary coding of the values of velocity.

that the binary information is decoded into the physical model parameters and the forward problem is solved. The resulting synthetic data is estimated, then compared with the actual observed data using the specific fitness criteria. Depending on the problem, the definition of fitness will vary. For example, we represent the normalized correlation function, recently used in the geophysical inversion,

$$F(\vec{\theta}) = \frac{u_0 \otimes u_s(\vec{\theta})}{(u_0 \otimes u_0)^{1/2} (u_s \otimes u_s(\vec{\theta}))^{1/2}}$$

and mean squared error function

$$F(\vec{\theta}) = \sum_{i=1}^n (u_{i0} - u_{is}(\vec{\theta}))^2,$$

where \otimes is a sign of correlation, u_0 and $u_s(\vec{\theta})$ correspond to the observed data and model data for parameter $\vec{\theta}$.

7.5.2 Selection

Selection pairs of individual models for reproduction based on their fitness values. Models with higher fitness values are more likely to get selection than models with low fitness values. Let us consider the fitness proportionate selection.

The most basic selection method uses the ratio of each model's fitness function to the sum of fitness of all the models in the population to define its probability of

solution i.e.

$$p_s(\theta_i) = \frac{F(\theta_i)}{\sum_{j=1}^n F(\theta_j)},$$

where n is the number of models in the population. Selection based on these probabilities proceeds until a subset of the original models have been paired.

In a basic GA, if the population originally contained 100 models, fifty pairs are selected based on their fitness values. We shall take the models, which satisfied to the inequality

$$p_s(\theta_i) \geq \delta,$$

where δ is a given threshold value. Let we obtain L models, which form $L/2$ pairs. Each pair of the models will now produce two offsprings using the genetic operators of crossover and mutation. This will result in a completely new population of individuals.

7.5.3 Crossover

Crossover is the mechanism that allows genetic information between the paired models to be shared. In the terminology of geophysical inversion, crossover causes the exchange of some information between the paired models thereby generating new models. Crossover can performed in to modes: single and multi-point. In a single point crossover, one bit position in the binary string is selected at random from the uniform distribution. All of the bits to the right of this bit are now exchanged between the two models, generating two new models (see Figure 7.4). In multi-point crossover, this operation is carried out independently for each model parameter in the string (see Figure 7.5).

Example. The pair velocity model parameters v_1 and v_2 are the extreme members for the coding scheme of 1500 m/s and 1810 m/s (see Figure 7.6 and Table 7.1).

Crossover position	v_1 m/s	v_2 m/s
0	1500	1810
1	1510	1800
2	1530	1780
3	1570	1740
4	1650	1660
5	1810	1500

Table 7.1: An example of the crossover positions.

7.5.4 Mutation

The last genetic operator is mutation. Mutation is a random alternation of a bit. It can be carried out during the crossover process. The mutation rate is also specified by a probability determined by the algorithm designer.

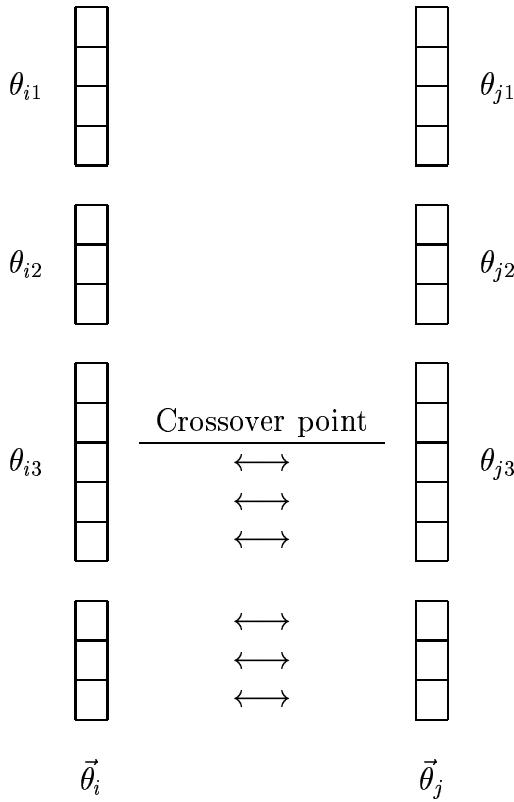


Figure 7.4: An example of the single-point crossover.

Example. The process of mutation for an initial velocity value of 1760 m/s (see Figure 7.7).

If random value ε , obtained from the random generator, less the threshold value P_m , then the mutation procedure is produced, if $\varepsilon > P_m$, then the mutation is not produced.

7.5.5 Choice

We choose from each of the $L/2$ pairs a model, which has the larger fitness function (see Figure 7.8). Thus we obtain $L/2$ models, which form randomly $L/4$ pairs. Then we produce the procedures: crossover, mutation and choice. This procedure is continued until we obtain the optimal model. The main advantage of GA is that makes possible to find the global maximum much faster than f.e. the Monte-Carlo method. We note that the local linearization methods can give only local maximum and they do not always lead to the global maximum.

Example 1: Test of GA on synthetic seismogram.

In the specific case of seismic signals the fitness function have the form

$$F(A_1, \dots, A_M, \tau_1, \dots, \tau_M) = \sum_{i=1}^n \left(u_i - \sum_{\mu=1}^M A_\mu \varphi(t_i - \tau_\mu) \right)^2,$$

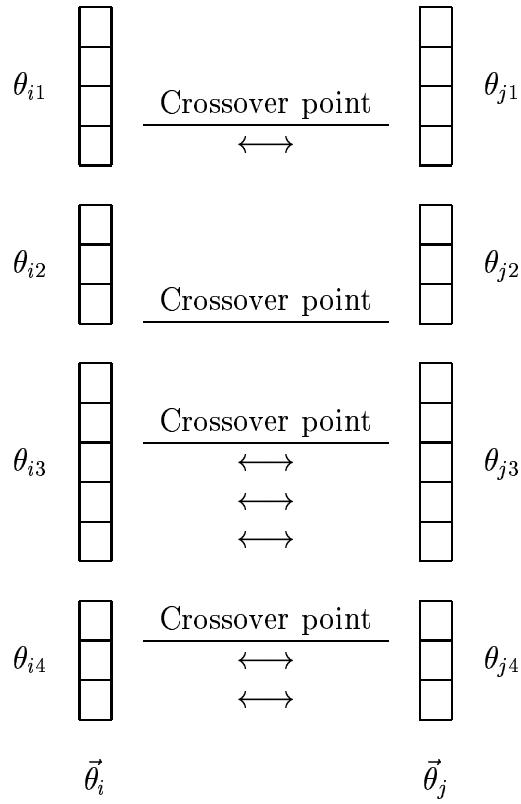


Figure 7.5: An example of the multi-point crossover

<table border="1"> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>5</td><td>4</td><td>3</td><td>2</td><td>1</td></tr> </table>	0	0	0	0	0	5	4	3	2	1	$v_1 = 1500 \text{ m/s}$
0	0	0	0	0							
5	4	3	2	1							
<table border="1"> <tr><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr> </table>	1	1	1	1	1	$v_2 = 1810 \text{ m/s}$					
1	1	1	1	1							

Figure 7.6: The pair of extreme positions for the coding scheme.

<table border="1"> <tr><td>5</td><td>4</td><td>3</td><td>2</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>0</td><td>1</td><td>0</td></tr> </table>	5	4	3	2	1	1	1	0	1	0	$v_1 = 1760 \text{ m/s}$
5	4	3	2	1							
1	1	0	1	0							
↓	$v_2 = 1740 \text{ m/s}$										

Figure 7.7: An illustration of mutation procedure.

where the shape of seismic wave is

$$\varphi = \exp\{-\beta|t|\} \cos \omega t$$

and A_μ is the amplitude, τ_μ is the arrival time of seismic wave with the number μ . Possible bounds for each of unknown parameters (A_μ, τ_μ) and discrete interval is determined a priori. The parameters of attenuation β and frequency ω are

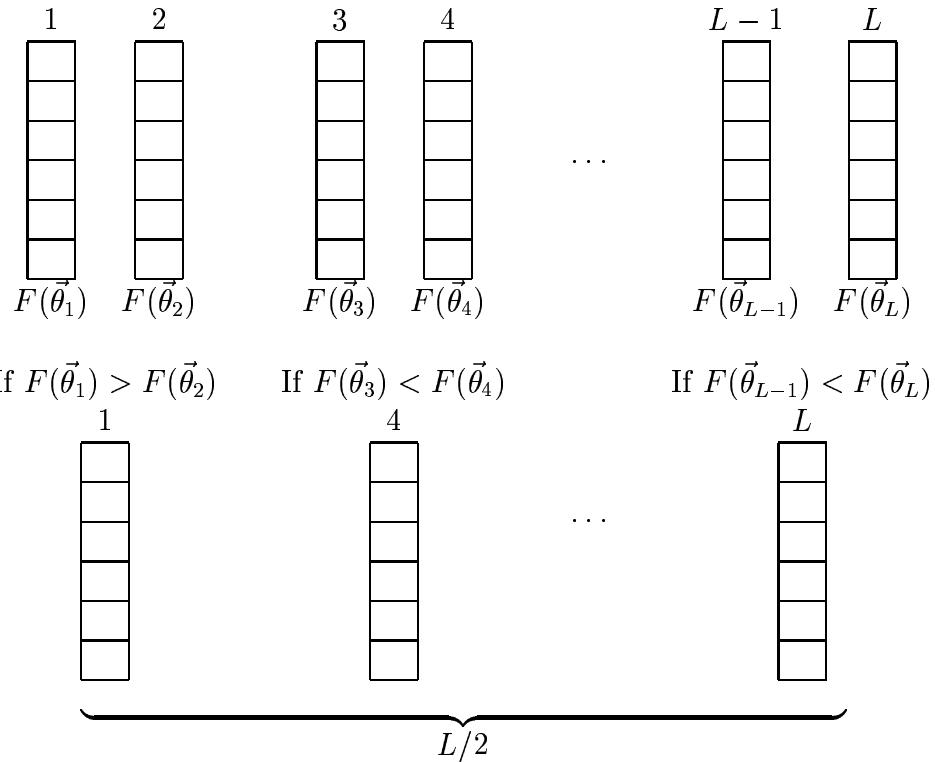


Figure 7.8: A reduction of the models during GA.

fixed also. The estimations of parameters to be find must minimize the fitness (or objective) function $F(A_1, \dots, A_M, \tau_1, \dots, \tau_M)$.

During applying of GA to optimization problem we choose the type of bit string encoding, the size of working population L and the values of probabilities of crossover and mutation.

Example 2: The size of population, probabilities of crossover and mutations are respectively:

$$L = 50, \quad P_c = 0.6, \quad P_m = 0.001.$$

Algorithm is tested under a variety population size, mutation and crossover rates.

In a case of two interfering seismic waves the threshold of correct waves separation is about $1/4$ wave period and signal/noise ratio must be equal or greater than two. For the case of three interfering seismic waves, if $|\tau_1 - \tau_2| = |\tau_2 - \tau_3| \geq 0.016s$. Three waves are separated with correct values of estimated parameters, which correspond to the value of waves displacement between $1/4$ and $1/3$ of wave period. Signal/noise ratio must be also greater or equal to two.

8

Statistical criteria for choice of models

The inverse geophysical problems are not uniquely defined, therefore we should use for the interpretation process a set of the medium models. We should find the inverse problem solution connected with our geological and geophysical a priori information for each model and choose by means of the test of statistical hypothesis the best model and the optimal reconstruction image.

8.1 Parametric hypothesis test

Let the vector parameter $\vec{\theta}_0$ be given. The hypothesis H_0 is a statement that a vector parameter $\vec{\theta}$ is equal to $\vec{\theta}_0$, and the hypothesis H_1 is a statement that a vector parameter $\vec{\theta}$ is not equal to $\vec{\theta}_0$:

$$\begin{aligned} H_0 &: \vec{\theta} = \vec{\theta}_0, \\ H_1 &: \vec{\theta} \neq \vec{\theta}_0. \end{aligned}$$

As a criterion it is used the likelihood ratio method

$$\lambda = -2 \ln \frac{L(\vec{u}, \vec{\theta}_0)}{L(\vec{u}, \hat{\vec{\theta}})} \in \chi_S^2.$$

The likelihood ratio λ belongs to the χ_S^2 -distribution with S freedom degree. The test can be written as

$$\begin{cases} \text{If } \lambda < \chi_{\alpha, S}^2 \text{ then hypothesis } H_0 \text{ does not contradict to data } \vec{u},. \\ \text{If } \lambda > \chi_{\alpha, S}^2 \text{ then hypothesis } H_0 \text{ is rejected.} \end{cases}$$

Example: we consider the model

$$\vec{u} = \begin{cases} \text{either } \vec{f}(\vec{\theta}_0) + \vec{\varepsilon}_0 & : H_0 \quad \vec{\varepsilon}_0 \in N(0, \sigma_0^2 I), \\ \text{or } \vec{f}(\vec{\theta}_1) + \vec{\varepsilon}_1 & : H_1 \quad \vec{\varepsilon}_1 \in N(0, \sigma_1^2 I). \end{cases}$$

The likelihood ratio will be written as

$$\lambda = -2 \ln \left[\frac{L(\vec{u}, \vec{\theta}_0, \sigma_0^2)}{L(\vec{u}, \hat{\vec{\theta}}_1, \sigma_1^2)} \right].$$

The estimates of mean squared errors are

$$\hat{\sigma}_0^2 = (1/n)(\vec{u} - \vec{f}(\vec{\theta}_0))^2, \quad \hat{\sigma}_1^2 = (1/n)(\vec{u} - \vec{f}(\vec{\theta}_1))^2.$$

Substituting these estimates to the λ -ratio, we shall obtain

$$\lambda = n \ln(\hat{\sigma}_0^2 / \hat{\sigma}_1^2).$$

8.2 Criterion of a posteriori probability ratio

We consider the model of a choice between two signals $\vec{f}(\vec{\theta}_0)$ and $\vec{f}(\vec{\theta}_1)$

$$\vec{u} = \begin{cases} \text{either } \vec{f}(\vec{\theta}_0) + \vec{\varepsilon} & : H_0, \\ \text{or } \vec{f}(\vec{\theta}_1) + \vec{\varepsilon} & : H_1. \end{cases}$$

The posterior ratio will be written in the form

$$\lambda = \ln \frac{P(1/\vec{u})}{P(0/\vec{u})} = \ln \frac{P(1)p(\vec{u}/1)}{P(0)p(\vec{u}/0)},$$

where $P(0)$ and $P(1)$ are a priori information about accordingly to hypothesis H_0 and H_1 . The criterion is represented as

$$\begin{cases} \text{If } \lambda < 0 \text{ then the hypothesis } H_0 \text{ does not contradict to data } \vec{u}, \\ \text{If } \lambda \geq 0 \text{ then the hypothesis } H_0 \text{ is rejected.} \end{cases}$$

We assume that $\vec{\varepsilon} \in N(0, R)$, then

$$\lambda = \frac{1}{2}(\vec{u} - \vec{f}(\vec{\theta}_0))^T R^{-1}(\vec{u} - \vec{f}(\vec{\theta}_0)) - \frac{1}{2}(\vec{u} - \vec{f}(\vec{\theta}_1))^T R^{-1}(\vec{u} - \vec{f}(\vec{\theta}_1)). \quad (8.1)$$

We shall find conditional mean values and variances:

if the hypothesis H_0 is valid and $\vec{u} = \vec{f}(\vec{\theta}_0) + \vec{\varepsilon}$, then $\langle \lambda_0 \rangle$ and $\sigma_{\lambda_0}^2$ will be determined;

if the hypothesis H_1 is valid and $\vec{u} = \vec{f}(\vec{\theta}_1) + \vec{\varepsilon}$, then $\langle \lambda_1 \rangle$ and $\sigma_{\lambda_1}^2$ will be determined.

To determine $\langle \lambda_0 \rangle$ we substitute in (8.1) instead of \vec{u} the model $\vec{f}(\vec{\theta}_0) + \vec{\varepsilon}$

$$\begin{aligned} \langle \lambda_0 \rangle &= \frac{1}{2} \langle \vec{\varepsilon}^T R^{-1} \vec{\varepsilon} \rangle - \frac{1}{2} (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T R^{-1} (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1)) - \\ &\quad - \frac{1}{2} \langle \vec{\varepsilon}^T R^{-1} \vec{\varepsilon} \rangle - \langle (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T R^{-1} \vec{\varepsilon} \rangle = \\ &= -\frac{1}{2} (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T R^{-1} (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1)). \end{aligned}$$

It is easily to check, substituting instead of \vec{u} the model $\vec{f}(\vec{\theta}_1) + \vec{\varepsilon}$, that

$$\langle \lambda_1 \rangle = -\langle \lambda_0 \rangle.$$

We determine now $\sigma_{\lambda_0}^2$, using instead of \vec{u} the model $\vec{f}(\vec{\theta}_0) + \vec{\varepsilon}$

$$\begin{aligned} \sigma_{\lambda_0}^2 &= \langle (\lambda_0 - \langle \lambda_0 \rangle)^2 \rangle = \langle (-\frac{1}{2}(\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T R^{-1} \times \\ &\quad \times (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))) - (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T R^{-1} \vec{\varepsilon} + \frac{1}{2}(\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T \times \\ &\quad \times R^{-1}(\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1)) \rangle = \\ &= (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T R^{-1} \langle \vec{\varepsilon} \vec{\varepsilon}^T \rangle R^{-1} (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1)) = \\ &= (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T R^{-1} (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1)). \end{aligned}$$

It is easily to prove that

$$\sigma_{\lambda_1}^2 = \sigma_{\lambda_0}^2.$$

Thus, we can write the conditional probabilities of the ratio λ

$$p(\lambda/0) = \frac{1}{\sqrt{2\pi}\sigma_{\lambda_0}} \exp\left\{-\frac{(\lambda - \lambda_0)^2}{2\sigma_{\lambda_0}^2}\right\},$$

if the hypothesis H_0 is valid and

$$p(\lambda/1) = \frac{1}{\sqrt{2\pi}\sigma_{\lambda_1}} \exp\left\{-\frac{(\lambda - \lambda_1)^2}{2\sigma_{\lambda_1}^2}\right\},$$

if the hypothesis H_1 is valid.

We can represent four special cases, analyzing our decisions (see Figure 8.1):

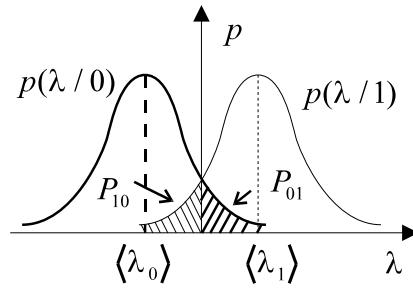


Figure 8.1: The density function for hypothesis H_0 and H_1 and errors for the decisions.

1. The hypothesis H_0 is true and it is not rejected.
2. The hypothesis H_0 is true, but it is rejected (error of the first kind). The probability P_{01} of this error is equal to the integral of $p(\lambda/0)$ into the interval

from 0 up to $+\infty$

$$\begin{aligned} P_{01} &= \int_0^\infty p(\lambda/0)d\lambda = \frac{1}{\sqrt{2\pi}\sigma_{\lambda_0}} \int_0^\infty \exp\left\{-\frac{(\lambda - \langle\lambda_0\rangle)^2}{2\sigma_{\lambda_0}^2}\right\} d\lambda = \\ &= \left[1 - \Phi\left(-\frac{\langle\lambda_0\rangle}{\sigma_{\lambda_0}}\right)\right], \end{aligned}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{t^2}{2}\right\} dt$$

is the Laplace function ($\Phi(-x) = 1 - \Phi(x)$).

- 3. The hypothesis H_0 is false and it is rejected.
- 4. The hypothesis H_0 is false, but it is not rejected (error of the second kind). The probability P_{10} of this error is equal to the integral of $p(\lambda/1)$ in the interval from 0 into $-\infty$.

$$\begin{aligned} P_{10} &= \int_{-\infty}^0 p(\lambda/1)d\lambda = \frac{1}{\sqrt{2\pi}\sigma_{\lambda_1}} \int_{-\infty}^0 \exp\left\{-\frac{(\lambda - \langle\lambda_1\rangle)^2}{2\sigma_{\lambda_1}^2}\right\} d\lambda = \\ &= \Phi\left(-\frac{\langle\lambda_1\rangle}{\sigma_{\lambda_1}}\right), \end{aligned}$$

Taking into account that $\sigma_{\lambda_0} = \sigma_{\lambda_1}$ and $\langle\lambda_0\rangle = -\langle\lambda_1\rangle$ we shall obtain for the total error

$$P_e = P(0)P_{01} + P(1)P_{10}.$$

In the case $P(0) = P(1) = 1/2$

$$\begin{aligned} P_e &= \frac{1}{2} \left[1 - \Phi\left(-\frac{\langle\lambda_0\rangle}{\sigma_{\lambda_0}}\right)\right] + \frac{1}{2} \Phi\left(\frac{\langle\lambda_0\rangle}{\sigma_{\lambda_0}}\right) = \Phi\left(\frac{\langle\lambda_0\rangle}{\sigma_{\lambda_0}}\right), \\ P_e &= \Phi(-\alpha/2) = 1 - \Phi(\alpha/2), \\ \alpha &= [(\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T R^{-1} (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))]^{1/2}. \end{aligned}$$

Graphic presentation of $\Phi(-\alpha/2)$ – see Figure 8.2.

For the total error 5 % we shall have the threshold condition

$$[(\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))^T R^{-1} (\vec{f}(\vec{\theta}_0) - \vec{f}(\vec{\theta}_1))]^{1/2} \geq \alpha_0 = 3.3.$$

Example: We consider the problem of a seismic wave detections. The model of a part of seismogram will be written as

$$u_{ki} = \begin{cases} \text{either } A\varphi(t_i - \tau - k\Delta x\gamma) + \varepsilon_{ki} & : H_0, \\ \text{or } \varepsilon_{ki} & : H_1, \end{cases}$$

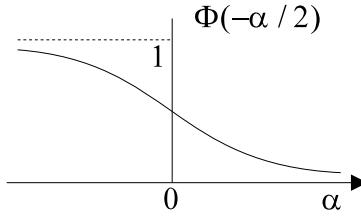
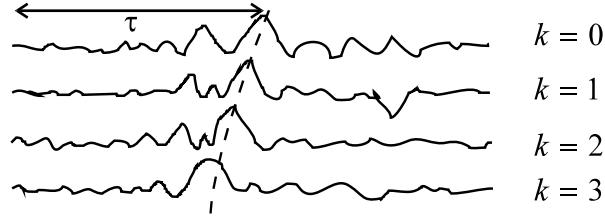
Figure 8.2: Graphic presentation of $\Phi(-\alpha/2)$.

Figure 8.3: An illustration for the seismic wave detection.

where A is an amplitude, τ is a travel time on the spur $k = 0$, γ is an apparent slowness, ε_{ki} is a random error, $\varphi(t)$ is a shape of seismic signal, $\varepsilon_{ki} \in N(0, R)$ (see Figure 8.3).

If the total error will be equal to 5 %, then the threshold condition is in our case

$$\alpha^2 = \frac{A^2}{\sigma^2} \sum_k \sum_i \sum_{i'} r_{ii'}^{-1} \varphi(t_i - \tau - k\Delta x \gamma) \varphi(t_{i'} - \tau - k\Delta x \gamma) \geq 3.3^2.$$

For uncorrelated random error $R = \sigma^2 I$

$$\alpha^2 = \frac{A^2}{\sigma^2} \sum_{i=1}^n \sum_{k=1}^K \varphi^2(t_i - \tau - k\Delta x \gamma) \geq 3.3^2.$$

If this condition is satisfied, we say that the seismic wave with parameters (A, τ, γ) exist under the 5 % total probability error.

8.3 The signal resolution problem

We consider the model for signal resolution

$$u = \begin{cases} \text{either } \vec{f}(\vec{\theta}_1) + \vec{f}(\vec{\theta}_2) + \vec{\varepsilon} & : H_0, \\ \text{or } \vec{f}(\vec{\theta}_0) + \vec{\varepsilon} & : H_1. \end{cases}$$

The hypothesis H_0 : we have two signals, the alternative hypothesis H_1 we have only one signal. We assume that $\varepsilon \in N(0, R)$, then we shall obtain the threshold condition.

If

$$\alpha^2 = (\vec{f}(\vec{\theta}_1) + \vec{f}(\vec{\theta}_2) - \vec{f}(\vec{\theta}_0))^T R^{-1} (\vec{f}(\vec{\theta}_1) + \vec{f}(\vec{\theta}_2) - \vec{f}(\vec{\theta}_0)) \geq \alpha_{0P_e}^2,$$

we can say that with the total error probability P_e , we have two signals.

Example 1: Let us consider the resolution of seismic signals. The model will be written in the form

$$u_{ki} = \begin{cases} \text{either } A_1\varphi(t_i - \tau_1 - k\Delta x \gamma_1) + A_2\varphi(t_i - \tau_2 - k\Delta x \gamma_2) + \varepsilon_{ki} & : H_0, \\ \text{or } A_0\varphi(t_i - \tau_0 - k\Delta x \gamma_0) + \varepsilon_{ki} & : H_1. \end{cases}$$

The α^2 -function has the following presentation

$$\begin{aligned} \alpha^2 = & \frac{1}{\sigma^2} \sum_k \sum_i \sum_{i'} r_{ii'}^{-1} (A_1\varphi(t_i - \tau_1 - k\Delta x \gamma_1) + \\ & + A_2\varphi(t_i - \tau_2 - k\Delta x \gamma_2) - A_0\varphi(t_i - \tau_0 - k\Delta x \gamma_0)) \\ & (A_1\varphi(t_{i'} - \tau_1 - k\Delta x \gamma_1) + A_2\varphi(t_{i'} - \tau_2 - k\Delta x \gamma_2) - A_0\varphi(t_{i'} - \tau_0 - k\Delta x \gamma_0)). \end{aligned}$$

We introduce for simplification of relations

$$\tau_1 = \tau_0 - \Delta\tau, \quad \gamma_1 = \gamma_0 - \Delta\gamma, \quad \tau_2 = \tau_0 + \Delta\tau, \quad \gamma_2 = \gamma_0 + \Delta\gamma, \quad A_0 = A_1/2 = A_2/2$$

and $R = \sigma^2 I$ (uncorrelated noise). Using these relations we shall obtain

$$\begin{aligned} \alpha^2 = & \frac{A_0^2}{\sigma^2} \sum_k \sum_i [\frac{1}{2}\varphi(t_i - (\tau_0 - \Delta\tau) - k\Delta x(\gamma_0 - \Delta\gamma)) + \\ & + \frac{1}{2}\varphi(t_i - (\tau_0 + \Delta\tau) - k\Delta x(\gamma_0 + \Delta\gamma)) - \varphi(t_i - \tau_0 - k\Delta x\gamma_0)]^2. \end{aligned}$$

In the simple case, if $\Delta\gamma = 0$, we can represent α^2 -function of $\Delta\theta$ (see Figure 8.4).

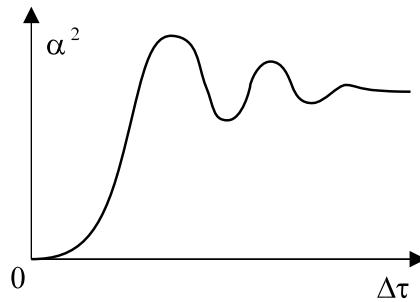


Figure 8.4: Graphic presentation of the α^2 -function.

Example 2: The resolution of magnetic fields from two vertical thin layer with the infinite embedded lower edge.

In this case the signals have the following presentation

$$\begin{aligned} f_{1k} &= \frac{I}{h} \left[\frac{1}{1 + (k\Delta x/h - \xi/h)^2} \right], & f_{2k} &= \frac{I}{h} \left[\frac{1}{1 + (k\Delta x/h + \xi/h)^2} \right], \\ f_{0k} &= \frac{2I}{h} \left[\frac{1}{1 + (k\Delta x/h)^2} \right], \end{aligned} \quad (8.2)$$

where I is the magnetization, ξ is the horizontal coordinate of the layer, h is the depth of the upper edge. The sum of the functions f_{0k} , f_{1k} and f_{2k} ($f = f_{0k} + f_{1k} + f_{2k}$) is presented at Figure 8.5. We can write α^2 -function and

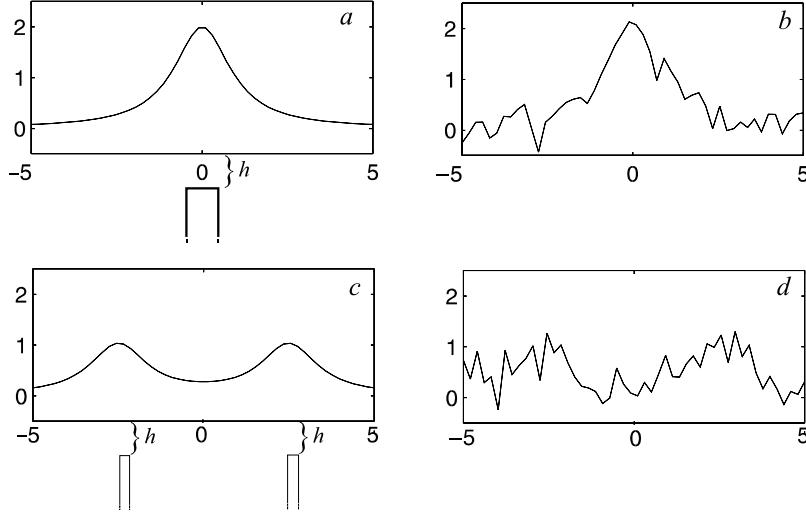


Figure 8.5: The magnetic fields produced by the vertical thin layers, defined in (8.2) ($I = 1$, $\xi = 2.5$, $h = 1.0$). a, c – without noise; b, d – with incorrelated Gaussian noise $N(0, 0.1)$; a, b – one layer; c, d – two layers.

the threshold condition

$$\left(\frac{2I}{\sigma h}\right)^2 \sum_{k=-K}^K \left[\left(2\left(1 + \left(\frac{k\Delta x}{h} - \frac{\xi}{h}\right)^2\right)\right)^{-1} + \left(2\left(1 + \left(\frac{k\Delta x}{h} + \frac{\xi}{h}\right)^2\right)\right)^{-1} - \left(1 + \left(\frac{k\Delta x}{h}\right)^2\right)^{-1} \right]^2 \geq \alpha_{0P_e}^2.$$

Under enough large h , if the inequality

$$\left(\frac{k\Delta x}{h} + \frac{\xi}{h}\right)^2 < 0.5$$

is satisfied, we can write the threshold condition approximately as

$$\left(\frac{\xi}{h}\right)^2 \geq \frac{\sigma h}{2I} \frac{\alpha_0}{\sqrt{2k+1}}.$$

The general threshold condition and its approximate analog determine the minimal values of ξ , which make possible still to distinguish the interference field of pair of layer from the field of single layer.

Example 3: We shall illustrate the application of the maximum likelihood method and the criterion of a posteriori probability ratio to the estimation of signal parameter of interfering fields.

The model of measurement data can be written as

$$\vec{u}_k = \sum_{\mu=1}^M \vec{f}_{\mu k}(\vec{\theta}_\mu) + \vec{\varepsilon}_k, \quad \vec{\varepsilon}_k \in N(0, R).$$

For the seismic case vector $\vec{f}_{\mu k}$ takes the form

$$\vec{f}_{\mu k} = \|A_\mu \varphi(t_i - \tau_\mu - k\Delta x \gamma_\mu)\|_{i=1}^n, \quad \vec{\theta}_\mu = \{A_\mu, \tau_\mu, \gamma_\mu\}, \quad k = 0, 1, \dots, K,$$

where A_μ is an amplitude, τ_μ is an arrival time on the trace $k = 0$, γ is an inverse apparent velocity of a seismic signal with number μ . An example of the interference phenomenon is presented at Figure 8.6

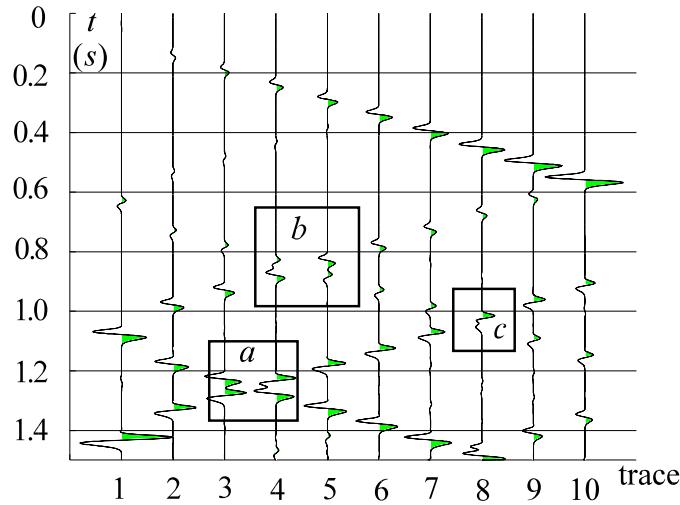


Figure 8.6: An illustration of the interference phenomenon of seismic waves. The model is the same as for Figure 5.1. The receiver points are: 0.003, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5. a, b, c are the regions with the interference phenomenon of seismic waves. (An amplitude amplifier is used.)

Let us consider the magnetic case. The interference magnetic field from unknown number of vertical thin layers with the infinite embedded lower edge reads as

$$\vec{u}_k = \sum_{\mu=1}^M \vec{f}_k(\vec{\theta}_\mu) + \vec{\varepsilon}_k, \quad \vec{\varepsilon}_k \in N(0, R),$$

$$f_k(\theta_\mu) = \frac{I_\mu h_\mu}{h_\mu^2 + (k\Delta x - \xi_\mu)^2}, \quad \vec{\theta}_\mu = \{I_\mu, h_\mu, \xi_\mu\},$$

where I_μ is the magnetization of the layer, ξ_μ is the horizontal coordinate of layer, h_μ is the depth of bedding of the upper edge of the layer with number μ .

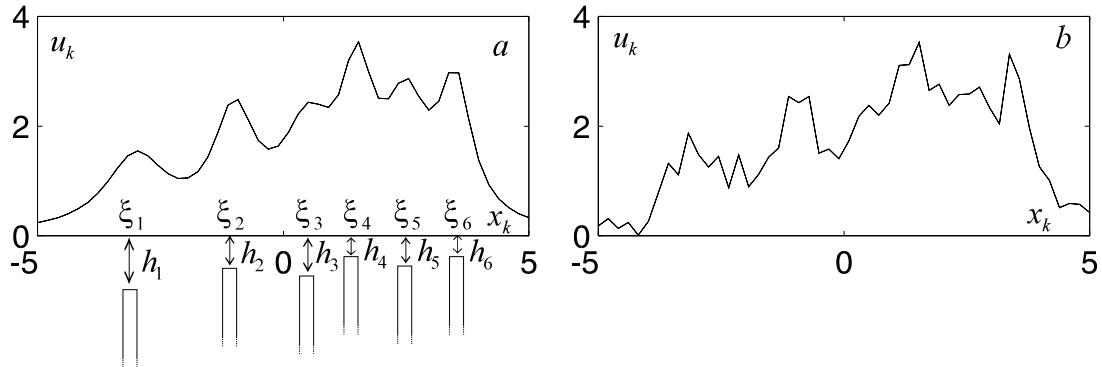


Figure 8.7: The magnetic field produced by the six thin layers ($I_\mu = 1$). The pairs (h_μ, ξ_μ) ($\mu = 1, \dots, 6$) are the next: $(0.75, -3.0)$, $(0.5, -1.0)$, $(0.6, 0.5)$, $(0.5, 1.5)$, $(0.5, 2.5)$, $(0.4, 3.5)$. a – without noise; b – with Gaussian uncorrelated noise $N(0, 0.2)$.

An example of the magnetic field produced by the six vertical layers is presented at Figure 8.7.

Our problem consist to estimate the parameter vector $\vec{\theta}_\mu$, ($\mu = 1, 2, \dots, M$) and the number of signals M

$$\vec{\theta}_\mu = ? \quad M = ?$$

Under our assumption a decision function from the maximum likelihood method can be written as

$$l(\vec{u}_k, \vec{\theta}) = -\frac{1}{2} \sum_k (\vec{u}_k - \sum_{\mu=1}^M \vec{f}_{\mu k}(\vec{\theta}_\mu))^T R^{-1} (\vec{u}_k - \sum_{\mu=1}^M \vec{f}_{\mu k}(\vec{\theta}_\mu)).$$

The parameter estimation will be found as

$$\hat{\vec{\theta}} = \arg \max_{\vec{\theta}} l(\vec{u}_k, \vec{\theta}), \quad \vec{\theta} = \|\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_M\|.$$

We leave in the function $l(\vec{u}_k, \vec{\theta})$ only terms, which are dependent on the vector parameter $\vec{\theta}$

$$\begin{aligned} g(\vec{u}_k, \vec{\theta}) &= \sum_{\mu=1}^M \sum_k [\vec{u}_k^T R^{-1} \vec{f}_{\mu k}(\vec{\theta}_\mu) - \frac{1}{2} \sum_{\mu' \neq \mu} \vec{f}_{\mu' k}^T(\vec{\theta}_{\mu'}) R^{-1} \vec{f}_{\mu k}(\vec{\theta}_\mu) - \\ &- \frac{1}{2} \vec{f}_{\mu k}^T(\vec{\theta}_\mu) R^{-1} \vec{f}_{\mu k}(\vec{\theta}_\mu)]. \end{aligned}$$

For the non-interfering signals $f_{\mu'}$ and f_μ the second term is vanished and the function $g(\vec{u}_k, \vec{\theta})$ will be written in the form

$$g_0(\vec{u}_k, \vec{\theta}) = \sum_{\mu=1}^M \sum_k [\vec{u}_k^T R^{-1} \vec{f}_{\mu k}(\vec{\theta}_\mu) - \frac{1}{2} \vec{f}_{\mu k}^T(\vec{\theta}_\mu) R^{-1} \vec{f}_{\mu k}(\vec{\theta}_\mu)].$$

We shall present the general function $g(\vec{u}_k, \vec{\theta})$ in the form by analogy with the function $g_0(\vec{u}_k, \vec{\theta})$. We shall distinguish in the function $g(\vec{u}_k, \vec{\theta})$ the term with the number m

$$\begin{aligned} g(\vec{u}_k, \vec{\theta}) = & [\sum_k \vec{u}_k^T R^{-1} \vec{f}_{mk}(\vec{\theta}_m) - \sum_{\substack{\mu=1 \\ \mu \neq m}}^M \sum_k \vec{f}_{\mu k}(\vec{\theta}_\mu) R^{-1} \vec{f}_{mk}(\vec{\theta}_m) - \\ & - \frac{1}{2} \sum_k \vec{f}_{mk}^T(\vec{\theta}_m) R^{-1} \vec{f}_{mk}(\vec{\theta}_m)] + [\sum_{\substack{\mu=1 \\ \mu \neq m}}^M \sum_k \vec{u}_k^T R^{-1} \vec{f}_{\mu k}(\vec{\theta}_\mu) - \\ & - \frac{1}{2} \sum_{\substack{\mu=1 \\ \mu \neq m}}^M \sum_{\substack{\mu'=1 \\ \mu' \neq m}}^M \sum_k \vec{f}_{\mu' k}^T(\vec{\theta}_{\mu'}) R^{-1} \vec{f}_{\mu k}(\vec{\theta}_\mu) - \frac{1}{2} \sum_{\substack{\mu=1 \\ \mu \neq m}}^M \sum_k \vec{f}_{\mu k}^T(\vec{\theta}_\mu) R^{-1} \vec{f}_{\mu k}(\vec{\theta}_\mu)]. \end{aligned}$$

Only the first bracket term depends on the parameter vector $\vec{\theta}_m$, therefore for the estimation of the vector $\vec{\theta}_m$ it is enough to minimize the function

$$\begin{aligned} g_m(\vec{u}_k, \vec{\theta}_m) &= \sum_k \vec{Y}_{km}^T R^{-1} \vec{f}_{mk}(\vec{\theta}_m) - \frac{1}{2} \sum_k \vec{f}_{mk}^T(\vec{\theta}_m) R^{-1} \vec{f}_{mk}(\vec{\theta}_m), \\ \vec{Y}_{km}^T &= \vec{u}_k^T - \sum_{\substack{\mu=1 \\ \mu \neq m}}^M \vec{f}_{k\mu}^T(\vec{\theta}_\mu) \end{aligned}$$

This presentation is a basis for the construction of the iteration procedure.

Step 1. We suggest that we have only one signal:

$$\begin{aligned} M = 1, \quad g_1(\vec{u}_k, \vec{\theta}_1) &= \sum_k [\vec{u}_k^T R^{-1} \vec{f}_{1k}(\vec{\theta}_1) - \frac{1}{2} \vec{f}_{1k}^T(\vec{\theta}_1) R^{-1} \vec{f}_{1k}(\vec{\theta}_1)] \\ \hat{\vec{\theta}}_1^{(1)} &= \max_{\vec{\theta}_1} g_1(\vec{u}_k, \vec{\theta}_1). \end{aligned}$$

We check the threshold condition:

If

$$\alpha_1^2 = \sum_k \vec{f}_{1k}^T(\hat{\vec{\theta}}_1^{(1)}) R^{-1} \vec{f}_{1k}(\hat{\vec{\theta}}_1^{(1)}) \geq \alpha_{0P_e}^2,$$

then the procedure will be continued.

If

$$\alpha_1^2 < \alpha_{0P_e}^2,$$

then the procedure will be stopped, and we say: we do not have any signal.

Step 2. We suggest that we have only two signals:

$$M = 2, \quad g_2(\vec{u}_k, \hat{\vec{\theta}}_1, \vec{\theta}_2) = \sum_k \vec{Y}_{k2}^T R^{-1} \vec{f}_{2k}(\vec{\theta}_2) - \frac{1}{2} \sum_k \vec{f}_{2k}^T(\vec{\theta}_2) R^{-1} \vec{f}_{2k}(\vec{\theta}_2)$$

$$\vec{Y}_{k2}^T = \vec{u}_k^T - \vec{f}_{1k}(\hat{\vec{\theta}}_1).$$

For parameter vector $\vec{\theta}_1$, we use the estimate received on the first step $\hat{\vec{\theta}}_1$

$$\hat{\vec{\theta}}_2^{(1)} = \max_{\vec{\theta}_2} g_2(\vec{u}_k, \hat{\vec{\theta}}_1, \vec{\theta}_2).$$

We check the threshold condition:

If

$$\alpha_2^2 = \sum_k \vec{f}_{2k}^T(\hat{\vec{\theta}}_2^{(1)}) R^{-1} \vec{f}_{2k}(\hat{\vec{\theta}}_2^{(1)}) \geq \alpha_{0P_e}^2,$$

then the procedure will be continued.

If

$$\alpha_2^2 < \alpha_{0P_e}^2,$$

then the procedure will be stopped, and we say: we have only one signal $\hat{\vec{\theta}}_1$.

We estimate the next iteration for the parameter vector $\vec{\theta}_1$ of the first signal, fixing the parameter vector $\vec{\theta}_2$ of the second signal by $\hat{\vec{\theta}}_2^{(1)}$

$$g_2(\vec{u}_k, \vec{\theta}_1, \hat{\vec{\theta}}_2^{(1)}) = \sum_k \vec{Y}_{k1}^T R^{-1} \vec{f}_{1k}(\vec{\theta}_1) -$$

$$- \frac{1}{2} \sum_k \vec{f}_{1k}^T(\vec{\theta}_1) R^{-1} \vec{f}_{1k}(\vec{\theta}_1),$$

$$\hat{\vec{\theta}}_1^{(1)} = \max_{\vec{\theta}_1} g_2(\vec{u}_k, \vec{\theta}_1, \hat{\vec{\theta}}_2^{(1)}), \quad \vec{Y}_{k1} = \vec{u}_k^T - \vec{f}_{2k}(\hat{\vec{\theta}}_2).$$

Then we estimate again $\hat{\vec{\theta}}_2^{(2)}$ and $\hat{\vec{\theta}}_1^{(2)}$. Each time transmitting to next iteration, we check the condition

If

$$\frac{g_2(\hat{\vec{\theta}}_1^{(n)}, \hat{\vec{\theta}}_2^{(n)}) - g_2(\hat{\vec{\theta}}_1^{(n-1)}, \hat{\vec{\theta}}_2^{(n-1)})}{g_2(\hat{\vec{\theta}}_1^{(n-1)}, \hat{\vec{\theta}}_2^{(n-1)})} < \delta \sim 10^{-2} \div 10^{-3},$$

we shall go to the next step $M = 3$.

If

$$\frac{\Delta g_2^{(n)}}{\Delta g_2^{(n-1)}} > \delta,$$

we continue the iteration procedure for vectors $\hat{\vec{\theta}}_2$ and $\hat{\vec{\theta}}_1$.

Step 3. We suggest that we have three signals

$$\begin{aligned} M = 3 \quad g_3(\vec{u}_k, \hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2, \hat{\vec{\theta}}_3) &= \sum_k \vec{Y}_{k3}^T R^{-1} \vec{f}_{k3}(\hat{\vec{\theta}}_3) - \\ &\quad - \frac{1}{2} \sum_k \vec{f}_{3k}^T(\hat{\vec{\theta}}_3) R^{-1} \vec{f}_{3k}(\hat{\vec{\theta}}_3) \\ \vec{Y}_{k3}^T &= \vec{u}_k^T - \vec{f}_{1k}^T(\hat{\vec{\theta}}_1) - \vec{f}_{2k}^T(\hat{\vec{\theta}}_2) \\ \hat{\vec{\theta}}_3^{(1)} &= \max_{\vec{\theta}_3} g_3(\vec{u}_k, \hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2, \hat{\vec{\theta}}_3). \end{aligned}$$

We check the threshold condition

If

$$\alpha_3^2 = \sum_k \vec{f}_{3k}^T(\hat{\vec{\theta}}_3^{(1)}) R^{-1} \vec{f}_{3k}(\hat{\vec{\theta}}_3^{(1)}) \geq \alpha_{0P_e}^2,$$

then the procedure will be continued.

If

$$\alpha_3^2 < \alpha_{0P_e}^2,$$

then the procedure will be stopped, and we say: we have any signals with parameters $\hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2$.

Then we reestimate parameters $\hat{\vec{\theta}}_2^{(1)}$ and $\hat{\vec{\theta}}_1^{(1)}$ and continue this iteration procedure, checking the satisfaction of the condition on each step

If

$$\frac{g_3(\hat{\vec{\theta}}_1^{(n)}, \hat{\vec{\theta}}_2^{(n)}, \hat{\vec{\theta}}_3^{(n)}) - g_3(\hat{\vec{\theta}}_1^{(n-1)}, \hat{\vec{\theta}}_2^{(n-1)}, \hat{\vec{\theta}}_3^{(n-1)})}{g_3(\hat{\vec{\theta}}_1^{(n-1)}, \hat{\vec{\theta}}_2^{(n-1)}, \hat{\vec{\theta}}_3^{(n-1)})} < \delta,$$

then we shall go to the next step $m = 4$.

If

$$\frac{\Delta g_3^{(n)}}{\Delta g_3^{(n-1)}} > \delta,$$

we continue the iteration procedure for vectors $\vec{\theta}_1, \vec{\theta}_2$ and $\vec{\theta}_3$.

Step m. We suggest that we have m signals.

$$\begin{aligned} M = m, \quad g_m(\vec{u}_k, \hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2, \dots, \hat{\vec{\theta}}_m) &= \sum_k \vec{Y}_{km}^T R^{-1} \vec{f}_{mk}(\hat{\vec{\theta}}_m) - \\ &\quad - \frac{1}{2} \sum_k \vec{f}_{mk}^T(\hat{\vec{\theta}}_m) R^{-1} \vec{f}_{mk}(\hat{\vec{\theta}}_m) \\ \vec{Y}_{km}^T &= \vec{u}_k^T - \sum_{\mu=1}^{m-1} \vec{f}_{\mu k}^T(\hat{\vec{\theta}}_\mu) \\ \hat{\vec{\theta}}_m^{(1)} &= \max_{\vec{\theta}_m} g_m(\vec{u}_k, \hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2, \dots, \hat{\vec{\theta}}_m). \end{aligned}$$

We check the threshold condition

If

$$\alpha_m^2 = \sum_k \vec{f}_{mk}^T(\hat{\vec{\theta}}_m^{(1)}) R^{-1} \vec{f}_{mk}(\hat{\vec{\theta}}_m^{(1)}) \geq \alpha_{0P_e}^2,$$

then the procedure will be continued.

If

$$\alpha_m^2 < \alpha_{0P_e}^2,$$

then the procedure will be stopped and we say: we have only $(m - 1)$ signals with parameters $\hat{\vec{\theta}}_1, \dots, \hat{\vec{\theta}}_{m-1}$.

Then we reestimate all parameters by the iteration procedure. We have $\hat{\vec{\theta}}_1^{(1)}, \dots, \hat{\vec{\theta}}_{m-1}^{(1)}, \hat{\vec{\theta}}_m^{(1)}$, then we implement the next iteration and to get $\hat{\vec{\theta}}_1^{(2)}, \dots, \hat{\vec{\theta}}_{m-1}^{(2)}, \hat{\vec{\theta}}_m^{(2)}$ and etc. Each time transmitting to the next iteration, we check the satisfaction of the condition.

If

$$\frac{g_m(\hat{\vec{\theta}}_1^{(n)}, \dots, \hat{\vec{\theta}}_{m-1}^{(n)}, \hat{\vec{\theta}}_m^{(n)}) - g_m(\hat{\vec{\theta}}_1^{(n-1)}, \dots, \hat{\vec{\theta}}_{m-1}^{(n-1)}, \hat{\vec{\theta}}_m^{(n-1)})}{g_m(\hat{\vec{\theta}}_1^{(n-1)}, \dots, \hat{\vec{\theta}}_{m-1}^{(n-1)}, \hat{\vec{\theta}}_m^{(n-1)})} < \delta,$$

then we shall go to the next step $M = m + 1$.

If

$$\frac{\Delta g_m^{(n)}}{\Delta g_m^{(n-1)}} > \delta,$$

we continue the iteration procedure for vectors $\vec{\theta}_1, \dots, \vec{\theta}_{m-1}, \vec{\theta}_m$.

For the practical application it is of great interest the modification of this iteration procedure without mutual correction of parameters. The modification procedure is the following.

Step 1. We suggest that we have only one signal:

$$M = 1, \quad g_1(\vec{u}_k, \vec{\theta}_1) = \sum_k [\vec{u}_k^T R^{-1} \vec{f}_{1k}(\vec{\theta}_1) - \frac{1}{2} \vec{f}_{1k}^T(\vec{\theta}_1) R^{-1} \vec{f}_{1k}(\vec{\theta}_1)]$$

$$\hat{\vec{\theta}}_1 = \max_{\vec{\theta}_1} g_1(\vec{u}_k, \vec{\theta}_1).$$

We check the threshold condition:

If

$$\alpha_1^2 = \sum_k \vec{f}_k^T(\hat{\vec{\theta}}_1) R^{-1} \vec{f}_k(\hat{\vec{\theta}}_1) \geq \alpha_{0P_e}^2,$$

then the procedure will be continued.

If

$$\alpha_1^2 < \alpha_{0P_e}^2,$$

then the procedure will be stopped, and we say: we do not have any signal.

Step 2. We suggest that we have only two signals:

$$M = 2, \quad g_2(\vec{u}_k, \hat{\vec{\theta}}_1, \vec{\theta}_2) = \sum_k \vec{Y}_{k2}^T R^{-1} \vec{f}_{2k}(\vec{\theta}_2) - \frac{1}{2} \sum_k \vec{f}_{2k}^T(\vec{\theta}_2) R^{-1} \vec{f}_{2k}(\vec{\theta}_2)$$

$$\vec{Y}_{k2}^T = \vec{u}_k^T - \vec{f}_{1k}^T(\hat{\vec{\theta}}_1).$$

For parameter vector $\vec{\theta}_1$, we use the estimate received on the first step $\hat{\vec{\theta}}_1$

$$\hat{\vec{\theta}}_2 = \max_{\vec{\theta}_2} g_2(\vec{u}_k, \hat{\vec{\theta}}_1, \vec{\theta}_2).$$

We check the threshold condition:

If

$$\alpha_2^2 = \sum_k \vec{f}_{2k}^T(\hat{\vec{\theta}}_2) R^{-1} \vec{f}_{2k}(\hat{\vec{\theta}}_2) \geq \alpha_{0P_e}^2,$$

then the procedure will be continued.

If

$$\alpha_2^2 < \alpha_{0P_e}^2,$$

then the procedure will be stopped, and we say: we have only one signal $\vec{\theta}_1$.

Step 3. We suggest that we have three signals

$$M = 3, \quad g_3(\vec{u}_k, \hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2, \vec{\theta}_3) = \sum_k \vec{Y}_{k3}^T R^{-1} \vec{f}_{3k}(\vec{\theta}_3) -$$

$$- \frac{1}{2} \sum_k \vec{f}_{3k}^T(\vec{\theta}_3) R^{-1} \vec{f}_{3k}(\vec{\theta}_3)$$

$$\vec{Y}_{k3}^T = \vec{u}_k^T - \vec{f}_{3k}^T(\hat{\vec{\theta}}_1) - \vec{f}_{3k}^T(\hat{\vec{\theta}}_2)$$

$$\hat{\vec{\theta}}_3 = \max_{\vec{\theta}_3} g_3(\vec{u}_k, \hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2, \vec{\theta}_3).$$

We check the threshold condition

If

$$\alpha_3^2 = \sum_k \vec{f}_k^T(\hat{\vec{\theta}}_3) R^{-1} \vec{f}_k(\hat{\vec{\theta}}_3) \geq \alpha_{0P_e}^2,$$

then the procedure will be continued.

If

$$\alpha_3^2 < \alpha_{0P_e}^2,$$

then the procedure will be stopped, and we say: we have any signals with parameters $\hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2$.

Step m. We suggest that we have m signals.

$$\begin{aligned} M = m, \quad g_m(\vec{u}_k, \hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2, \dots, \hat{\vec{\theta}}_m) &= \sum_k \vec{Y}_{km-1}^T R^{-1} \vec{f}_{mk}(\vec{\theta}_m) - \\ &\quad - \frac{1}{2} \sum_k \vec{f}_{mk}^T(\vec{\theta}_m) R^{-1} \vec{f}_{mk}(\vec{\theta}_m) \\ \vec{Y}_{km-1}^T &= \vec{u}_k^T - \sum_{\mu=1}^{m-1} \vec{f}_{\mu k}^T(\hat{\vec{\theta}}_\mu) \\ \hat{\vec{\theta}}_m &= \max_{\hat{\vec{\theta}}_m} g_m(\vec{u}_k, \hat{\vec{\theta}}_1, \hat{\vec{\theta}}_2, \dots, \hat{\vec{\theta}}_m). \end{aligned}$$

We check the threshold condition

If

$$\alpha_m^2 = \sum_k \vec{f}_{mk}^T(\hat{\vec{\theta}}_m) R^{-1} \vec{f}_{mk}(\hat{\vec{\theta}}_m) \geq \alpha_{0P_e}^2,$$

then the procedure will be continued.

If

$$\alpha_m^2 < \alpha_{0P_e}^2,$$

then the procedure will be stopped and we say: we have only $(m-1)$ signals with parameters $\hat{\vec{\theta}}_1, \dots, \hat{\vec{\theta}}_{m-1}$.

Example 5: For seismic case, we shall write the decision function and shall estimate the parameters for the arbitrary step m of the modified procedure. We suggest for simplicity, that random component is uncorrelated $R = \sigma^2 I$, and the signal is presented by

$$\vec{f}_k(\vec{\theta}_m) = \|A_m \varphi(t_i - \tau_m - k\Delta x \gamma_m)\|_{i=1}^n,$$

$$\begin{aligned} M = m, \quad g_m(\vec{u}_k, A_m, \tau_m, \gamma_m) &= \frac{A_m}{\sigma^2} \sum_i \varphi(t_i) \sum_k y(t_i + \tau_{mk}) - \\ &\quad - \frac{A_m^2}{2\sigma^2} K \sum_i \varphi^2(t_i), \end{aligned}$$

where

$$y(t_i + \tau_{mk}) = u(t_i + \tau_{mk}) - \sum_{\mu' \neq m} A_{\mu'} \varphi(t_i - \tau_{\mu'k} - \tau_{\mu k}).$$

We can find an explicit form of the amplitude estimate

$$\frac{\partial g_m}{\partial A_m} = 0 \Rightarrow \hat{A}_m = \frac{\sum_i \varphi(t_i) \sum_k y(t_i + \tau_{mk})}{\sum_i \varphi^2(t_i)}.$$

Substituting the estimate \hat{A}_m to the decision function $g_m(\vec{u}_k, A_m, \tau_m, \gamma_m)$, we shall obtain

$$g_m(\tau_m, \gamma_m) = \frac{1}{\sigma^2} \frac{[\sum_i \varphi(t_i) \sum_k y(t_i + \tau_{mk})]^2}{K \sum_i \varphi^2(t_i)}.$$

The estimates of the travel time τ_m and inverse apparent velocity γ_m of the signal m can be found by

$$\hat{\tau}_m, \hat{\gamma}_m = \max_{\tau_m, \gamma_m} g_m(\vec{u}_k, \hat{A}_m, \tau_m, \gamma_m).$$

The threshold condition is written as

If

$$\alpha_m^2 = \frac{\hat{A}_m^2}{\sigma^2} \sum_k \sum_i \varphi^2(t_i - \hat{\tau}_m - k\Delta x \hat{\gamma})) \geq \alpha_{0P_e}^2,$$

then we will go to the next step $M = m + 1$.

If

$$\alpha_m^2 < \alpha_{0P_e}^2,$$

the procedure will be stopped.

We shall take as the parameter vector

$$\hat{\theta}\{\hat{A}_1, \hat{\tau}_1, \hat{\gamma}_1, \hat{A}_2, \hat{\tau}_2, \hat{\gamma}_2, \dots, \hat{A}_m, \hat{\tau}_m, \hat{\gamma}_m\}.$$

Example 6: For magnetic case, we shall write the decision function and the estimate of parameters for the arbitrary step m again that the random component ε is uncorrelated $R = \sigma^2 I$ and the signal number μ is presented by

$$f_{k\mu} = \frac{I_\mu h_\mu}{h_\mu^2 + (x_k - \xi_\mu^2)} = I_\mu \varphi_k(h_\mu, \xi_\mu),$$

where

$$\varphi_k(h_\mu, \xi_\mu) = \frac{h_\mu}{h_\mu^2 + (x_k - \xi_\mu^2)}.$$

$$\begin{aligned} M = m, \quad g_m(u_k, I_m, h_m, \xi_m) &= \frac{I_m}{\sigma^2} \sum_k y_{km} \varphi(h_m, \xi_m) - \\ &- \frac{1}{2} \frac{I_m^2}{\sigma^2} \sum_k \varphi_k^2(h_m, \xi_m), \end{aligned}$$

where

$$y_{km} = u_k - \sum_{\mu' \neq m} I_{\mu'} \varphi_k(h_{\mu'}, \xi_{\mu'}).$$

We can find an explicit form of the magnetization estimate:

$$\frac{\partial g_m}{\partial I_m} = 0, \quad \hat{I}_m = \frac{\sum_k y_{km} \varphi_k(h_m, \xi_m)}{\sum_k \varphi_k^2(h_m, \xi_m)}.$$

Substituting the estimate \hat{I}_m into decision function $g_m(u_k, I_m, h_m, \xi_m)$, we shall obtain

$$g_m(h_m, \xi_m) = \frac{1}{\sigma^2} \frac{\sum_k y_{km} \varphi_k(h_m, \xi_m)}{\sum_k \varphi_k^2(h_m, \xi_m)}$$

and the estimate of the parameter h_m (depth of the upper edge) and ξ_m (horizontal coordinate) of the thin vertical magnetic layer with the number m will be found by the maximization of the function g_m

$$h_m, \xi_m = \max g_m(u_k, \hat{I}_m, h_m, \xi_m).$$

The threshold condition will be written as

If

$$\alpha_m^2 = \frac{\hat{I}_m^2}{\sigma_m^2} \sum_k \varphi_k^2(\hat{h}_m, \hat{\xi}_m) \geq \alpha_{0P_e}^2,$$

then we will go to the next step $M = m + 1$.

If

$$\alpha_m^2 < \alpha_{0P_e}^2,$$

the procedure will be stopped.

We shall take as the estimate of the parameter vector

$$\hat{\theta} = \{\hat{I}_1, \hat{h}_1, \hat{\xi}_1, \hat{I}_2, \hat{h}_2, \hat{\xi}_2, \dots, \hat{I}_m, \hat{h}_m, \hat{\xi}_m\}.$$

9

Ray geophysical tomography method

9.1 Definition of ray tomography

Let $f(\vec{r})$ be an unknown function and u_i is an integral projection from the pair number i source and receiver $(s, r)_i$. The relation between u_i and $f(\vec{r})$ can be written as an integral along the ray L_i

$$u_i = \int_{L_i} f(\vec{r}) dl.$$

The integral can be rewritten in the form

$$u_i = \int_{\Omega} f(\vec{r}) \delta(L_i(\vec{r})) d\vec{r},$$

where Ω is an integration domain (surface or volume), $\delta(\cdot)$ is a delta-function. We denote the integral operator as \tilde{R}

$$u_i = \tilde{R}_i[f(\vec{r})], \quad \tilde{R}_i[\cdot] \equiv \int d\vec{r} \delta(L_i(\vec{r}))[\cdot].$$

Introducing the polar coordinates for 2-D case we can write the unknown function $f(\vec{r})$ as $f(r, \varphi)$ by the use of $\vec{r} = (r, \varphi)$. Going over to other pair of parameters (n, θ) we shall obtain the direct Radon transform

$$(n, \theta), \quad u(n, \theta) = \int_L f(\sqrt{l^2 + n^2}, \theta - \arctan(l/n)) dl.$$

The parameters relating to the transformations concerning the ray tomography is shown at Figure 9.1.

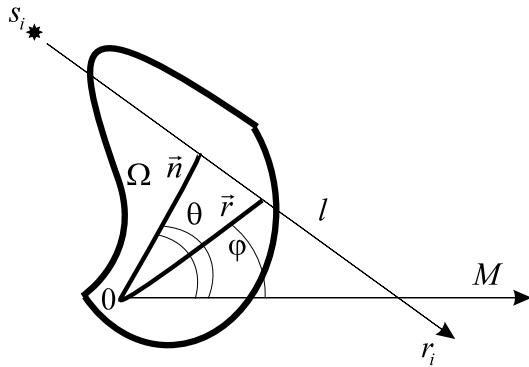


Figure 9.1: The main parameters relating to transformation concerning the ray tomography.

9.2 Radon transform

The direct Radon transform has been defined above

$$u(n, \theta) = \tilde{R}[f(r, \varphi)].$$

The inversion Radon transform can be written in the form

$$f(r, \varphi) = \frac{1}{2\pi^2} \int_{-\infty}^{\infty} \int_0^{\pi} \frac{1}{r \cos(\theta - \varphi) - n} \frac{\partial u(n, \theta)}{\partial n} dnd\theta,$$

or in the operator form:

$$f(r, \varphi) = \tilde{R}^{-1}[u(n, \theta)].$$

The inverse Radon operator can be presented as a sequence of the operators:

$$\tilde{R}^{-1} = -\frac{1}{2\pi} PHD,$$

where

$$D : q(n, \theta) = D[u(n, \theta)] = \lim_{\Delta n \rightarrow 0} \frac{u(n + \Delta n, \theta) - u(n, \theta)}{\Delta n}$$

is a *differential operator* of the first derivative;

$$H : h(n', \theta) = H[q(n', \theta)] = -\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0} \left\{ \int_{-\infty}^{n' - \varepsilon} \frac{q(n, \theta)}{n' - n} dn + \int_{n' + \varepsilon}^{\infty} \frac{q(n, \theta)}{n' - n} dn \right\}$$

is an operator of *Hilbert transform*;

$$P : p(r, \varphi) = P[h(n', \theta)] = \int_0^\pi h(r \cos(\theta - \varphi), \theta) d\theta$$

is an operator of *back projection*.

The Radon inversion is used rarely, because we never have a continuous field of data. We consider the most important algebraic methods.

9.3 Algebraic methods

We shall represent the unknown function $f(\vec{r})$ as an expansion on the basis functions

$$f(\vec{r}) = \sum_{j=1}^M \theta_j \psi_j(\vec{r}).$$

The observation data u_i can be written in this case as

$$u_i = \sum_{j=1}^M C_{ij} \theta_j,$$

where

$$C_{ij} = \tilde{R}_i[\psi_j(\vec{r})].$$

The simplest basis function are characteristic functions such that

$$\psi_j(\vec{r}) = \begin{cases} 1 & \text{for } \vec{r} \in \Omega_j, \\ 0 & \text{for } \vec{r} \notin \Omega_j. \end{cases}$$

The elements C_{ij} of the projective matrix in this case will be represented as

$$C_{ij} = \begin{cases} C & \text{for } L_i \cap \Omega_j \neq \emptyset, \\ 0 & \text{for } L_i \cap \Omega_j = \emptyset. \end{cases}$$

The main elements of a scheme of the algebraic methods is presented at Figure 9.2.

The model of the measurement data can be written as

$$\vec{u} = C\vec{\theta} + \vec{\varepsilon},$$

where $\vec{\varepsilon}$ is the random component ($\vec{\varepsilon} \in N(0, R_\varepsilon)$) and $\vec{\theta}$ is the random parameter vector ($\vec{\theta} \in N(\langle \vec{\theta} \rangle, R_\theta)$).

Our problem is to estimate of parameter vector $\vec{\theta}$. We shall use for the estimation the method of a posteriori probability:

$$\tilde{\vec{\theta}} = \max_{\vec{\theta}} p(\vec{\theta}) p(\vec{u}/\vec{\theta}).$$

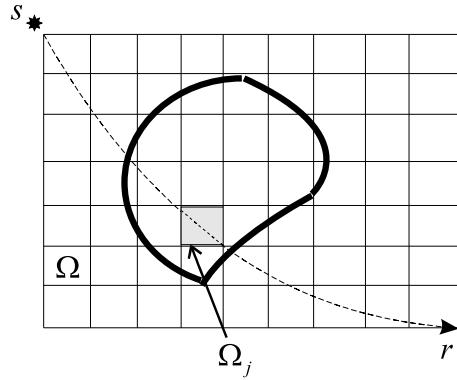


Figure 9.2: The main elements of a scheme of application of the algebraic methods.

In the case of the normal distribution of $\vec{\varepsilon}$ and $\vec{\theta}$ this criterion is written as

$$\tilde{\vec{\theta}} = \min_{\vec{\theta}} [(\vec{u} - C\vec{\theta})^T R_\varepsilon^{-1} (\vec{u} - C\vec{\theta}) + (\vec{\theta} - \langle \vec{\theta} \rangle)^T R_\theta^{-1} (\vec{\theta} - \langle \vec{\theta} \rangle)].$$

The explicit form of the estimate and its covariance matrix can be represented in two forms:

$$1. \quad \hat{\vec{\theta}} = (C^T R_\varepsilon^{-1} C + R_\theta^{-1})^{-1} (C^T R_\varepsilon^{-1} \vec{u} + R_\theta^{-1} \langle \vec{\theta} \rangle), \\ R(\hat{\vec{\theta}}) = (C^T R_\varepsilon^{-1} C + R_\theta^{-1})^{-1}.$$

$$2. \quad \tilde{\vec{\theta}} = \langle \vec{\theta} \rangle + R_\theta C^T (C R_\theta C^T + R_\varepsilon)^{-1} (\vec{u} - C \langle \vec{\theta} \rangle), \\ R(\tilde{\vec{\theta}}) = R_\theta - R_\theta C^T (C R_\theta C^T + R_\varepsilon)^{-1} C R_\theta.$$

In the particular case of the uncorrelated random noise $R_\varepsilon = \sigma_\varepsilon^2 I$ and the elements of the parameter vector $R_\theta = \sigma_\theta^2 I$, the estimates and relevant covariance matrix will be written as

$$1. \quad \hat{\vec{\theta}} = (C^T C + \alpha I)^{-1} (C^T \vec{u} + \alpha \langle \vec{\theta} \rangle), \\ R(\hat{\vec{\theta}}) = \sigma_\varepsilon^2 (C^T C + \alpha I)^{-1},$$

where $\alpha = \sigma_\varepsilon^2 / \sigma_\theta^2$ is the Tikhonov regularization parameter.

$$2. \quad \tilde{\vec{\theta}} = \langle \vec{\theta} \rangle + C^T (C C^T + \alpha I)^{-1} (\vec{u} - C \langle \vec{\theta} \rangle), \\ R(\tilde{\vec{\theta}}) = \sigma_\theta^2 - \sigma_\theta^2 C^T (C C^T + \alpha I)^{-1} C.$$

The covariance matrices $R(\hat{\vec{\theta}})$ and $R(\tilde{\vec{\theta}})$ can be used for the estimate for the interpretation quality. The variances of the recovery error lie on the main diagonal of the matrices $R(\hat{\vec{\theta}})$ and $R(\tilde{\vec{\theta}})$ and the correlation connections of the recovery errors in the different space points is described by off-diagonal elements.

We consider a recurrent algorithm which takes into account automatically the presence in the matrix C many zero-elements, which is based on the estimate $\tilde{\vec{\theta}}$.

As an initial parameter vector $\vec{\theta}^{(0)}$ and covariance matrix $R_{\theta}^{(0)}$, we take a priori vector

$$\vec{\theta}^{(0)} = \langle \vec{\theta} \rangle$$

and a priori covariance matrix

$$R_{\theta}^{(0)} = R_{\theta}.$$

The first approximation of the estimate $\vec{\theta}$ is calculated on the formula

$$\vec{\theta}^{(1)} = \vec{\theta}^{(0)} + \alpha^{(1)} \Delta \vec{\theta}^{(1)},$$

where

$$\begin{aligned} \Delta \vec{\theta}^{(1)} &= R_{\theta}^{(0)} C^{(1)T}, & \alpha^{(1)} &= e_1 / \delta_1^2, \\ e_1 &= u_1 - C^{(1)} \vec{\theta}^{(1)}, & \delta_1^2 &= \sigma_{\varepsilon 1}^2 + C^{(1)} \Delta \vec{\theta}^{(1)}, \end{aligned}$$

where $C^{(1)}$ is the first row of the matrix C . The first approximation of the covariance matrix is given by

$$R_{\theta}^{(1)} = R_{\theta}^{(0)} - \Delta \vec{\theta}^{(1)} \Delta \vec{\theta}^{(1)T} / \delta_1^2.$$

Let the $(i-1)$ approximation be

$$\vec{\theta}^{(i-1)}, \quad R_{\theta}^{(i-1)}.$$

We shall obtain i -th approximation of the estimate $\vec{\theta}$

$$\vec{\theta}^{(i)} = \vec{\theta}^{(i-1)} + \alpha^{(i)} \Delta \vec{\theta}^{(i)},$$

where

$$\alpha^{(i)} = e_i / \delta_i^2, \quad e_i = u_i - C^{(i)} \vec{\theta}^{(i-1)}, \quad \delta_i^2 = \sigma_{\varepsilon i}^2 + C^{(i)} \Delta \vec{\theta}^{(i)}, \quad \Delta \vec{\theta}^{(i)} = R_{\theta}^{(i-1)} C^{(i)T},$$

$C^{(i)}$ is the i -th row of the matrix C . The i -th approximation of the covariance matrix will be written as

$$R_{\theta}^{(i)} = R_{\theta}^{(i-1)} - \Delta \vec{\theta}^{(i)} \Delta \vec{\theta}^{(i)T} / \delta_i^2.$$

The number of approximations is different by the number of rays.

We can estimate the approximation quality with respect to a priori representation on each step i using the information about the parameter θ_j

$$I_j^{(i)} = \frac{1}{2} \ln \frac{R_{\theta jj}^{(0)}}{R_{\theta jj}^{(i)}}, \quad j = 1, 2, \dots, M.$$

The contribution of the new ray to the information is given by the information field

$$I_j^{(i-1) \rightarrow (i)} = \frac{1}{2} \ln \frac{R_{\theta jj}^{(i-1)}}{R_{\theta jj}^{(i)}}, \quad j = 1, 2, \dots, M.$$

9.4 Slowness and absorption coefficient reconstruction

Let sources be located on the plane $z = 0$ and geophones in the borehole. We assume there are information about the background medium in the form of a slowness $S_0(r) = 1/v_0(r)$ and absorption coefficient $\alpha_0(r)$. Some practical schemes of application of the ray tomography is presented at Figure 9.3.

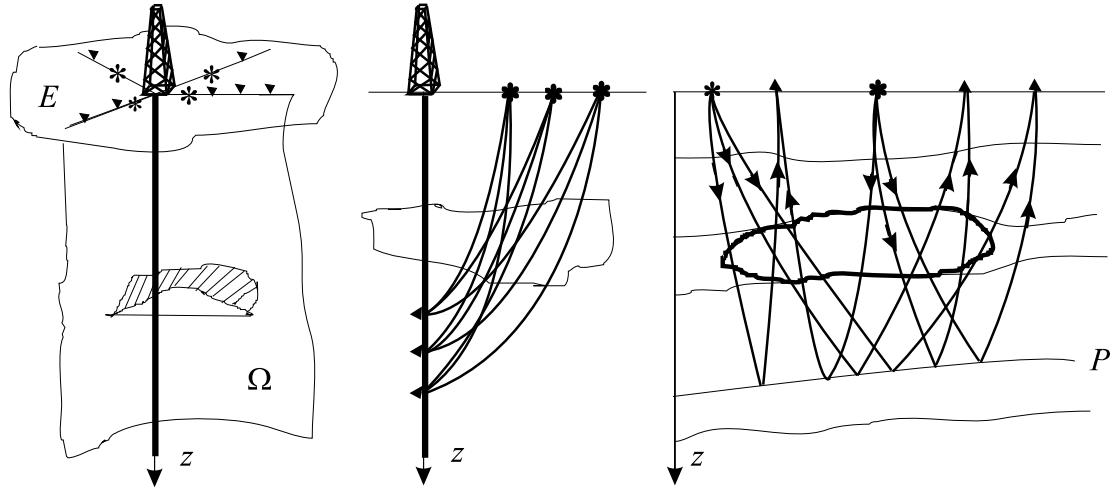


Figure 9.3: The schemes of applications of the ray tomography.

We can calculate the spectrum of seismic field for each pair source-receiver in the background medium with the functions $S_0(r)$ and $\alpha_0(r)$:

$$u_{0i}(\omega) = A_i(\omega, r_{ri}, r_{si}, \alpha_0(r), S_0(r)) \exp(-i\omega\tau_0(r_{ri}, r_{si})),$$

where (r_{ri}, r_{si}) are the coordinate accordingly of receiver and source, τ_0 is a wave travel time. We note, the functional A_i describes the forward problem. We assume that the deviations $\delta\alpha$ and δS of the real function $\alpha(r)$ and $S(r)$ from the $\alpha_0(r)$ and $S_0(r)$ are small and we can use the perturbation method

$$\begin{aligned} \delta\alpha(r) &= \alpha(r) - \alpha_0(r), & \delta\alpha(r) &\ll \alpha_0(r). \\ \delta S(r) &= S(r) - S_0(r), & \delta S(r) &\ll S_0(r). \end{aligned}$$

We multiply the function on

$$\exp(-i\omega\tau_1),$$

where a complex quantity $\tau_1 = \tau_{1\Re} + i\tau_{1\Im}$ with the real part $\tau_{1\Re}$ determines the correction of the wave travel time, and imaginary part $\tau_{1\Im}$ determines the correction of the absorption coefficient. We assume that the complete time

$$\tau = \tau_0 + \tau_1$$

satisfies to the eikonal equation

$$(\vec{\nabla}\tau)^2 = S^2(r),$$

where $S(r) = S_0(r) + S_1(r)$ and $S_1(r) = S_{1\Re} + S_{1\Im}$, where $S_{1\Re}$ and $S_{1\Im}$ are accordingly the real part and imaginary part of the complex function $S_1(r)$. Thus we can write

$$(\vec{\nabla}(\tau_0 + \tau_1))^2 = (S_0 + S_1)^2$$

and

$$(\vec{\nabla}\tau_0)^2 + 2\vec{\nabla}\tau_0\vec{\nabla}\tau_1 + (\vec{\nabla}\tau_1)^2 = S_0^2 + 2S_0S_1 + S_1^2.$$

Taking into account that

$$(\vec{\nabla}\tau_1)^2 \approx 0, \quad S_1^2 \approx 0, \quad (\vec{\nabla}\tau_0)^2 = S_0^2,$$

we can write

$$\vec{\nabla}\tau_0\vec{\nabla}\tau_1 = S_0S_1.$$

Granting that

$$\vec{\nabla}\tau_0 = S_0\vec{l}$$

where \vec{l} is an identity vector of the ray tangent, we shall obtain the equation

$$\vec{\nabla}_l\tau_1 = S_1 \Rightarrow \vec{\nabla}_l\tau_{1\Re} = S_{1\Re}, \quad \vec{\nabla}_l\tau_{1\Im} = S_{1\Im},$$

where $\vec{\nabla}_l$ is the derivative operator on direction \vec{l} .

Integrating the equations along the ray L we obtain

$$\tau_{1\Re} = \int_L S_{1\Re} dl, \quad \tau_{1\Im} = \int_L S_{1\Im} dl.$$

The model of the observed travel time is written as

$$\tau_i - \tau_{i0} = \int_{L_i} S_{1\Re}(r) dl + \varepsilon_i.$$

The model of the observed amplitude spectrum

$$|u_i(\omega)|_{\max} = |u_{0i}(\omega)|_{\max} \exp(-\alpha_i(l)L_i) + \varepsilon_i,$$

where

$$\begin{aligned} \alpha_i(l) &= \frac{1}{L_i} \int_{L_i} \alpha_l dl, \quad \varepsilon_i \ll |u_{0i}(\omega)|_{\max} \exp(-\alpha_i L_i), \\ \ln \frac{|u_i(\omega)|_{\max}}{|u_{0i}(\omega)|_{\max}} &= - \int_{L_i} \alpha_l(r) dl + \tilde{\varepsilon}_i, \end{aligned}$$

where

$$\tilde{\varepsilon}_i = \frac{\varepsilon_i}{|u_{0i}(\omega)|_{\max} \exp(-\alpha_i L_i)}, \quad \ln(1 + \tilde{\varepsilon}_i) \approx \tilde{\varepsilon}_i.$$

We can use these models for the recovery of the slowness $S(r)$ and the absorption coefficient $\alpha_l(r)$.

9.5 Radon transform for seismogram processing

The direct Radon transform of seismogram is written as

$$s(p, \tau) = \int_{-\infty}^{\infty} u(x, px + \tau) dx$$

(see Figure 9.4). The inverse Radon transform is represented in the form

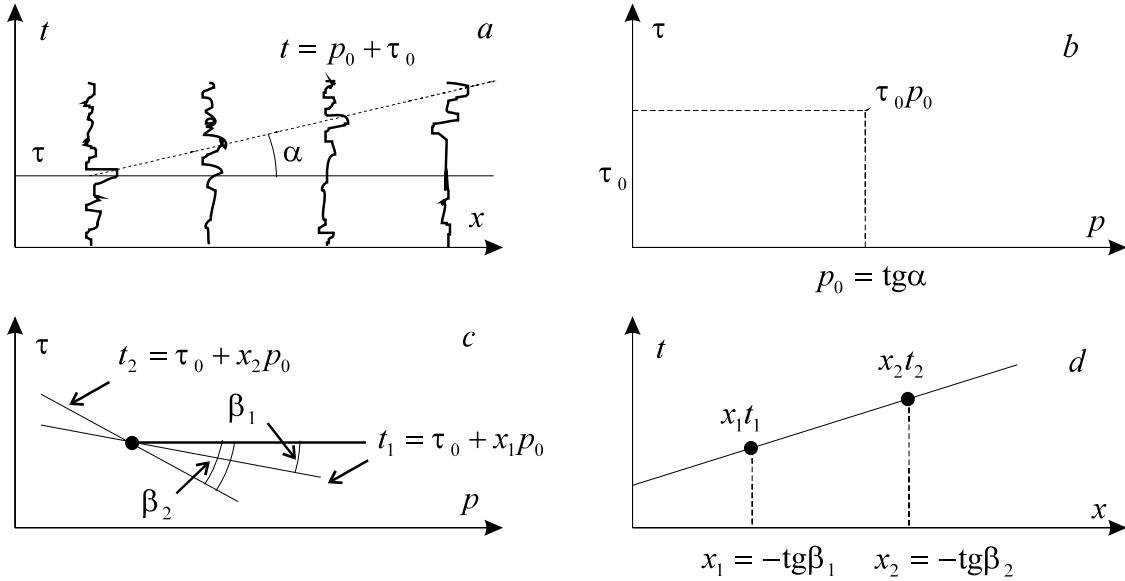


Figure 9.4: The Radon transform and inverse Radon transform. A line in space-time coordinate (a) and a point in the plane – angle coefficient – reflection time. A point in the plane with the coordinates angle coefficient – reflection time (c) and a line in the space-time coordinates (d).

$$u(x, t) = - \int_{-\infty}^{\infty} H \frac{d}{d\tau} s(p, \tau) dp = -PHD[s(p, \tau)]$$

or

$$u(x, t) = - \int_{-\infty}^{\infty} \left[-\frac{1}{\pi\tau} * \frac{\partial}{\partial\tau} s(p, \tau) \right] dp.$$

We shall write the spectrum of seismogram

$$u(k_x, \omega) = \iint_{-\infty}^{\infty} u(x, t) \exp[i(\omega t - k_x x)] dx dt$$

(see Figure 9.5). Using the relation $k_x = p\omega$, we can rewrite the spectrum $u(k_x, \omega)$ in

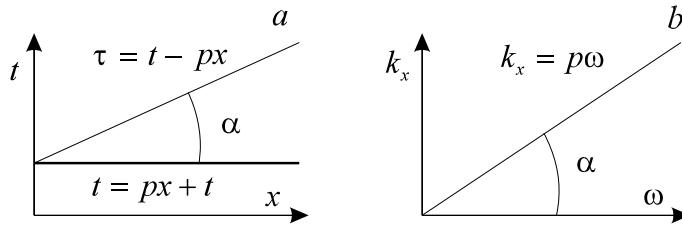


Figure 9.5: A line in space-time coordinate (a) and a line in the plane with coordinate wave number – frequency.

the form

$$u(p\omega, \omega) = \iint_{-\infty}^{\infty} u(x, t) \exp[i(\omega t - p\omega x)] dx dt$$

or, introducing new variable $t = \tau + px$, we shall have

$$\begin{aligned} u(p\omega, \omega) &= \int_{-\infty}^{\infty} e^{i\omega\tau} \left[\int_{-\infty}^{\infty} u(x, \tau + px) dx \right] d\tau = \\ &= \int_{-\infty}^{\infty} e^{i\omega\tau} s(p, \tau) d\tau = FR[u(x, t)], \end{aligned}$$

where $s(p, \tau) = \int_{-\infty}^{\omega} u(x, \tau + px) dx$ is the Radon transform of a seismogram u . Thus the 2-D Fourier transform can be represented as a successive application of the Radon transform and 1-D Fourier transform. The filtering in the $(k - \omega)$ domain is reduced to the integration in limits from k_{min} to k_{max} (see Figure 9.6)

$$u(x) = \frac{1}{2\pi} \int_{k_{min}}^{k_{max}} u(k) \exp(ikx) dk,$$

where $p = |\omega|p$ or

$$u(x) = \frac{|\omega|}{2\pi} \int_{p_{min}}^{p_{max}} u(p) \exp(i\omega p x) dp \quad (dk = |\omega|dp),$$

The reconstruction of seismogram with the filtering can be represented as

$$\begin{aligned} u(x, t) &= \frac{1}{4\pi^2} \int_{\omega_{min}}^{\omega_{max}} \int_{p_{min}}^{p_{max}} u(\omega, p) |\omega| \exp[-i\omega(t - px)] dp d\omega = \\ &= \frac{1}{2\pi} \int_{p_{min}}^{p_{max}} dp \left[\frac{1}{2\pi} \int_{\omega_{min}}^{\omega_{max}} \exp[-i\omega(t - px)] \times \right. \end{aligned}$$

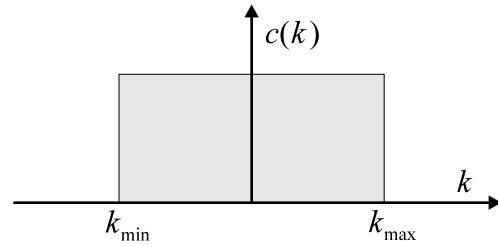


Figure 9.6: A window function for integration in k -domain.

$$\begin{aligned} & \times \left. FR[u(x, t)] |\omega| d\omega \right] = \\ & = \frac{1}{2\pi} \int_{p_{\min}}^{p_{\max}} R[u(x, t)] * (F^{-1}[|\omega|]) dp \end{aligned}$$

(see Figure 9.7).

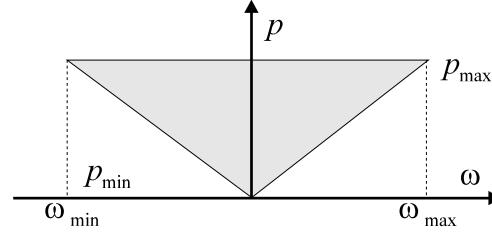


Figure 9.7: A region for integration in (ω, k) plane.

9.6 Hilbert transform and analytical signal

We assume that the function $S(t)$ has the boundary norm L^p , i.e.

$$S(t) \in L^p(-\infty, \infty), \quad \int_{-\infty}^{\infty} |S(t)|^p dt < \infty, \quad p \geq 1.$$

In this case, the pair of the Hilbert transform will be written as

$$Q(t) = -\frac{1}{\pi} \text{V.P.} \int_{-\infty}^{\infty} \frac{S(\tau)}{t - \tau} d\tau = -S(t) * \frac{1}{\pi t}$$

and

$$S(t) = \frac{1}{\pi} \text{V.P.} \int_{-\infty}^{\infty} \frac{Q(\tau)}{t - \tau} d\tau,$$

where *V.P.* means that we take the Cauchy principle value at $t = \tau$. The Fourier spectrum of the Hilbert transform $Q(t)$ becomes

$$Q(\omega) = \int_{-\infty}^{\infty} Q(t) \exp(i\omega t) dt = -\frac{1}{\pi} \int_{-\infty}^{\infty} S(\tau) \int_{-\infty}^{\infty} \frac{\exp(i\omega t)}{\tau - t} dt d\tau$$

and changing the variable $\tau - t = u$ we continue:

$$\begin{aligned} Q(\omega) &= -\frac{1}{\pi} \int_{-\infty}^{\infty} S(\tau) \exp(i\omega\tau) \int_{-\infty}^{\infty} \frac{\exp(i\omega u)}{u} du d\tau = \\ &= -S(\omega) \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\exp(i\omega u)}{u} du = \\ &= -S(\omega) \frac{2i}{\pi} \int_0^{\infty} \frac{\sin \omega u}{u} du = -iS(\omega) \operatorname{sign} \omega, \end{aligned}$$

since

$$\int_0^{\infty} \frac{\sin \omega u}{u} du = \frac{\pi}{2} \operatorname{sign} \omega$$

(see Figure 9.8). We note that

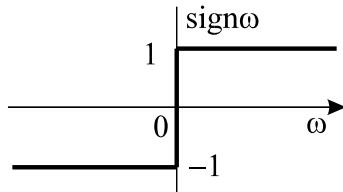


Figure 9.8: Graphic presentation of a function sign ω .

$$|Q(\omega)| = |S(\omega)|, \quad \arg Q(\omega) = \arg S(\omega) \pm \pi/2.$$

The quantity $Q(\omega)$ is a quadrature filter, which changes the phase by $\pm\pi/2$, but has no effect on the amplitude spectrum.

Let $Z(t)$ be a complex time function or analytical signal:

$$Z(t) = S(t) + iQ(t) \quad \text{or} \quad Z(t) = a(t) \exp(i\varphi(t)),$$

where

$$\begin{aligned} S(t) &= \Re Z(t) = a(t) \cos \varphi(t), \\ Q(t) &= \Im Z(t) = a(t) \sin \varphi(t). \end{aligned}$$

The envelope and phase function can be written accordingly as

$$a(t) = (S^2(t) + Q^2(t))^{1/2} \quad \text{and} \quad \varphi(t) = \arctan(q(t)/S(t)).$$

The envelope function and the phase function of the analytical signal are used for the interpretation of the data.

10

Diffraction tomography method

Diffraction tomography is a geophysical imaging technique with high potential to achieve maximum spatial resolution with minimum image distortion for seismic or acoustic data processing. In diffraction tomography the scattered field (or diffraction field) is used to reconstruct the object function.

10.1 Diffraction tomography in the scalar case

To demonstrate the basic principle of diffraction tomography we consider the case of the acoustic wave equation in the source free region

$$\Delta u(\vec{x}, t) - \frac{1}{v^2(\vec{x})} \frac{\partial^2}{\partial t^2} u(\vec{x}, t) = 0, \quad x \in R^3,$$

where $u(\vec{x}, t)$ is a field of pressure, $v(\vec{x})$ is unknown acoustic velocity. We assume a constant background velocity $v(\vec{x}) = \text{const}$ (see Figure 10.1). The Fourier spectrum

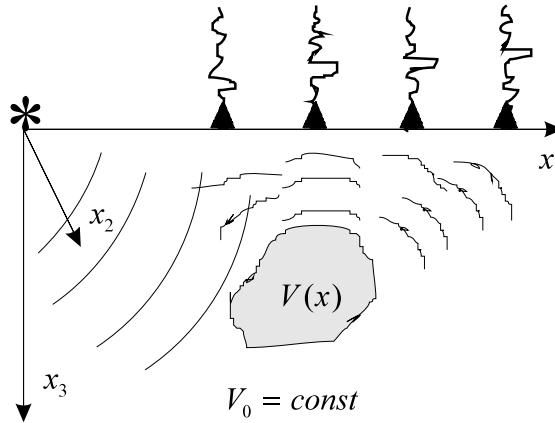


Figure 10.1: A scheme of application of the diffraction tomography method.

of pressure field reads as following

$$u(\vec{x}, \omega) = \int_{-\infty}^{\infty} u(\vec{x}, t) \exp(i\omega t) dt.$$

Consider the *perturbation method* for scalar wave equation. The inverse Fourier transform will be written as

$$u(\vec{x}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(\vec{x}, \omega) \exp(-i\omega t) d\omega.$$

Substituting this expression for $u(\vec{x}, t)$ into scalar wave equation, we shall obtain the Helmholtz equation

$$\Delta u(\vec{x}, \omega) + \frac{\omega^2}{v^2(\vec{x})} u(\vec{x}, \omega) = 0.$$

Adding a term $(\omega^2/v^2(\vec{x}))u(\vec{x}, \omega)$ to the right and left parts of the equation we shall set

$$\Delta u(\vec{x}, \omega) + \frac{\omega^2}{v_0^2} u(\vec{x}, \omega) = \beta(\vec{x}, \omega) u(\vec{x}, \omega),$$

where

$$\beta(\vec{x}, \omega) = \frac{\omega^2}{v_0^2} \left(1 - \frac{v_0^2}{v^2(\vec{x})} \right) = \frac{\omega^2}{v_0^2} \beta_0(\vec{x}).$$

We can consider the term $\beta(\vec{x}, \omega)u(\vec{x}, \omega)$ as a source function and seek the solution by means of the perturbation method:

$$u(\vec{x}, \omega) = u_0(1 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots),$$

where ε is a small parameter ($\varepsilon \ll 1$). Substituting the expansion $u(\vec{x}, \omega)$ to the Helmholtz equation, we shall obtain

$$\begin{aligned} \left(\Delta + \frac{\omega^2}{v_0^2} \right) u_0(1 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots) &= \\ &= \varepsilon \tilde{\beta} u_0(1 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots), \end{aligned}$$

where $\tilde{\beta} = \beta/\varepsilon \sim O(1)$.

Equating the terms with the equal power of ε

$$\begin{aligned} \varepsilon^0 : \quad & \left(\Delta + \frac{\omega^2}{v_0^2} \right) u_0(\vec{x}, \omega) = 0, \\ \varepsilon^1 : \quad & \underbrace{\left(\Delta + \frac{\omega^2}{v_0^2} \right) u_0(\vec{x}, \omega)}_{=0} u_1(\vec{x}, \omega) = \tilde{\beta} u_0(\vec{x}, \omega), \end{aligned}$$

and taking into account

$$\begin{aligned} \Delta(u_0 u_1) &= u_1 \Delta u_0 + u_0 \Delta u_1 + 2(\vec{\nabla} u_0, \vec{\nabla} u_1), \\ u_1 \underbrace{(\Delta + \frac{\omega^2}{v_0^2}) u_0}_{=0} + u_0 \Delta u_1 + 2(\vec{\nabla} u_0, \vec{\nabla} u_1) &= \tilde{\beta} u_0, \end{aligned}$$

we shall get

$$\varepsilon^1 : \quad \Delta u_1 + \frac{2(\vec{\nabla} u_0, \vec{\nabla} u_1)}{u_0} = \tilde{\beta}.$$

It is easily to obtain the recursion relation for arbitrary number of approximations

$$\begin{aligned}\Delta u_0 + \frac{\omega^2}{v_0^2} u_0 &= 0, \\ \Delta u_1 + \frac{2(\vec{\nabla} u_0, \vec{\nabla} u_1)}{u_0} &= \tilde{\beta}, \\ \Delta u_2 + \frac{2(\vec{\nabla} u_0, \vec{\nabla} u_2)}{u_0} &= \tilde{\beta} u_1, \\ \dots &\dots \\ \Delta u_n + \frac{2(\vec{\nabla} u_0, \vec{\nabla} u_n)}{u_0} &= \tilde{\beta} u_{n-1}.\end{aligned}$$

Consider an incident of a plane wave u_0 (see Figure 10.2) on an object located inside an homogeneous infinite medium:

$$u_0 = a_0 \exp[i(\vec{k}_0, \vec{x})], \quad \vec{k}_0 = \vec{n}\omega/v_0, \quad \vec{n} = \vec{\nabla} u / |\vec{\nabla} u|.$$

We restrict ourselves to the first order approximations:

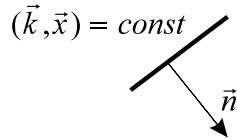


Figure 10.2: Graphic presentation of the relation between vectors \vec{k}_0 , \vec{x} and \vec{n} .

$$\Delta u_1 + 2i(\vec{k}_0, \vec{\nabla})u_1 = \tilde{\beta}.$$

To write corresponding equation for the Green function

$$(\Delta + 2i(\vec{k}_0, \vec{\nabla}))G(\vec{x}, \vec{x}') = \delta(\vec{x} - \vec{x}').$$

Substituting the inverse Fourier transform for the Green function and δ -function

$$\begin{aligned}G(\vec{x}, \vec{x}') &= \frac{1}{(2\pi)^3} \iiint_{-\infty}^{\infty} G(\vec{k}, \vec{x}') \exp[-i(\vec{k}, \vec{x})] d\vec{k}, \\ \delta(\vec{x} - \vec{x}') &= \frac{1}{(2\pi)^3} \iiint_{-\infty}^{\infty} \exp[-i(\vec{k}, \vec{x} - \vec{x}')] d\vec{k},\end{aligned}$$

into equation for the Green function, we shall obtain the equation with respect to the spectrum $G(\vec{k}, \vec{x})$

$$(-k^2 + 2(\vec{k}_0, \vec{k}))G(\vec{k}, \vec{x}') = \exp[i(\vec{k}, \vec{x}')]$$

and

$$G(\vec{k}, \vec{x}') = \frac{\exp[i(\vec{k}, \vec{x}')]])}{-k^2 + 2(\vec{k}_0, \vec{k})}.$$

Taking the inverse Fourier transform, we shall get the Green function for the first order approximation

$$G(\vec{x}, \vec{x}') = \frac{1}{(2\pi)^3} \iiint_{-\infty}^{\infty} \frac{\exp\{i[(\vec{k}, \vec{x}') - (\vec{k}, \vec{x})]\}}{-k^2 + 2(\vec{k}_0, \vec{k})} d\vec{k}.$$

Changing the variables $\vec{k}' = \vec{k} - \vec{k}_0$, $\vec{k} = \vec{k}' + \vec{k}_0$ we get

$$\begin{aligned} G(\vec{x}, \vec{x}') &= \frac{1}{(2\pi)^3} \iiint_{-\infty}^{\infty} \frac{\exp[i(\vec{k}', \vec{x}' - \vec{x})] \exp[i(\vec{k}_0, \vec{x}' - \vec{x})]}{-k'^2 - 2(\vec{k}', \vec{k}_0) - k_0^2 + 2k_0^2 + 2(\vec{k}_0, \vec{k}')} dk' = \\ &= \frac{\exp[i(\vec{k}_0, \vec{x}' - \vec{x})]}{(2\pi)^3} \iiint_{-\infty}^{\infty} \frac{\exp[i(\vec{k}', \vec{x}' - \vec{x})]}{k_0^2 - k'^2} dk'. \end{aligned}$$

We note that the last integral is the inverse Fourier transform from the well-known Green function for the Helmholtz equation:

$$\frac{1}{(2\pi)^3} \iiint_{-\infty}^{\infty} \frac{\exp[-i(\vec{k}', \vec{x}' - \vec{x})]}{k_0^2 - k'^2} dk' = -\frac{1}{(4\pi)} \frac{\exp[i|\vec{k}_0| |\vec{x} - \vec{x}'|]}{|\vec{x} - \vec{x}'|}.$$

Using this expression, we shall obtain

$$\begin{aligned} G(\vec{x}, \vec{x}') &= \exp[-i(\vec{k}_0, \vec{x} - \vec{x}')] \left(\frac{-1}{4\pi} \frac{\exp[i|\vec{k}_0| |\vec{x} - \vec{x}'|]}{|\vec{x} - \vec{x}'|} \right) = \\ &= \frac{-1}{4\pi} \frac{1}{|\vec{x} - \vec{x}'|} \exp[i|\vec{k}_0| |\vec{x} - \vec{x}'| (1 - \cos \theta)], \end{aligned}$$

where θ is the angle between vectors \vec{k}_0 and $\vec{x} - \vec{x}'$. Thus, we can write the first approximation of the scattered field as

$$\begin{aligned} u_1(\vec{x}, \omega) &= \iiint_{-\infty}^{\infty} G(\vec{x}, \vec{x}') \tilde{\beta}(\vec{x}') d\vec{x}' = \\ &= -\frac{1}{4\pi} \iiint_{-\infty}^{\infty} \frac{\tilde{\beta}(\vec{x}') \exp[i|\vec{k}_0| |\vec{x} - \vec{x}'| (1 - \cos \theta)]}{|\vec{x} - \vec{x}'|} d\vec{x}'. \end{aligned}$$

The Born approximation will be represented in the following form

$$\begin{aligned} u_B(\vec{x}, \omega) &= u_0(\vec{x}, \omega)(1 + \varepsilon u_1(\vec{x}, \omega)) = \\ &= a_0 \exp[i(\vec{k}_0, \vec{x})] + a_0 Q(\vec{x}, \omega, \vec{k}_0), \end{aligned}$$

where

$$\begin{aligned} Q(\vec{x}, \omega, \vec{k}_0) &= -\frac{\exp[i(\vec{k}_0, \vec{x})]}{4\pi} \iiint_{-\infty}^{\infty} \frac{\beta(\vec{x}, \omega)}{|\vec{x} - \vec{x}'|} \times \\ &\quad \times \exp[i|\vec{k}_0||\vec{x} - \vec{x}'|(1 - \cos\theta)] d\vec{x}'. \end{aligned}$$

The Born approximation requires the weak scatters (the parameter $\beta(x, \omega)$ should be small enough).

10.2 Born approximation for seismic trace model

The diffraction tomography can provide resolution on the order of one wavelength or less. The Born approximation is used to reconstruct the distribution of velocity or $\beta_0(x) = 1 - v_0^2/v^2(x)$ function for a weak scattering object.

The Born approximation holds when

$$\frac{\delta v}{v} \frac{d}{\lambda} \ll 1,$$

where $\delta v/v$ is the relative velocity contrast, d is the measure of the size of the scattering region and λ is a wavelength.

We construct the model of seismogram in the case of the Born approximation

$$\tilde{u}(t, \vec{x}_r, \vec{n}) = u_{mB}(t, \vec{x}_r, \vec{n}, \beta_0(\vec{x})) + \xi(t, \vec{x}_r, \vec{n}).$$

Here u_{mB} is the Born approximation of the scattered field

$$u_{mB}(t, \vec{x}_r, \vec{n}, \beta_0(\vec{x})) = h(t) * F_{\omega}^{-1}[u_B(\vec{x}, \omega, \beta_0(\vec{x}))],$$

where $h(t)$ is an apparatus function of the measurement channel, F_{ω}^{-1} is the operator of inverse Fourier transform,

$$u_B(\vec{x}, \omega) = a_0 \exp[i(\vec{k}_0, \vec{x})] + a_0 Q(\vec{x}_r, \omega, \vec{k}_0).$$

The main notations for introducing the Born approximation is presented at (see Figure 10.3). The function $Q(\vec{x}_r, \omega, \vec{k}_0)$ can be written in our case as

$$\begin{aligned} Q(\vec{x}_r, \omega, \vec{k}_0) &= -\frac{1}{4\pi} \iiint_{-\infty}^{\infty} \frac{\beta_0(\vec{x}) \omega^2}{v_0^2 |\vec{x} - \vec{x}'|} \exp[i|\vec{k}_0||\vec{x}_r - \vec{x}'|] \times \\ &\quad \times \exp[i(\vec{k}_0, \vec{x})] \exp[-i(\vec{k}_0, \vec{x}_r - \vec{x}')] d\vec{x}' = \\ &= \frac{-1}{4\pi v_0^2} \iiint_{-\infty}^{\infty} \frac{\beta_0(\vec{x}') \omega^2}{|\vec{x}_r - \vec{x}'|} \exp\{i[|\vec{k}_0||\vec{x}_r - \vec{x}'| + (\vec{k}_0, \vec{x}')] \} d\vec{x}' = \\ &= \frac{1}{4\pi v_0^2} \iiint_{-\infty}^{\infty} \frac{(i\omega)^2}{|\vec{x}_r - \vec{x}'|} \beta_0(\vec{x}) \exp\left\{i\omega\left[\frac{|\vec{x}_r - \vec{x}'|}{v_0} + \frac{(\vec{n}, \vec{x}')}{{v_0}}\right]\right\} d\vec{x}'. \end{aligned}$$

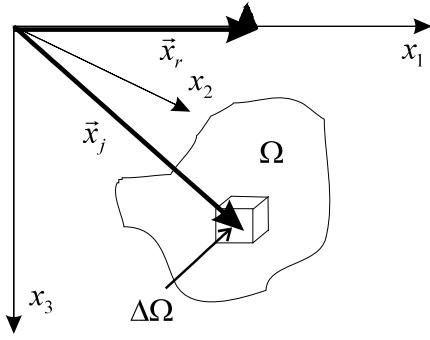


Figure 10.3: The main notations for introducing the Born approximation.

The unknown function $\beta_0(\vec{x})$ can be presented in the expansion on basic functions

$$\beta_0(\vec{x}) = \left(1 - \frac{v_0^2}{v^2(\vec{x})}\right) = \sum_{j=1}^M \alpha_j \varphi_j(\vec{x}),$$

where

$$\varphi_j(\vec{x}) = \begin{cases} 1 & \text{for } \vec{x} \in \Delta\Omega_j, \\ 0 & \text{for } \vec{x} \notin \Delta\Omega_j. \end{cases}$$

Substituting the expansion for $\beta_0(\vec{x})$ into the integral, we shall obtain

$$Q(\vec{x}_r, \omega, \vec{k}_0) = (i\omega)^2 \sum_{j=1}^M \alpha_j A_j \exp[i\omega\tau_j],$$

where

$$\tau_j = \left[\frac{|\vec{x}_r \vec{x}_j|}{v_0} + \frac{(\vec{n}, \vec{x}_j)}{v_0} \right], \quad A_j = \frac{\Delta\Omega_j}{4\pi v_0^2 |\vec{x}_r - \vec{x}_j|}.$$

Using this expression for $Q(\vec{x}_r, \omega, k_0)$ we can rewrite u_{mB} in the form

$$\begin{aligned} u_{mB} &= a_0 h(t) * F_\omega^{-1} \left[\exp[i(\vec{k}_0, \vec{x}_r)] \right] + a_0 h(t) * F_\omega^{-1} \left[(i\omega)^2 \sum_{j=1}^M \exp[i\omega\tau_j] A_j \alpha_j \right] = \\ &= a_0 h(t) * \delta \left(t - \frac{(\vec{n}, \vec{x}_r)}{v_0} \right) + a_0 h(t) * \sum_{j=1}^M \delta^{(2)}(t - \tau_j) \alpha_j A_j = \\ &= a_0 h \left(t - \frac{(\vec{n}, \vec{x}_r)}{v_0} \right) + a_0 \sum_{j=1}^M \alpha_j A_j h''(t - \tau_j). \end{aligned}$$

We can rewrite the model for $\tilde{u}(t, \vec{x}_r, \vec{n})$ taking into account the obtained expression for $u_{mB}(t, \vec{x}_r, \vec{n}, \beta_0(\vec{x}))$:

$$u(t, \vec{x}_r, \vec{n}) = \sum_{j=1}^M \alpha_j C_j(t, \vec{x}_r, \vec{n}) + \varepsilon(t, \vec{x}_r, \vec{n}),$$

where

$$u(t, \vec{x}_r, \vec{n}) = \tilde{u}(t, \vec{x}_r, \vec{n}) - a_0 h(t - (\vec{n}, \vec{x}_r)/v_0), \quad C_j(t, \vec{x}_r, \vec{n}) = A_j h(t - \tau_j).$$

Taking into account that the seismogram is registered in the discrete points x_{ri} ($i = 1, 2, \dots, n$), we shall write the model in the vector-matrix form

$$\vec{u} = C\vec{\alpha} + \vec{\varepsilon},$$

where $\vec{u} = \|u(t, x_{ri}, \vec{n})\|_{i=1}^n$ is the measurement vector and $C = \|C_j(t, x_{ri}, \vec{n})\|_{i=1, j=1}^{n M}$. For the estimation of the parameter vector $\vec{\alpha}$ we can use any statistical criterion, f.e. statistical regularization or singular analysis.

10.3 Diffraction tomography for vector displacement field

10.3.1 Model of measurement

The unknown parameter fields (elastic Lame's parameters $\lambda(\vec{x})$, $\mu(\vec{x})$ and density $\rho(\vec{x})$, or velocities p - and s -waves) are elements of $\theta(\vec{x})$ of the functional space $\Phi(R^3)$

$$\theta(\vec{x}) \in \Phi(\vec{x} \in R^3) \Rightarrow (\lambda(\vec{x}), \mu(\vec{x}), \rho(\vec{x})).$$

The measurement space is an n -dimensional Euclidean space (R^n) is determined by the mapping of the functional space into the measurement space

$$\Phi(\vec{x} \in R^3) \xrightarrow{\mathcal{P}} R^n$$

There, experimental data $\{u(x_i, t_i), i = 1 \div n\} \in R^n$, $x \in R^3$ (3-component seismogram) are noise-including values of the n -functionals of the sought for fields $\theta(\vec{x})$

$$u_i = \mathcal{P}_i(\theta) + \varepsilon_i, \quad \varepsilon_i \in N(0, K_\varepsilon),$$

where $\mathcal{P}_i(\theta)$ is a transition operator from the functional space $\Phi(\vec{x} \in R^3)$ into the measurement space R^n , ε_i is a normal distributed random noise with a covariance matrix K_ε and a mathematical expectation 0. The tomography experiment consist of the registration of the sounding signals, produced by the source f , and have passed through the examined medium. We may assume that the process of propagation is described by the linear operator L_θ

$$\begin{aligned} L_\theta \varphi &= f, \\ L_\theta : L_\theta(\alpha\varphi + \beta\psi) &= \alpha L_\theta \varphi + \beta L_\theta \psi. \end{aligned}$$

The transformation of the sounding signal by the registering channel is described by the integral linear operator of convolution

$$H : H_i \varphi \leftrightarrow \int_T \int_\Omega \int_V d\vec{x} d\tau d\Omega h_i(\vec{n}, \vec{n}', t - \tau) \varphi(\vec{x}, \vec{n}, \tau) \delta(\vec{x} - \vec{x}_i),$$

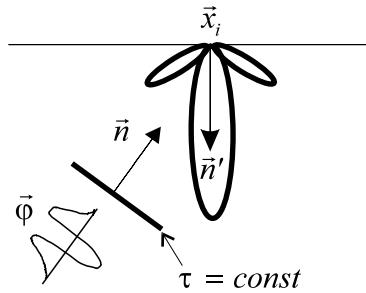


Figure 10.4: An incident field and receiver diagram.

where Ω is a solid angle, \vec{n}' is a main vector of the direction characteristic of receiver (is a vector of the direction of the main maximum of a receiver diagram – see (see Figure 10.4)).

The formal model of the experimental data can be written as

$$u_i = H_i L_\theta^{-1} \vec{f} + \varepsilon_i.$$

In this case the transition operator is

$$\mathcal{P}_i(\theta) = H_i L_\theta^{-1} \vec{f}.$$

It is impossible to find an accurate solution from the measurement data. Therefore, the approximate method is used. We assume that for a certain value of θ_0 we can build the solution

$$\varphi_0 = L_0^{-1} f.$$

We consider that sought for θ is close to θ_0 , i.e.

$$\theta = \theta_0 + \delta\theta,$$

where $\delta\theta \ll \theta_0$.

The formal solution for φ is determined by

$$\varphi = \varphi_0 + L_0^{-1} \delta L_\theta \varphi,$$

where $\delta L_\theta = L_0 - L_\theta$ is a perturbation operator. We can rewrite the model in the form

$$u_i = H_i [\varphi_0 + L_0^{-1} \delta L_\theta \varphi] + \varepsilon_i.$$

The model is nonlinear relative to $\delta\theta$. If the condition

$$\frac{\|H_i L_0^{-1} \delta L_\theta (\varphi - \varphi_0)\|^2}{E(\varepsilon_i^2)} \ll 1$$

is satisfied (the model error is much less than the measurement error), then φ in the model can be replaced by φ_0

$$u_i = H_i [\varphi_0 + L_0^{-1} \delta L_\theta \varphi_0] + \tilde{\varepsilon}_i,$$

where $\tilde{\varepsilon}_i$ includes both the random error ε_i and the error relevant to the determined part of the model. We rewrite the model introducing the scalar product

$$\tilde{u}_i = \langle h_i | L_0^{-1} \delta L_\theta \varphi_0 \rangle_{V,T,\Omega} + \tilde{\varepsilon}_i$$

where

$$\langle \xi | \eta \rangle_{T,V,\Omega} = \int_{\Omega} \int_V \xi(\vec{x}, t, \Omega) * \eta(\vec{x}, t, \Omega) d\vec{x} d\Omega$$

denote the scalar product, $*$ is the convolution sign over time,

$$\tilde{u}_i = u_i - \langle h_i | \varphi_0 \rangle_{T,V,\Omega}$$

is a difference between real data u_i and calculated in the background medium $\theta_0(\vec{x})$ data.

10.3.2 Definition of tomography functionals

We distinguish in the perturbation operator δL the monotonous function $\nu(\delta\theta)$ relative to which the perturbation operator will be linear

$$\begin{aligned} \tilde{u}_i &= \langle L_0^{-1*} h_i | \delta L_\theta \varphi_0 \rangle_{V,T,\Omega} + \tilde{\varepsilon}_i = \\ &= \langle \langle G^* h_i | L'_\nu \varphi_0 \rangle_{T,\Omega} | \nu(\delta\theta) \rangle_V + \tilde{\varepsilon}_i, \end{aligned}$$

where $G = L_0^{-1}$ is the Green operator, G^* is the conjugate Green operator, L'_ν is the Freshe derivative such that

$$L'_\nu : \frac{\delta u}{\delta \nu} = \langle G^* h | L'_\nu \varphi_0 \rangle_{T,\Omega}.$$

The integral kernel of the measurement functional relative to $\nu(\delta\theta)$ takes the name of tomography functional

$$p_i = \langle \varphi_{out} | S | \varphi_{in} \rangle_{T,\Omega},$$

where

$\varphi_{in} = \varphi_0 = Gf$ is the incoming field in a known background medium θ_0 ,

$\varphi_{out} = G^* h_i$ is the inverted outgoing field from the receiver,

$S = L'_\nu$ is the interaction operator of the fields φ_{in} and φ_{out} .

Taking into account the definition of the tomography functionals, we shall write the model as

$$\tilde{u} = P\nu + \tilde{\varepsilon},$$

where $\tilde{u} = \|\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_n\|^T$ is the vector of observed data, $\tilde{\varepsilon} = \|\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \dots, \tilde{\varepsilon}_n\|^T$ is the random error vector,

$$\nu = \left\| \begin{array}{c} |\nu_1\rangle \\ \dots \\ |\nu_M\rangle \end{array} \right\|$$

is the functional vector of unknown parameters,

$$P = \begin{vmatrix} \langle p_{11} | & \dots & \langle p_{1\mu} | & \dots & \langle p_{1M} | \\ \dots & \dots & \dots & \dots & \dots \\ \langle p_{n1} | & \dots & \langle p_{n\mu} | & \dots & \langle p_{nM} | \end{vmatrix}$$

is the tomography functional matrix.

The tomography functional p_i determines all of the elements of the spatial region in the formation of the experimental value u_i

$$u_i = \langle p_i | \nu(\delta\theta) \rangle_V.$$

In traditional ray tomography the tomography functional

$$p_i = \delta(L_i(\vec{x}))$$

is singular one and is localized along the ray connecting the source and the receiver (see Figure 10.5).

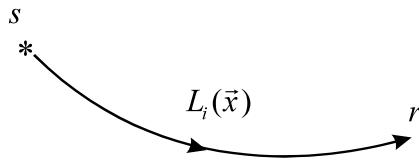


Figure 10.5: In the ray tomography the tomography functional is localized along the ray connecting the source and the receiver.

In the diffraction tomography an every element of a volume has its own weight, which is determined by the interaction of the fields φ_{in} and φ_{out} . The basic content of the tomography experiment is relevant to the overlapping of the tomography functional supports.

10.3.3 Tomography functional for scalar wave equation

The operator of propagation in the background medium:

$$L_0 = -\Delta + \frac{1}{v_0^2(x)} \frac{\partial^2}{\partial t^2}.$$

The operator of propagation in the unknown medium:

$$L_\theta = -\Delta + \frac{1}{v^2(x)} \frac{\partial^2}{\partial t^2},$$

where $\theta_0 = v_0$, $\theta = v(x)$, $x \in R^3$,

$$\nu = \frac{1}{v_0^2} \left(1 - \frac{v_0^2}{v^2(x)} \right)$$

is the unknown parameter function,

$$S = -\frac{\partial^2}{\partial t^2}$$

is the operator of interaction,

$$p = \langle \varphi_{out} | \frac{\partial^2}{\partial t^2} | \varphi_{in} \rangle_T$$

is the tomography functional.

The tomography functional support in the homogeneous background medium is concentrated in the paraboloid layer in the case of a plane incoming wave, and is concentrated in an ellipsoid layer in the case of a spherical incoming wave (under condition of a point-type receiver) (see Figure 10.6).

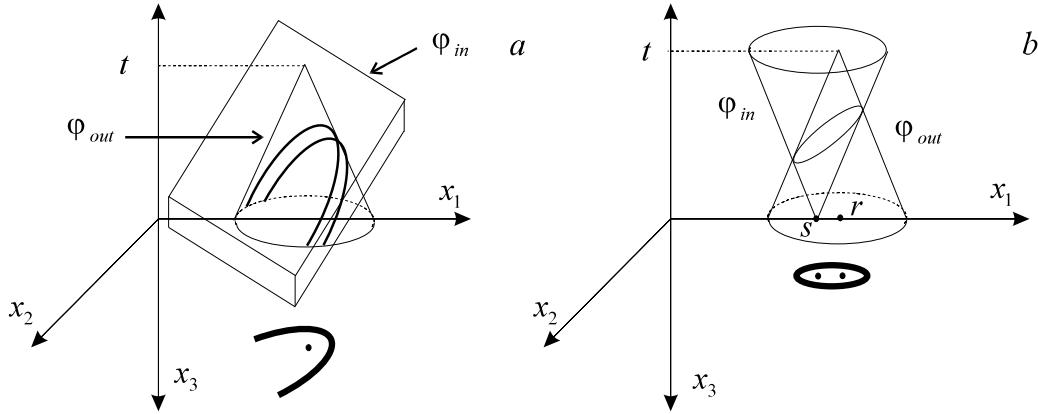


Figure 10.6: Graphic presentation of the interaction between φ_{in} and φ_{out} . a – plane incoming wave; b – spherical incoming wave.

10.3.4 Tomography functional for Lame equation

The operator of propagation in the background elastic medium:

$$\begin{aligned} L_0 \vec{\varphi} &= \rho_0 \frac{\partial^2}{\partial t^2} \vec{\varphi} - [(\lambda_0 + \mu_0) \vec{\nabla} \vec{\nabla} \cdot \vec{\varphi} + \mu \Delta \vec{\varphi} + \vec{\nabla} \lambda_0 \vec{\nabla} \cdot \vec{\varphi} + \\ &+ \vec{\nabla} \mu_0 \times \vec{\nabla} \times \vec{\varphi} + 2(\vec{\nabla} \mu_0 \cdot \vec{\nabla}) \vec{\varphi}]. \end{aligned}$$

The operator of propagation in the unknown elastic medium:

$$\begin{aligned} L \vec{\varphi} &= \rho \frac{\partial^2}{\partial t^2} \vec{\varphi} - [(\lambda + \mu) \vec{\nabla} \vec{\nabla} \cdot \vec{\varphi} + \mu \Delta \vec{\varphi} + \vec{\nabla} \lambda \vec{\nabla} \cdot \vec{\varphi} + \\ &+ \vec{\nabla} \mu \times \vec{\nabla} \times \vec{\varphi} + 2(\vec{\nabla} \mu \cdot \vec{\nabla}) \vec{\varphi}]. \end{aligned}$$

The quantities θ_0 and θ

$$\theta_0 = \begin{vmatrix} |\lambda_0(x)\rangle \\ |\mu_0(x)\rangle \\ |\rho_0(x)\rangle \end{vmatrix} \quad \theta = \begin{vmatrix} |\lambda(x)\rangle \\ |\mu(x)\rangle \\ |\rho(x)\rangle \end{vmatrix}$$

are the parameter vectors in the background medium and the unknown medium accordingly, where $\lambda(x)$ and $\mu(x)$ are the Lame parameters and $\rho(x)$ is density. The tomography functional vector and the interaction operator with block-type structure accordingly look like

$$p = \|\langle p_\lambda | \langle p_\mu | \langle p_\rho | \| \quad \text{and} \quad S = \|S_\lambda| S_\mu| S_\rho\|.$$

The perturbation operator can be presented as

$$\delta L_\theta = L_0 - L = \delta L_\lambda + \delta L_\mu + \delta L_\rho.$$

We shall write the unknown parameter in the form of the perturbation of the background medium

$$\lambda(x) = \lambda_0(x) + \delta\lambda(x), \quad \mu(x) = \mu_0(x) + \delta\mu(x), \quad \rho(x) = \rho_0(x) + \delta\rho(x).$$

The explicit form of the perturbation operator δL_λ :

$$\delta L_\lambda : \delta L_\lambda \vec{\varphi} = \delta\lambda \vec{\nabla} \vec{\nabla} \cdot \vec{\varphi} + \vec{\nabla} \delta\lambda \vec{\nabla} \cdot \vec{\varphi} = \vec{\nabla} (\delta\lambda \vec{\nabla} \cdot \vec{\varphi}),$$

where an equality $\vec{\nabla}(\phi\psi) = \psi \vec{\nabla}\phi + \phi \vec{\nabla}\psi$ is used. The explicit form of the perturbation operator δL_μ :

$$\begin{aligned} \delta L_\mu : \quad \delta L_\mu \vec{\varphi} &= \delta\mu \vec{\nabla} \vec{\nabla} \cdot \vec{\varphi} + \delta\mu \Delta \vec{\varphi} + \vec{\nabla} \delta\mu \times \vec{\nabla} \times \vec{\varphi} + \\ &+ 2(\vec{\nabla} \delta\mu \cdot \vec{\nabla}) \vec{\varphi} = 2\vec{\nabla} \cdot (\delta\mu \vec{\nabla} \vec{\varphi}) + \vec{\nabla} \times (\delta\mu \vec{\nabla} \times \vec{\varphi}). \end{aligned}$$

The last relation has obtained due to relation

$$2\vec{\nabla} \cdot (\delta\mu \vec{\nabla} \vec{\varphi}) = 2(\vec{\nabla} \delta\mu \cdot \vec{\nabla}) \vec{\varphi} + 2\delta\mu \Delta \vec{\varphi},$$

which follows from an identity

$$\vec{\nabla} \cdot (\varphi \vec{F}) = \vec{\nabla} \varphi \cdot \vec{F} + \varphi \vec{\nabla} \cdot \vec{F}$$

and relation

$$\begin{aligned} \vec{\nabla} \times (\delta\mu \underbrace{\vec{\nabla} \times \vec{\varphi}}_{\vec{\nabla} \delta\mu \times \vec{\nabla} \times \vec{\varphi}}) &= \vec{\nabla} \delta\mu \times \vec{\nabla} \times \vec{\varphi} + \delta\mu \vec{\nabla} \times \vec{\nabla} \times \vec{\varphi} = \\ &= \vec{\nabla} \delta\mu \times \vec{\nabla} \times \vec{\varphi} + \delta\mu \vec{\nabla} \vec{\nabla} \cdot \vec{\varphi} - \delta\mu \Delta \vec{\varphi} \end{aligned}$$

which follows from the identities

$$\begin{aligned} \vec{\nabla} \times (\phi \vec{F}) &= \vec{\nabla} \phi \times \vec{F} + \phi \vec{\nabla} \times \vec{F}, \\ \vec{\nabla} \times \vec{\nabla} \times \vec{F} &= \vec{\nabla} \vec{\nabla} \cdot \vec{F} - \Delta \vec{F}. \end{aligned}$$

The explicit form of the perturbation operator δL_ρ :

$$\delta L_\rho : \quad \delta L_\rho \vec{\varphi} = -\delta \rho \partial^2 \vec{\varphi} / \partial t^2.$$

Taking into account the presentation of the tomography functionals, model of measurements data can be written in the form

$$\vec{u}_i = u_i - u_{i0} = \langle p_{i\lambda} | \delta \lambda \rangle_V + \langle p_{i\mu} | \delta \mu \rangle_V + \langle p_{i\rho} | \delta \rho \rangle_V + \varepsilon_i = \langle p_{i\theta} | \delta \theta \rangle_V + \varepsilon_i,$$

where we denote

$$\langle p_{i\theta} | \delta \theta \rangle_V = \int_V \vec{\varphi}_{out} \odot \delta L_\theta \vec{\varphi}_{in} d\vec{x} = \langle \langle \vec{\varphi}_{out} | S_\theta | \vec{\varphi}_{in} \rangle_T | \delta \theta \rangle_V.$$

The sign \odot has the sense of the convolution $*$ and the scalar product (\cdot, \cdot)

$$\vec{\xi} \odot \vec{\eta} \equiv (\vec{\xi}(t) * \vec{\eta}(t)).$$

Taking into account the presentation of the perturbation operators obtained above, we shall receive the explicit form for the interaction operators:

$$\begin{aligned} S_\lambda : \quad & \int_V \vec{\varphi}_{out} \odot \delta L_\lambda \vec{\varphi}_{in} d\vec{x} = \int_V \vec{\varphi}_{out} \odot \vec{\nabla}(\delta \lambda \vec{\nabla} \cdot \vec{\varphi}_{in}) d\vec{x} = \\ & = \int_V \vec{\nabla} \cdot (\vec{\varphi}_{out} * \delta \lambda \vec{\nabla} \cdot \vec{\varphi}_{in}) d\vec{x} - \int_V \vec{\nabla} \cdot \vec{\varphi}_{out} * \delta \lambda \vec{\nabla} \cdot \vec{\varphi}_{in} d\vec{x}, \end{aligned}$$

using the identity

$$\vec{\nabla}(\phi \vec{\xi}) = \vec{\xi} \cdot \vec{\nabla} \phi + \phi \vec{\nabla} \cdot \vec{\xi}.$$

According to Ostrogradskii-Gauss theorem we get

$$\int_V \vec{\nabla} \cdot (\vec{\varphi}_{out} * \delta \lambda \vec{\nabla} \cdot \vec{\varphi}_{in}) d\vec{x} = \int_{\partial V} (d\vec{S} \cdot \vec{\varphi}_{out} * \delta \lambda \vec{\nabla} \cdot \vec{\varphi}_{in}) = 0.$$

Choosing the volume of integration to be large enough, so that it can be considered that $\delta \lambda|_{\partial V} \equiv 0$ and the surface integral vanishes. Finally, we shall obtain the explicit form of the tomography functional p_λ

$$S_\lambda : \quad p_\lambda = \langle \vec{\varphi}_{out} | S_\lambda | \vec{\varphi}_{in} \rangle_T = -\vec{\nabla} \cdot \vec{\varphi}_{out} * \vec{\nabla} \cdot \vec{\varphi}_{in}.$$

The interaction operator S_μ can be written in the general form

$$S_\mu : \quad 2 \int_V \vec{\varphi}_{out} \odot \vec{\nabla} \cdot (\delta \mu \vec{\nabla} \vec{\varphi}_{in}) d\vec{x} + \int_V \vec{\varphi}_{out} \odot \vec{\nabla} \times (\delta \mu \vec{\nabla} \vec{\varphi}_{in}) d\vec{x},$$

using the identities

$$\vec{\nabla} \vec{\varphi} \oplus \vec{\nabla} \vec{\psi} = \sum_{i=1}^3 \sum_{j=1}^3 \nabla_i \varphi_j * \nabla_i \psi_j$$

$$\vec{\varphi} \circledast \vec{\nabla} \vec{\psi} = \sum_{i=1}^3 \varphi_i * \nabla_i \psi_j, \quad \vec{\nabla} \cdot (\vec{\xi} \cdot \mathbf{T}) = \vec{\nabla} \vec{\xi} \cdot \mathbf{T} + \vec{\xi} \cdot \vec{\nabla} \cdot \mathbf{T},$$

we shall obtain for the first integral

$$\begin{aligned} 2 \int_V \vec{\varphi}_{out} \odot \vec{\nabla} \cdot (\delta \mu \vec{\nabla} \vec{\varphi}_{in}) d\vec{x} &= \\ &= 2 \int_V \vec{\nabla} \cdot (\vec{\varphi}_{out} \circledast \delta \mu \vec{\nabla} \vec{\varphi}_{in}) d\vec{x} - 2 \int_V \vec{\nabla} \vec{\varphi}_{out} \oplus \delta \mu \vec{\nabla} \vec{\varphi}_{in} d\vec{x} = \\ &= -2 \int_V \vec{\nabla} \vec{\varphi}_{out} \oplus \vec{\nabla} \vec{\varphi}_{in} \delta \mu d\vec{x}. \end{aligned}$$

Since according to Ostrogradskii-Gauss theorem, the first integral vanishes by the condition $\delta \mu|_{\partial V} = 0$

$$\int_V \vec{\nabla} \cdot (\vec{\varphi}_{out} \circledast \delta \mu \vec{\nabla} \vec{\varphi}_{in}) d\vec{x} = \int_{\partial V} d\vec{S} \cdot (\vec{\varphi}_{out} \circledast \delta \mu \vec{\nabla} \vec{\varphi}_{in}) = 0.$$

The second integral will be written as

$$\begin{aligned} \int_V \vec{\varphi}_{out} \odot \vec{\nabla} \times (\delta \mu \vec{\nabla} \times \vec{\varphi}_{in}) d\vec{x} &= \\ &= - \int_V \vec{\nabla} \cdot (\vec{\varphi}_{out} \otimes (\delta \mu \vec{\nabla} \times \vec{\varphi}_{in})) d\vec{x} + \\ &\quad + \int_V (\delta \mu \vec{\nabla} \times \vec{\varphi}_{in}) \odot (\vec{\nabla} \times \vec{\varphi}_{out}) d\vec{x} = \\ &= \int_V \delta \mu (\vec{\nabla} \times \vec{\varphi}_{out}) \odot (\vec{\nabla} \times \vec{\varphi}_{in}) d\vec{x}, \end{aligned}$$

where it was used identity

$$\vec{\nabla} \cdot (\vec{\xi} \times \vec{\eta}) = \vec{\eta} \cdot \vec{\nabla} \times \vec{\xi} - \vec{\xi} \cdot \vec{\nabla} \times \vec{\eta}, \quad \vec{\eta} = \vec{\varphi}_{out}, \quad \vec{\xi} = (\delta \mu \vec{\nabla} \times \vec{\varphi}_{in})$$

and the notation

$$\otimes : \vec{\xi} \otimes \vec{\eta} = \int \vec{\xi}(\tau) \times \vec{\eta}(t - \tau) d\tau.$$

After an application of the Ostrogradskii-Gauss theorem the first integral vanishes

$$\begin{aligned} \int_V \vec{\nabla} \cdot (\vec{\varphi}_{out} \otimes (\delta \mu \vec{\nabla} \times \vec{\varphi}_{in})) d\vec{x} &= \\ &= \int_{\partial V} d\vec{S} \cdot (\vec{\varphi}_{out} \otimes (\delta \mu \vec{\nabla} \times \vec{\varphi}_{in})) = 0 \end{aligned}$$

due to condition $\delta\mu|_{\partial V} = 0$. Taking into account the obtained presentation for the interaction operator S_μ , we shall write the perturbation operator δL_μ in the form

$$\begin{aligned}\delta L_\mu : & \int_V \vec{\varphi}_{out} \odot \delta L_\mu \vec{\varphi}_{in} d\vec{x} = \\ & = \int_V [(\vec{\nabla} \times \vec{\varphi}_{out}) \odot (\vec{\nabla} \times \vec{\varphi}_{in}) - 2\vec{\nabla} \vec{\varphi}_{out} \oplus \vec{\nabla} \vec{\varphi}_{in}] \delta\mu d\vec{x} = \\ & = \int_V p_\mu(\vec{x}) \delta\mu(\vec{x}) d\vec{x} = \langle \langle \vec{\varphi}_{out} | S_\mu | \vec{\varphi}_{in} \rangle_T | \delta\mu \rangle_V.\end{aligned}$$

Finally, the tomography functional p_μ and interaction operator S_μ will be

$$\boxed{S_\mu : p_\mu = \langle \langle \vec{\varphi}_{out} | S_\mu | \vec{\varphi}_{in} \rangle_T = (\vec{\nabla} \times \vec{\varphi}_{out}) \odot (\vec{\nabla} \times \vec{\varphi}_{in}) - 2\vec{\nabla} \vec{\varphi}_{out} \oplus \vec{\nabla} \vec{\varphi}_{in}}.$$

The action of the operator S_ρ is determined by the expression

$$S_\rho : \int_V \vec{\varphi}_{out} \odot \delta L_\rho \vec{\varphi}_{in} d\vec{x} = - \int_V \vec{\varphi}_{out} \odot \delta\rho \frac{\partial^2}{\partial t^2} \vec{\varphi}_{in} d\vec{x}.$$

Using the integration by parts we shall obtain

$$\begin{aligned}\vec{\varphi}_{out} \odot \frac{\partial^2}{\partial t^2} \vec{\varphi}_{in} &= \int_{-\infty}^{\infty} \vec{\varphi}_{out}(t_i - t) \cdot \frac{\partial^2}{\partial t^2} \vec{\varphi}_{in}(t) dt = \\ &= \vec{\varphi}_{out} \cdot \frac{\partial}{\partial t} \vec{\varphi}_{in} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\partial}{\partial t} \vec{\varphi}_{out}(t_i - t) \cdot \frac{\partial}{\partial t} \vec{\varphi}_{in}(t) dt.\end{aligned}$$

Taking into account the conditions

$$\vec{\varphi}_{out} \Big|_{\infty} = 0, \quad \frac{\partial}{\partial t} \vec{\varphi}_{in} \Big|_{-\infty} = 0,$$

finally, the tomography functional p_ρ and interaction operator S_ρ will be written in the form

$$\boxed{S_\rho : p_\rho = \langle \langle \vec{\varphi}_{out} | S_\rho | \vec{\varphi}_{in} \rangle_T = \frac{\partial}{\partial t} \vec{\varphi}_{out} \odot \frac{\partial}{\partial t} \vec{\varphi}_{in}}.$$

To illustrate the geometry of the tomography functional support, we consider the case of plane incoming wave. The cones of the field $\vec{\varphi}_{out}$, generated by the receiver at the moment t_i and propagating in the reverse time. The outer conical surface corresponds to the propagation with the compressional wave velocity (p -wave), the inner one with the transverse wave velocity (s -wave). The field $\vec{\varphi}_{in}$ is propagated in the positive direction of the x -axis. The tomography functional support is concentrated between the rotation paraboloid, corresponding to the intersection of the outer cone $\vec{\varphi}_{out}$ and the field $\vec{\varphi}_{in}$ and the rotation ellipsoid, corresponding to the intersection of the inner cone $\vec{\varphi}_{out}$ and $\vec{\varphi}_{in}$ (see Figure 10.7). If the incoming field is a transverse wave (s -wave), then the tomography functional support is concentrated in the region between two rotation paraboloids (see Figure 10.8).

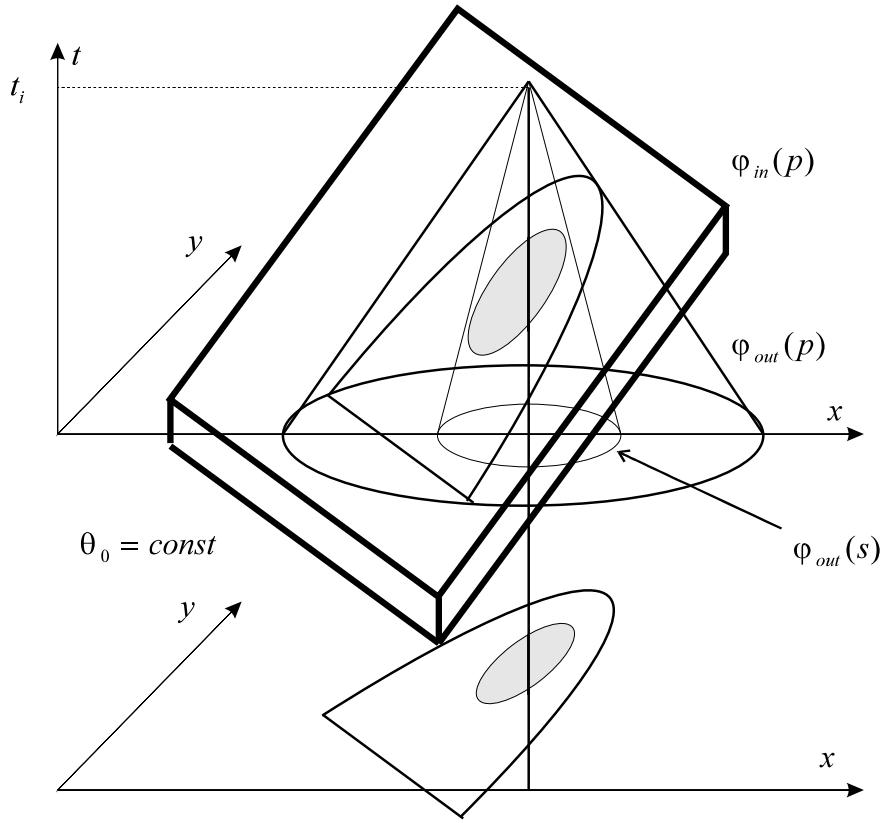


Figure 10.7: Graphic presentation of the interaction between $\vec{\varphi}_{in}$ and $\vec{\varphi}_{out}$. $\vec{\varphi}_{in}$ is p -wave.

10.4 The reconstruction algorithm

To build the reconstruction algorithm, we write down the model using the tomography functional

$$u = P\nu + \varepsilon, \quad u_i = \sum_{\mu=1}^M \langle p_{\mu i} | \nu_{\mu}(\delta\theta) \rangle_V + \varepsilon_i.$$

For an arbitrary linear functional $l(\nu)$, we shall seek for a solution in the form of a linear combination of measurements

$$l(\nu) = \sum_{i=1}^n \alpha_i u_i = (\vec{\alpha} \cdot \vec{u}),$$

where $l(\nu) \equiv \langle l | \nu \rangle_V$ is a linear functional, for example, mean value in the elementary volume (sphere or parallelepiped – see Figure 10.9):

$$\begin{aligned} l^X(\nu) &= \left(\frac{4}{3} \pi r^3 \right)^{-1} H(r - |x - X|), \\ l^X(\nu) &= \frac{1}{abc} H(a/2 - |x_1 - X_1|) H(b/2 - |x_2 - X_2|) H(c/2 - |x_3 - X_3|), \end{aligned}$$

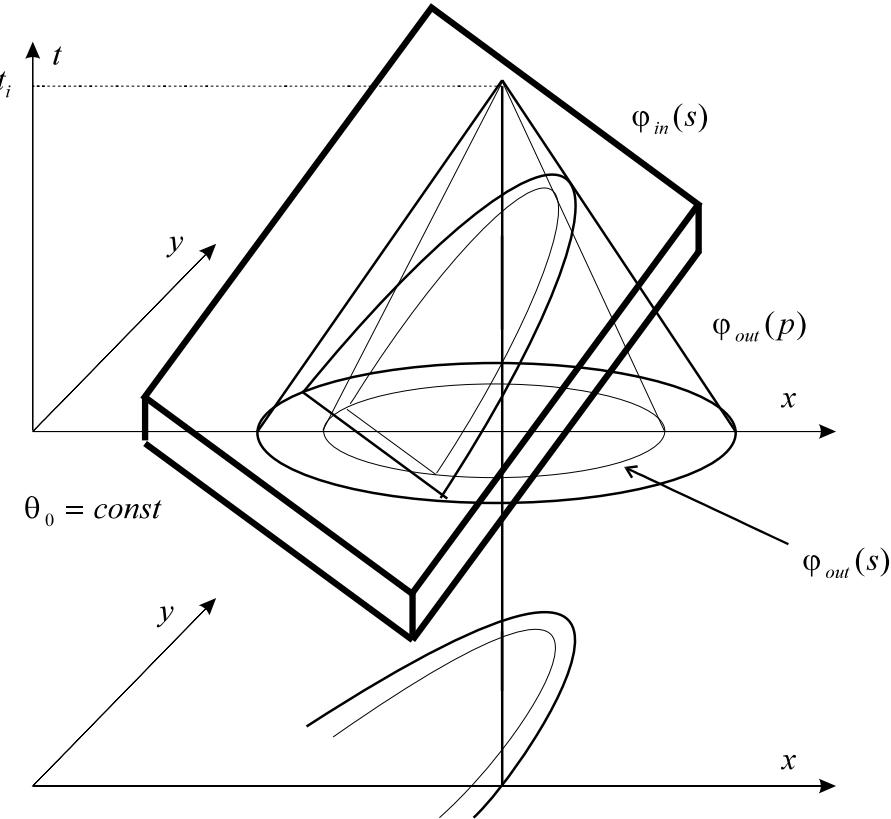


Figure 10.8: Graphic presentation of the interaction $\vec{\varphi}_{in}$ and $\vec{\varphi}_{out}$. $\vec{\varphi}_{in}$ is s -wave.

where

$$H(x) = \begin{cases} 1 & \text{for } x > 0, \\ 0 & \text{for } x < 0. \end{cases}$$

Graphic presentation of the $H(x)$ function – see Figure 10.10. The solution error

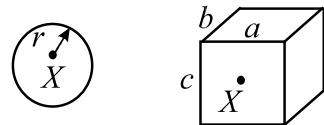


Figure 10.9: The examples of the elementary volumes.

can be written as

$$\begin{aligned} \eta_\mu^X &= l(\nu_\mu) - \sum_{\mu'=1}^M (\vec{\alpha} \cdot \langle \vec{p}_{\mu'} | \nu_{\mu'} \rangle_V) - (\vec{\alpha} \cdot \vec{\varepsilon}) = \\ &= \langle l - (\vec{\alpha} \cdot \vec{p}_\mu) | \nu_\mu \rangle_V - \sum_{\mu' \neq \mu} \langle (\vec{\alpha} \cdot \vec{p}_{\mu'}) | \nu_{\mu'} \rangle_V - (\vec{\alpha} \cdot \vec{\varepsilon}). \end{aligned}$$

We consider to be the solution the function

$$\hat{l}(\nu) = (\vec{\alpha} \cdot \vec{u}),$$

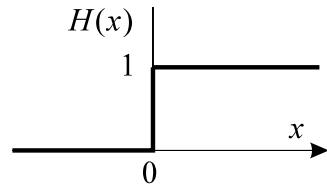


Figure 10.10: Graphic presentation of the $H(x)$ function.

which minimizes the mean square of the retrieval error

$$E(\eta_\mu^X)^2 = (\vec{\Lambda} - Q\vec{\alpha})^* K (\vec{\Lambda} - Q\vec{\alpha}),$$

where $Q = \|P|I\|$ and I is an identity operator of the dimension of the number of measurements

$$\begin{aligned} \alpha^T Q^* &= \|\alpha_1 \dots \alpha_n\| \times \\ &\times \left\| \begin{array}{ccccccccc} \langle p_{11}(x) | & \dots & \langle p_{1\mu}(x) | & \dots & \langle p_{1M}(x) | & 1 & \dots & 0 \\ \dots & \dots \\ \langle p_{i1}(x) | & \dots & \langle p_{i\mu}(x) | & \dots & \langle p_{iM}(x) | & 0 & \dots & 0 \\ \dots & \dots \\ \langle p_{n1}(x) | & \dots & \langle p_{n\mu}(x) | & \dots & \langle p_{nM}(x) | & 0 & \dots & 1 \end{array} \right\|, \\ \Lambda^* &= \left\| \begin{array}{ccccccccc} \langle 0 & \dots & \langle l(x) & \dots & \langle 0 & 0 & \dots & 0 \\ 1 & \dots & \mu & \dots & M & M+1 & \dots & M+n \end{array} \right\|. \end{aligned}$$

The matrix K is the correlation operator

$$K = \left\| \begin{array}{cccc} K_{\nu_1 \nu_1} & \dots & K_{\nu_1 \nu_M} & K_{\nu_1 \varepsilon} \\ \dots & \dots & \dots & \dots \\ K_{\nu_M \nu_1} & \dots & K_{\nu_M \nu_M} & K_{\nu_M \varepsilon} \\ K_{\varepsilon \nu_1} & \dots & K_{\varepsilon \nu_M} & K_{\varepsilon \varepsilon} \end{array} \right\|,$$

$$K : \langle \varphi | K_{\nu_\mu \nu_{\mu'}} | \psi \rangle = E \langle \varphi | \nu_\mu \rangle \langle \nu_{\mu'} | \psi \rangle,$$

$$K_{\varepsilon \varepsilon} = E(\vec{\varepsilon} \cdot \vec{\varepsilon}^T), \quad K_{\varepsilon \nu_\mu} : E(\varepsilon \langle \nu_\mu | \psi \rangle) = K_{\varepsilon \nu_\mu} | \psi \rangle.$$

An optimal estimation of $\vec{\alpha}$ can be presented as

$$\hat{\vec{\alpha}} = (QKQ^*)^{-1} QK\vec{\Lambda}.$$

In a particular case, when statistical relations between various fields ν_μ and $\nu_{\mu'}$ are absent, as well as ν_μ and ε , the optimal estimate will take the form

$$\hat{\vec{\alpha}}_\mu = (P_\mu K_{\nu_\mu \nu_\mu} P_\mu^* + \sum_{\mu' \neq \mu} P_{\mu'} K_{\nu_{\mu'} \nu_{\mu'}} P_{\mu'}^* + K_{\varepsilon \varepsilon})^{-1} \langle p_\mu K_{\nu_\mu \nu_\mu} | l \rangle,$$

where

$$\begin{aligned} P_\mu K_{\nu_\mu \nu_\mu} P_\mu^* &= \left\| \begin{array}{c} \langle p_{\mu 1} | \\ \dots \\ \langle p_{\mu n} | \end{array} \right\| K_{\nu_\mu \nu_\mu} \left\| |p_{\mu 1}\rangle \dots |p_{\mu n}\rangle \right\| = \\ &= \left\| \begin{array}{ccc} \langle p_{\mu 1}(x) | K_{\nu_\mu \nu_\mu}(x, x') | p_{\mu 1}(x') \rangle & \dots & \langle p_{\mu 1}(x) | K_{\nu_\mu \nu_\mu}(x, x') | p_{\mu n}(x') \rangle \\ \dots & \dots & \dots \\ \langle p_{\mu n}(x) | K_{\nu_\mu \nu_\mu}(x, x') | p_{\mu 1}(x') \rangle & \dots & \langle p_{\mu n}(x) | K_{\nu_\mu \nu_\mu}(x, x') | p_{\mu n}(x') \rangle \end{array} \right\| \end{aligned}$$

is the covariance matrix of the signal which is effective with respect to the retrieved μ -th parameter field,

$$\left(\sum_{\mu' \neq \mu} P_{\mu'} K_{\nu_{\mu'} \nu_{\mu'}} P_{\mu'} + K_{\varepsilon \varepsilon} \right)$$

is the covariance matrix of the effective noise.

We find the error variance

$$\begin{aligned} E(\eta_\mu)^2 &= \langle l | K_{\nu_\mu \nu_\mu} - K_{\nu_\mu \nu_\mu} P_\mu^* (P_\mu K_{\nu_\mu \nu_\mu} P_\mu^* + \\ &\quad + \sum_{\mu' \neq \mu} P_{\mu'} K_{\nu_{\mu'} \nu_{\mu'}} P_{\mu'}^* + K_{\varepsilon \varepsilon})^{-1} P_\mu K_{\nu_\mu \nu_\mu} | l \rangle \end{aligned}$$

or, introducing the Fisher operator,

$$E(\eta_\mu)^2 = \langle l | F_\mu^{-1} | l \rangle,$$

where

$$\begin{aligned} F_\mu &= [K_{\nu_\mu \nu_\mu} - K_{\nu_\mu \nu_\mu} P_\mu^* (P_\mu K_{\nu_\mu \nu_\mu} P_\mu^* + \\ &\quad + \sum_{\mu' \neq \mu} P_{\mu'} K_{\nu_{\mu'} \nu_{\mu'}} P_{\mu'}^* + K_{\varepsilon \varepsilon})^{-1} P_\mu K_{\nu_\mu \nu_\mu}]^{-1}. \end{aligned}$$

It follows from the last expression that the quality of the estimation of the μ -th field is the better, the higher is sensitivity of the processed data to variations of the μ -th field and the weaker influence of the variations of the other fields $\mu' \neq \mu$ on the experimental data.

10.5 Analysis of recovery errors and comparison the reconstruction tomography algorithm with Backus-Gilbert method

The presence in the observed data of the random noise leads to the situation, when any procedure of the recovery of the fields $\nu_\mu(\delta\theta)$ is the procedure of the statistical estimation.

We rewrite the error of the estimation for one field $\nu(\delta\theta)$

$$\eta_l = \underbrace{\langle l - (\vec{\alpha} \cdot \vec{p}) | \nu \rangle}_b - \underbrace{(\vec{\alpha} \cdot \vec{\varepsilon})}_n.$$

The error η_l consist of two parts: bias (b) and noise (n).

We write down the optimality criterion in the sense relevant to *BG*-method, in which the absence of a priori information on field of parameters $\nu(\delta\theta)$ is announced.

The first part of the criterion W_{BG} is a “deltaformity” L e.i. the singular functional

$$l_{BG}^x = \delta(x - X)$$

is estimated, and if one denotes

$$\hat{l} \equiv (\vec{\alpha}_x \cdot \vec{p}) = (\vec{\alpha}(X) \cdot \vec{p}(x)),$$

then

$$L = \langle l_{BG} - \hat{l} | K_{BG}^X | l_{BG} - \hat{l} \rangle,$$

where the integral operator K_{BG}^X has a singular kernel of the kind

$$k_{BG}^X(x', x'') = c(x' - X)(x'' - X)\delta(x' - x'').$$

The L -criterion can be written in an explicit form containing the estimation point X :

$$L = (\vec{\alpha}^T S \vec{\alpha}),$$

where

$$\begin{aligned} S_{ij} &= c \int \int dx' dx'' p_i(x') p_j(x'') (x' - X)(x'' - X) \delta(x' - x'') = \\ &= c \int p_i(x) p_j(x) (x - X)^2 dx. \end{aligned}$$

The value of L is relevant to the resolution of the method over the coordinate x .

Substituting instead of the linear combination $(\vec{\alpha} \cdot \vec{p})$ a uniform distribution in the Δ -interval (see Figure 10.11):

$$\begin{aligned} L &= c \int (\vec{\alpha} \cdot \vec{p})^2 (x - X)^2 dx \Rightarrow c \int_{X-\Delta/2}^{X+\Delta/2} \frac{1}{\Delta^2} (x - X)^2 dx = \\ &= \frac{c}{3\Delta^2} (x - X)^3 \Big|_{X-\Delta/2}^{X+\Delta/2} = \frac{\Delta \cdot c}{12}, \\ &c = 12 \Rightarrow L = \Delta. \end{aligned}$$

The value of L is numerically equal to the value of the Δ -interval. Here Δ is interpreted as *resolution length* over the coordinate x .

The second part of the criterion W_{BG} is relevant to the measurement error ε

$$N = (\vec{\alpha}^T K_{\varepsilon\varepsilon} \vec{\alpha}),$$

where $K_{\varepsilon\varepsilon} = E(\varepsilon\varepsilon^T)$ is a covariance matrix of the measurement error.

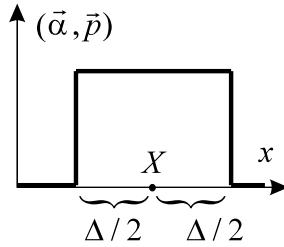


Figure 10.11: Graphic presentation of the window function.

Finally, as the criterion of the optimality of estimation of the functional l^X , it is proposed to minimize the sum of weighted quadratic over α forms of L and N

$$W_{BG} = (1 - \gamma)L + \gamma N \quad (\gamma \in [0, 1])$$

or

$$W_{BG} = (1 - \gamma)\langle l - \hat{l} | K_{BG}^X | l - \hat{l} \rangle + \gamma(\vec{\alpha}^T K_{\varepsilon\varepsilon} \vec{\alpha})$$

under the condition

$$\langle \hat{l} | 1 \rangle = 1 \Rightarrow (\vec{\alpha}^T \cdot \vec{p}) = 1.$$

To solve this problem we shall use the Lagrange multipliers method

$$\begin{aligned} F(\vec{\alpha}) &= (1 - \gamma)\vec{\alpha}^T S \vec{\alpha} + \gamma \vec{\alpha}^T K_{\varepsilon\varepsilon} \vec{\alpha} - 2\lambda(\vec{\alpha}^T \cdot \vec{p} - 1), \\ \frac{\partial F}{\partial \alpha_i} &= 0, \quad i = 1, 2, \dots, n, \\ 2[(1 - \gamma)S + \gamma K_{\varepsilon\varepsilon}] \vec{\alpha} &= 2\lambda \vec{p}. \end{aligned}$$

We introduce the notation

$$B = [(1 - \gamma)S + \gamma K_{\varepsilon\varepsilon}], \quad \hat{\vec{\alpha}} = B^{-1} \vec{p} \lambda.$$

We multiply the estimation $\hat{\vec{\alpha}}$ on the vector \vec{p}^T

$$\vec{p}^T \mid \hat{\vec{\alpha}} = B^{-1} \vec{p} \lambda \Rightarrow \vec{p}^T \hat{\vec{\alpha}} = \vec{p}^T B^{-1} \vec{p} \lambda.$$

Using the condition

$$\vec{p}^T \vec{\alpha} = 1,$$

we shall obtain for λ

$$1 = \vec{p}^T B^{-1} \vec{p} \lambda \Rightarrow \lambda = [\vec{p}^T B^{-1} \vec{p}]^{-1}.$$

Finally we shall write the estimation

$$\hat{\vec{\alpha}} = \frac{B^{-1} \vec{p}}{[\vec{p}^T B^{-1} \vec{p}]}.$$

Let us compare the W_{BG} optimality criterion and the reconstruction tomography criterion W_{RT} , which can be written as

$$W_{RT} = \langle l_{RT}^X - \hat{l} | K_{RT}^X | l_{RT}^X - \hat{l} \rangle + (\vec{\alpha}^T K_{\varepsilon\varepsilon} \vec{\alpha}).$$

The a priori information about $\nu(\delta\theta)$ is given by the operator K_{RT} , which has the sense of the a priori covariance operator in the case of the Gaussian distribution

$$K_{RT}^X = K_{\nu\nu}.$$

It is easy to see that in the case of W_{BG} the first part is a quadratic form of bias by analogy with W_{RT} , the operator K_{BG} should be responsible for “a priori information” in the class of solution. But the operator K_{BG} has a singular kernel, which depends on the estimated point X .

Since the choice of the estimated point is arbitrary, the “a priori” representation in the BG -method is voluntarily transformed along with the change the point X , which is not permissible in the processing of a fixed volume of data.

The linear condition $\langle \hat{l} | 1 \rangle$ used in W_{BG} , is brought about artificially ε due to two factors:

1. The estimate functional is estimated at the point X .
2. The kernel of K_{BG}^X becomes zero at the same point.

The considered above reconstruction tomography algorithm W_{RT} is free from these limitations of the W_{BG} -criterion. The criterion W_{RT} takes full account of both the a priori notions of the elastic fields and the statistical structure of the observed data. At the Figure a kernel K_{BG} is shown. Kernel K_{RT} for the cases of uncorrelated process and for Markov process are shown at Figure b and Figure c respectively. This presentation characterizes the giving of a priori information in the BG -method and RT -method (see Figure 10.12).

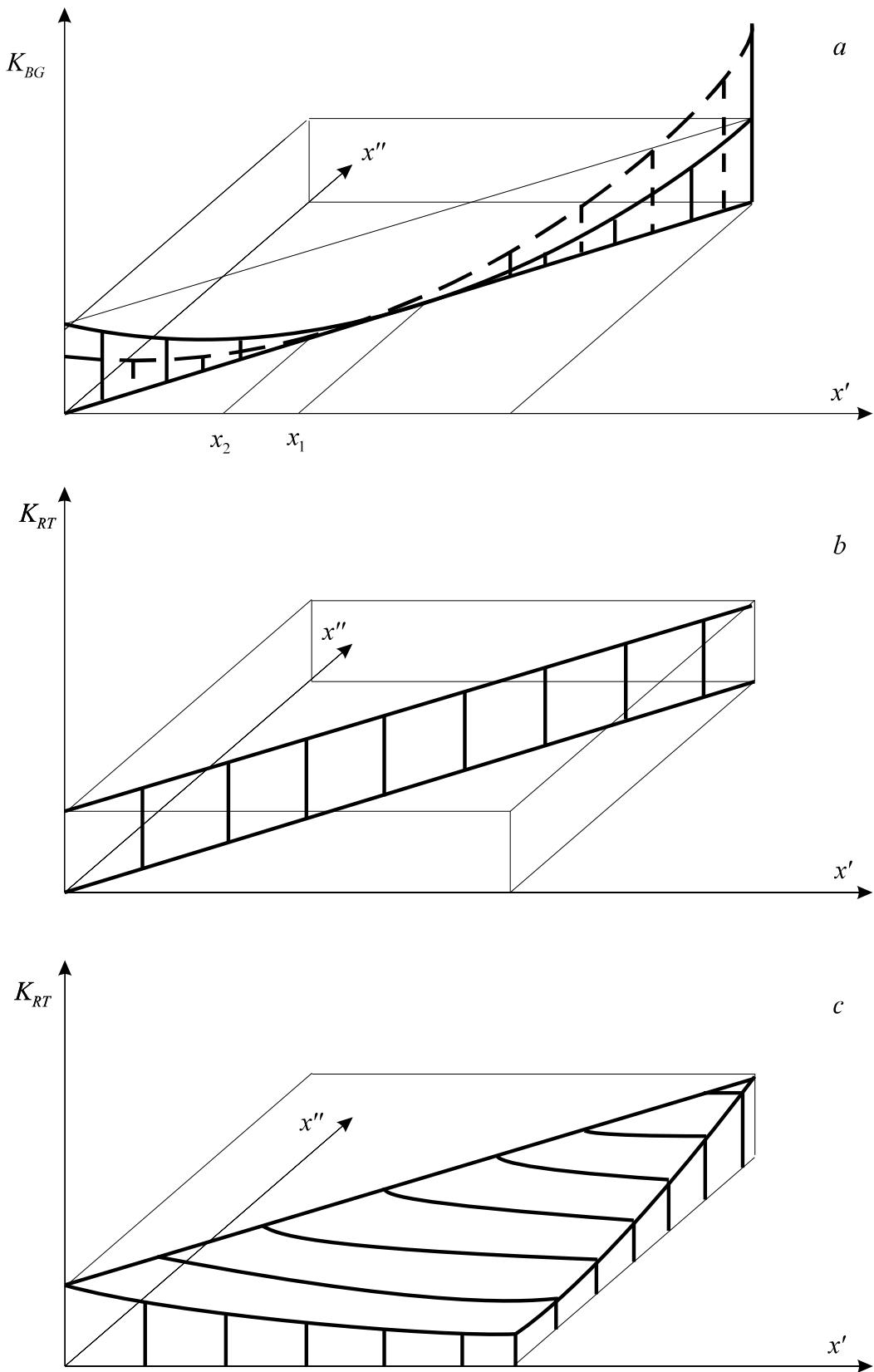


Figure 10.12: The kernels K_{BG} (a) and K_{RT} in the case of uncorrelated process (b) and Markov process (c).

11

Multifactor analysis and processing of time series

We consider a time series

$$\{x_i\}_{i=1}^N \Rightarrow x_i = f((i-1)\Delta t), \quad i = 1, 2, \dots, N.$$

The function $f(t)$ might be random.

I step. Evolvent of 1-D time series to multidimensional one

We choose some number $M < N$, named *Maximum Number of Principal Components* – $MNPC$, and present the first M values of the time series x_i as a first row of the matrix X . As elements of the second row we take values from x_2 to x_{M+1} . The last row with the number $k = N - M + 1$ will be last M elements of the time series: x_k, x_{k+1}, \dots, x_N

$$X = (x_{ij})_{i,j=1}^{k,M} = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_M \\ x_2 & x_3 & x_4 & \dots & x_{M+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & x_{k+2} & \dots & x_N \end{bmatrix}. \quad (11.1)$$

This matrix, with elements $x_{ij} = x_{i+j-1}$, can be considered as M -dimensional sampling with k sample size.

II step. Analysis of principal components: Singular decomposition of the sample correlation matrix

1. Calculation of the mean value and the standard deviation along the columns of the matrix X :

$$\bar{x}_j = \frac{1}{k} \sum_{i=1}^k x_{i+j-1}, \quad s_j = \sqrt{\frac{1}{k} \sum_{i=1}^k (x_{i+j-1} - \bar{x}_j)^2}, \quad (11.2)$$

\bar{x}_j has meaning of the moving average, s_j has meaning of the standard with the rectangular window with the width k .

2. We introduce matrix X^* , that is centered along

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{s_j}, \quad i = 1, \dots, k; \quad j = 1, \dots, M. \quad (11.3)$$

3. Calculation of the sample correlation matrix

$$R = \frac{1}{k} X^{*T} X^* \quad (11.4)$$

with elements

$$r_{ij} = \frac{1}{k} \sum_{l=1}^k \frac{1}{s_i s_j} (x_{i+l-1} - \bar{x}_i)(x_{j+l-1} - \bar{x}_j).$$

4. Calculation of eigenvalues and eigenvectors of the matrix R (*Singular Value Decomposition – SVD*):

$$R = P \Lambda P^T \quad (11.5)$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_M \end{bmatrix}$$

is a diagonal matrix of eigenvalues and

$$P = (p_1, p_2, \dots, p_M) = \begin{bmatrix} p_{11} & p_{21} & \dots & p_{M1} \\ p_{12} & p_{22} & \dots & p_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1M} & p_{2M} & \dots & p_{MM} \end{bmatrix}$$

is an orthogonal matrix of the eigenvectors of the matrix R .

In this case the following relations are realized:

$$P^T = P^{-1}, \quad P^T P = P P^T = I_M,$$

$$\Lambda = P^T R P, \quad \sum_{i=1}^M \lambda_i = M, \quad \prod_{i=1}^M \lambda_i = \det R.$$

Interpretation of the matrices P and Λ . The matrix P might be considered as a transfer function from the matrix X^* to the *Principal Components*

$$X^*P = Y = (y_1, y_2, \dots, y_M).$$

If we consider random time series x_i the eigenvalues of matrix R are sampling variances of the *Principal Components* and square roots of its are sampling standards. The sampling standards are proportionally to the semiaxes of the ellipsoid of the concentration, described by matrix R .

The graphic presentation of eigenvalues is used in the method APC for analysis of the structure of the time series and for selection and the interpretation of the principal components.

In particular, from the eidentity

$$\sum_{j=1}^M \lambda_j = M$$

it follows, that by multiplication all eigenvalues on 100 % it is obtained parts M of variances, which are used by the interpretation of separate principal components.

If the considered time series $\{x_i\}$ is stationary, in the sense that any part of time series with the length $s = \min(M, k)$ carries out all basic information about the structure of the process $\{x_i\}$, it is possible to average the matrix R along diagonals and to obtain autocorrelation function $r(\tau)$:

$$r(\tau) = \frac{1}{M-\tau} \sum_{j=1}^{M-\tau} r_{j,j+\tau}, \quad \tau = 1, \dots, M. \quad (11.6)$$

III step. Selection of the principal components

As it was defined above the principal components is defined as

$$Y = X^*P = (y_1, y_2, \dots, y_M).$$

We normalize the principal components on the sampling standards (by $\lambda_M \neq 0$)

$$Y^* = Y\Lambda^{-1/2} = (y_1^*, y_2^*, \dots, y_M^*).$$

In this case the principal components are orthonormalized:

$$Y^{*T}Y^* = I_M,$$

i.e. we obtain the decomposition of the initial M -dimensional process on orthogonal components.

The transform

$$y_j = X^*p_j$$

is very close to the linear transform of the initial time series by means of discrete convolution operator

$$\begin{aligned} y_j(l) &= \sum_{q=1}^M X_{lq}^* p_{jq} = \sum_{q=1}^M (x_{l+q-1} - \bar{x}_q) \frac{p_{jq}}{s_q} = \\ &= \sum_{q=1}^M x_{l+q-1} \frac{p_{jq}}{s_q} - \sum_{q=1}^M \bar{x}_q \frac{p_{jq}}{s_q}. \end{aligned}$$

Thus, the procedure multifactor analysis creates a set of the linear filters, in this case the eigenvectors are the transition functions of linear filters.

The visual analysis of the eigenvectors and principal components, obtained after linear filtering can give a lot of interesting information about the structure of the process and properties of components of the process.

It is very useful the visual information, which gives a two-dimensional diagram (on the x -axis – p_i or y_i , on the y -axis p_j or y_j).

Example. If we have on the 2-dimensional diagram for p_i and p_{i+1} presentation close to the circle it means that we can interpret these eigenvectors as a pair of $\sin - \cos$ in the Fourier transform. We can estimate the frequency of this function:

$$\begin{cases} x(t) = a(t) \cos(\omega(t)t + \psi(t)), \\ y(t) = a(t) \sin(\omega(t)t + \psi(t)), \end{cases}$$

$a(t)$, $\omega(t)$ and $\psi(t)$ are slowly varied functions

$$\begin{cases} x(t) = r(t) \cos(\alpha(t)t), \\ y(t) = r(t) \sin(\alpha(t)t), \end{cases}$$

where

$$\begin{aligned} r(t) &= \sqrt{x^2(t) + y^2(t)} = a(t), \\ \alpha(t) &= \arctan \frac{y(t)}{x(t)} = \omega(t)t + \psi(t). \end{aligned}$$

We choose two sequences with the interval τ and find a differential of a polar angle:

$$\begin{aligned} \Delta\alpha &= \alpha(t + \tau) - \alpha(t) = \omega(t + \tau)(t + \tau) + \\ &+ \psi(t + \tau) - \omega(t)t - \psi(t) = \\ &= (\omega(t + \tau) - \omega(t))t + \psi(t + \tau) - \psi(t) = \\ &= (\omega'(t)\tau + o(\tau))t + (\omega(t) + \\ &+ \omega'(t)\tau + o(\tau))\tau + \psi'(t)\tau + o(\tau). \end{aligned}$$

If $\omega(t)$ and $\psi(t)$ are slowly varied function such that $\omega'(t)$ and $\psi'(t)$ have an order $O(\tau^2)$, we can write

$$\begin{aligned} \Delta\alpha(t) &= \omega(t)\tau = \frac{2\pi\tau}{T(t)}, \\ \omega(t) &= \frac{\Delta\alpha(t)}{\tau} \quad \text{and} \quad T(t) \approx \frac{2\pi\tau}{\Delta\alpha(t)}. \end{aligned}$$

Thus, calculating the differential of the polar angle, we can estimate an instantaneous frequency $\omega(t)$ and an instantaneous period $T(\tau)$ of a “harmonic oscillation” corresponding to chosen pair of eigenvectors.

It should be noted, it is impossible to estimate the frequency more than

$$\omega_N = \frac{\pi}{\tau}, \quad \hat{\omega}(t) \text{ (should be)} < \omega_N.$$

IV step. Recovery of 1-D time series

Taking into account the orthogonality of the matrix P we have

$$\begin{aligned} X^* &= YP^T = (y_1, y_2, \dots, y_M) \cdot \begin{pmatrix} p_1^T \\ p_2^T \\ \vdots \\ p_M^T \end{pmatrix} = \\ &= \sum_{l=1}^M y_l P_l^T = \sum_{l=1}^M X_l^*. \end{aligned}$$

Denormalization and decentering of the matrix X^*

$$X = \bar{x} I_k^T + X^* S = X_0^* + \sum_{l=1}^M X_l^* S = \sum_{l=0}^M X_l^* S.$$

As a result it is obtained the initial matrix as a sum of $(M + 1)$ matrices.

The transfer to the initial time series might be realized by averaging along the secondary diagonals. We denote by \mathcal{A} the average operator

$$x = \mathcal{A}(X) = \sum_{l=0}^M \mathcal{A}(X_l^* S). \quad (11.7)$$

Thus, it is realized the decomposition of initial time series as a sum of $(M + 1)$ time series.

The important feature of the suggested algorithm is its interactivity, i.e. the dialog principle between a researcher and PC in the process of application.

Selection of parameters

The main control parameter is the MNPC – M (M should be $< N/2$).

The choice of the length M depends on solving problems:

1. *The search of latent periodicity.*

First of all, we calculate eigenvalues by the maximum M and estimate number of eigenvalues $\lambda_i > 0$ (l). On the next step we carry out our analysis using $M = l$.

2. *Smoothing of the time series.*

In this case we can interpret action of the caterpillar-algorithm as filtering. The

reconstruction of some principal component is reduced to the filtering of the time series by the transition function which is equal to the eigenvector. The more M , the more narrower will be the band of filter.

3. *Recovery of the periodicity with the known period.*

M should be equal to the period T and N should be multiply to the period T .

It should be noted that method is very stable to the variation of the length M .

Intermediate results for interpretation

1. Eigenvalues of the correlation matrix of M -dimensional presentation of the time series.
2. Eigenvectors of the correlation matrix. They take up the orthogonal system.
3. Principal components of M -dimensional presentation. They also form the orthogonal system.
4. Reconstructed time series on the different sets of the principal components is operator of transfer from principal components to the M -dimensional matrix X and operator \mathcal{A} for transition from M -dimensional matrix X to 1-dimensional time series $\{x_i\}_{i=1 \div N}$.

It should be pointed out two extreme cases:

- $M \ll N$. We interpret the eigenvectors as transition functions of linear filters, and principal components as results of the action these filters.
- $M \simeq N/2$. The method can be interpreted as an approximation method of an initial time series. For example, the search of harmonic components.

Appendix A

About MATLAB

The first version of MATLAB has appeared in New-Mexico University and Stanford University at the end of 70's. The MATLAB was intended for studying of the theory of matrices, linear algebra and numerical analysis.

At present MATLAB is an effective language for scientific and engineering calculations. It supports mathematical calculations, visualization of the scientific graphics and also programming in language close to mathematical one. MATLAB is an interactive system, the basic object of which is an array, for which it is not required to specify dimensions in an explicit form. It allows to solve many numerical problems, connected with the vector-matrix formulations, reducing time, which would be required for programming in "scalar" languages as C or Fortran. It allows to use MATLAB as a convenient tool for realization of the numerical exercises¹ in the field of the solution of the geophysical inverse problems.

MATLAB Command Window

The Command Window is made active any time when is started MATLAB. It has heading: MATLAB Command Window. This is a basic window for interaction between the user and MATLAB. After activation, in a window there is some help information about the help commands and demonstrations, and also invitation to work ">>", which means, that MATLAB waits for command from the user. The command is entered, if enter is pressed.

Menu bar of the Command Window. Command Window contains the menu bar, which allows to expedite some frequently executed commands of MATLAB. We shall consider in brief a part from commands of the menu, which can be required hereinafter.

Option File. The first option File of the menu includes commands, which will organize work with so-called *M-files*. *M-files* represents a usual text file with extension

¹The following computer exercises are prepared by Dr. Yu.V. Kiselev, (St.Petersburg State University) during of stay at Lausanne University. Authors are thankful to Professor R. Olivier for favourable conditions and friendly spirit.

".m", which contains a sequence of commands of MATLAB language.

Command **New** contains command **M-file**. Command **M-file** allows to create new *M-file*.

Command **Open M-file** bring in on the screen a window of a choice of files. Using this window the user can load *M-file* to the text editor.

The command **Run M-file** shows a window of a choice of files. Using this window the user can run *M-file*.

Option Edit. The second option **Edit** contains commands of editing of the text in the **Command Window**.

Some operations with matrices

The elementary operations. All operations of MATLAB carries out only with one type of objects - rectangular matrices (MATLAB means *matrix laboratory*). The elements of matrices can be the integer, real or complex numbers. To a scalar variable corresponds a matrix 1×1 . The vectors are matrices with one row or column. Definition of variable is not required the instructions about dimensions.

Let we want to multiply two matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 7 & 8 & 1 \\ 9 & 0 & 2 \end{bmatrix}.$$

To create matrices A and B the following commands is entered:

```
A=[1 2; 3 4; 5 0];
B=[7 8 1; 9 0 2];
```

To multiply entered matrices:

```
C=A*B
```

After input of this command there will be the next result on the screen:

```
C=
25    8    5
57   24   11
35   40    5
```

We shall note, that at the end of the first two commands there is the mark ";". It is necessary to finish those commands, the results of which are not required immediately to be displayed on the screen. Now all three matrices are kept in the memory and are accessible to further using.

To transpose a matrix C:

```
D=C'
```

There will be the answer at the screen:

```
D=
25 57 35
 8 24 40
 5 11 5
```

It is possible to obtain eigenvalues and eigenfunctions with the help of function `eig`:

```
v=eig(C)
v=
    57.9031
    0.0000
   -3.9031
```

As a result we have a vector `v`, containing eigenvalues. To calculate the determinant of a matrix it is enough to enter a command

```
d=det(C)
```

and to get the answer

```
d=
 0
```

Division of matrices.

In MATLAB is available two versions of operation of division of matrices. It is a result of non-uniqueness of operation of a matrices multiplication in a general case. The division are designated by symbols "/" (right division) and "\\" (left division). If `D` a nonsingular square matrix, then

```
X=B/D
```

is the solution of the matrix equation

```
X*D=B
```

A statement

```
X=D\B
```

gives the solution of the matrix equation

```
D*X=B
```

If the solution does not exist, i.e. the system of the equations is over-determined, as the answer the best possible solution in sense of a method of the least squares undertakes. If the solution is non-unique, i.e. the system of the linear equations is under-determined, then undertakes the solution in sense of the method of the least squares with a minimum length.

Necessary condition of successful fulfillment of all operations with matrices is compatibility of dimensions of all matrix operands to known rules of linear algebra. In other case the message about an error will be given out.

Updating matrices.

In MATLAB there is the opportunity of addition and removal of rows and columns of a matrix. For an illustration we shall enter a matrix **A** from the previous example

```
A=[1 2;3 4;5 0];
```

and we shall enter a vector **b**

```
b=[7 8];
```

We shall add now vector **b** as a row to a matrix **A**

```
A=[A;b];
```

As a result we obtain

```
A=
1 2
3 4
5 0
7 8
```

We can delete rows or columns of a matrix or "to extract submatrices" from large matrices with the help of a symbol ":" (colon). For example, we shall extract the first three rows from created above matrix *A*. To enter a command

```
A=(1: 3, :)
```

and thus we return back to a matrix **A**

```
A=
1 2
3 4
5 0
```

It is possible to choose one element of a matrix

```
A(3,1)
```

and to get the answer

```
ans=
```

```
5
```

Variable **ans** is generated automatically, if expression without a symbol of equality is entered. MATLAB allows complex expressions and brackets. Thus here are the rules of priorities, similar those in the known programming language Fortran.

Symbol ":"

Above we saw, as with the help of a colon ":" it is possible to extract the rows from a matrix. In MATLAB a symbol ":" is the most used at creation of vectors and matrices. It also represents a convenient way of the address to elements of matrices and manipulations with them by means of indexation, i.e. implementation of various operations with indexes of elements. We shall consider these two operations in brief.

Creation of vectors. The colon is very conveniently to use for creation of vectors with the elements, which change under the laws of the arithmetic progression. For example, operator

```
x=1:6;
```

create a row vector

```
x=
    1 2 3 4 5 6
```

The elements of a created vector x change with a step + 1 by default. To set a step, different from a unit, it should be inserted between initial and final elements. For example, by entering a row

```
y=6:-1:1;
```

we get

```
y=
    6 5 4 3 2 1
```

We note, that a step it is possible to set not only constant, but also expression. We shall create, for example, vector with a step of $\pi/4$

```
z=0:pi/4:pi;
```

and as a result we shall get

```
z=
    0.000 0.7854 1.5708 2.3562 3.1416
```

Similarly it is possible to set a vector-column, by conjugation. For example, the operator

```
y=(0.0:0.5:2.5)'
```

creates a vector-column

```
A=
    0.0000
    0.5000
    1.0000
    1.5000
    2.0000
    2.5000
```

Indexation. One of the ways of the manipulation with an element of a matrix is the manipulation with its indexes in round brackets. For example, if we enter a matrix

```
A=
 1  2  8
 3  4  9
 5  6  1
 7  2  2
```

then to reach, for example, the second element of the first row and to implement with it any operations it is possible as

```
c=A(1,2)+0.5
c=
 2.5000
```

Index in MATLAB can be not only number, but also vector and this vector-index is very conveniently to set with the help of a colon “:”. We shall consider the following command

```
c=A(1:3,2)
c=
 2
 4
 6
```

Now the first index represents a vector of three elements: [1 2 3] and result of implementation of the command is a vector *c* with the elements [A (1,2) A (2,2) A (3,2)], i.e. vector, equal to three elements of the second column of the matrix *A*. Similarly it is possible to extract a square matrix 2x2 from a matrix *A*:

```
c=A(3:4,2:3)
c=
 6  1
 2  2
```

If instead of an index only one colon “:” be situated, it means, that all appropriate elements of a row or column are returned

```
d=A(:,4)
d=
 8
 9
 1
 2
```

The vector *d* is the fourth column of a matrix *A*. The operator *A(:, :)*, hence, returns the full matrix *A*.

The conditional operators and loops.

MATLAB has the operators of managing logic: the loops FOR and WHILE, operator of conditional statement IF and switch SWITCH. The scope of each operator of managing logic should be limited by a word `end`. MATLAB allows embedded loops and opportunity of a preschedule exit from them with the help of a command `break`. Example of the operator FOR:

```
n=10;
for i=1:n
    x(i)=2*i-1;
end
```

Example of the operator WHILE

```
n=10;
i=1;
while i <= n
    x(i)=2*i-1;
    i=i+1;
end
```

We shall consider an example of the operator SWITCH. Let variable `num` can have different values, including values: -1,0,1. In dependence on value of variable `num` we want to print on the screen "minus one", "zero", "plus one" respectively or, if the value of variable `num` is not equal -1,0,1, is printed "another value". An example of the operator "switch" is given below:

```
switch num
    case - 1
        disp ("minus one")
    case 0
        disp ("zero")
    case + 1
        disp ("plus one")
    otherwise
        disp ("another value")
end
```

Arithmetic element by element operations exist in MATLAB, These operations can be used instead of some loop operations. For example, element-wise multiplication, element-wise division, element-wise rising to a power. We shall consider examples of such operations. Let two vectors are given

```
a= [2 1 2];
b= [1 2 1];
```

Element-wise division:

```
c=a./b
c=
2 0.5 2
```

Element-wise multiplication:

```
c=a.*b
c=
2 2 2
```

Element-wise rising to a power:

```
C= (a.*b) .^ 2
c=
4 4 4
```

The functions and M-files

MATLAB contains a wide set of functions: elementary mathematical functions, special mathematical functions, functions of matrices manipulations and etc.. Some of functions are “built in MATLAB” and their texts are not accessible to the users. The other functions are accessible and make of a library of so-called *M-files*. *M-file* represents a usual text file, containing a set of the operators in the language MATLAB.

Functional M-files. The name of *M-file*, containing function should coincide with a name of this function with addition of extension ”.m”. The first line of a *M-file* should contain a key word ”function”, which specifies, that it is a functional *M-file*. Then there is the body of function, i.e. program, realizing algorithm, written in the MATLAB language.

We shall consider for an example of one of the base functions MATLAB with a name `mean`, calculating a mean value. The function is contained in a file `mean.m`.

```
function y = mean(x)
% For vectors, MEAN(X) is the mean value of the elements in X.
% For matrices, MEAN(X) is a row vector containing the mean value
% of each column.
%
[m,n] = size(x);
if m == 1
    m = n;
end
y = sum(x) / m;
```

We shall consider each line of this text. The first line after a key word `function` contains the announcement of function: its name, list of input arguments in round brackets (in our case one argument `x`) and the list of target arguments before the sign of equality (in our case one argument `y`). If the function returns a few arguments, their lists are located in the square brackets and the arguments are separated by commas, for example:

```
function [mean, stdev] =stat (x)
```

The following four lines begin with a symbol of the line comment.

In the sixth line are entered two variable `m` and `n`, the values of which are returned by function `size` of calculation of a size of a matrix. As the function `size` returns two values, variables in the left part of assignment statement are placed into square brackets.

The conditional statement `if` processing the case when `x` is $1 \times n$ matrix.

Calculation of a value of returned argument `y` finishes the body of function

Variables `m`, `n` and `y` are local variables for function `mean`. It means, that their scope is limited by the body of function and they do not exist after its completion.

Script-file The *M-file* can consist of a fragment of the program, i.e. from some sequence of the operators of MATLAB language, not being taken as a function. Such files have a name: script-file. They should not contain a word `function`.

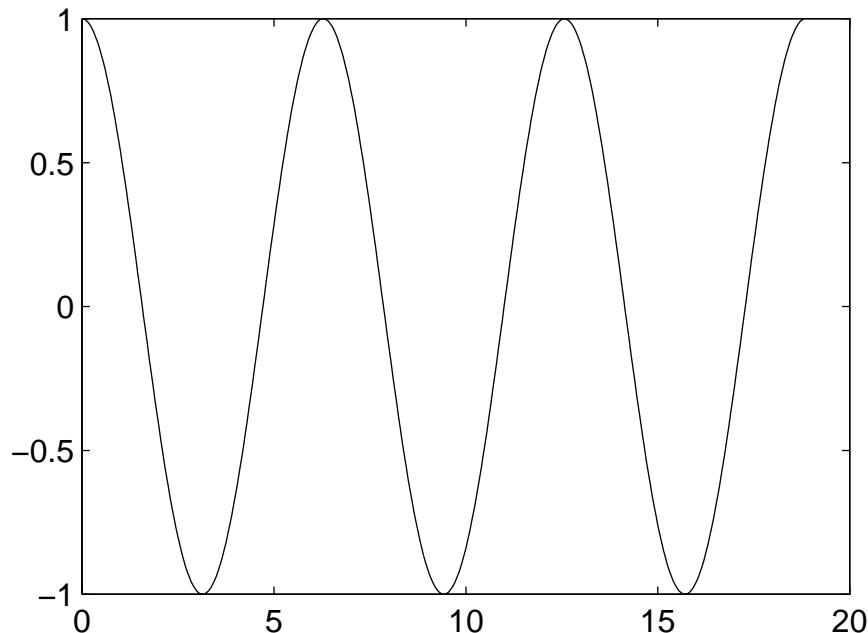
Creation of the elementary graphics

We shall illustrate some elementary graphic opportunities of MATLAB language on several simple examples.

If it is required to construct the diagram of any simple function, for example $y=\cos(x)$, it can be made by writing the following sequence of the operators:

```
x=0:0.05:6 * pi;
y=cos (x);
plot (x, y)
```

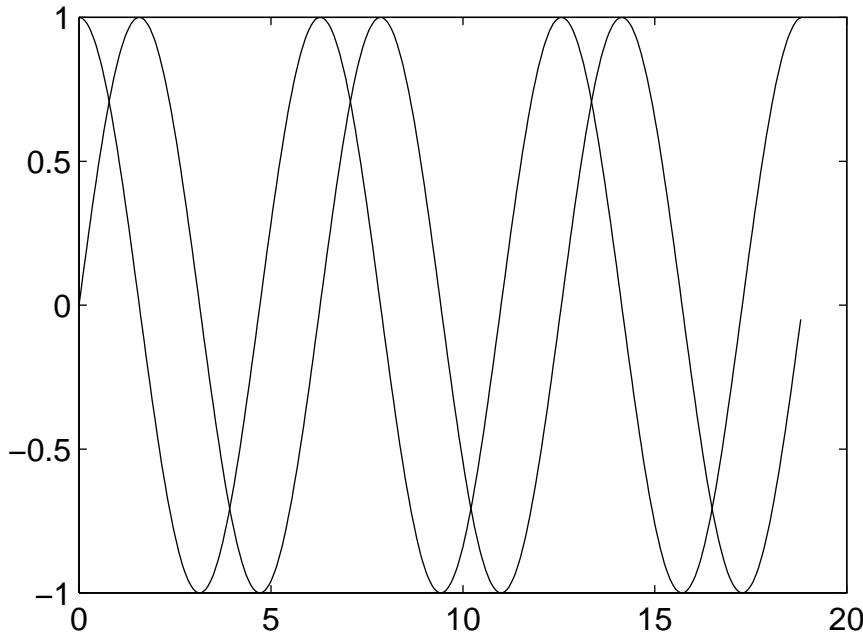
As the result at the screen we shall see the next figure.



The figure can be constructed by the several curves. In this case each dependence should be set by a pair of input arguments of function `plot`:

```
t=0:0.01:6 * pi;
y1=cos (t);
y2=sin (t);
plot (t, y1, t, y2)
```

As a result we have the next diagram of two functions.



MATLAB allows to set a colour and type of curves on the diagram, which differs from established by default. For example a command

```
plot (x, y, " * r ")
```

draws a curve of y as a function of x by symbols $*$ of a red colour.

For an output of the diagrams in logarithmic and half-logarithmic scale the following functions are available:

`loglog (x, y)` is diagram with logarithmic scale on both axes;

`semilogx (x, y)` is diagram with a logarithmic axis x and linear y ;

`semilogy (x, y)` is diagram with a logarithmic axis y and linear x .

MATLAB gives an opportunity of division of a graphic window by the parts, for a simultaneous output of the several diagrams. The operator

```
subplot (m, n, p)
```

breaks a window by a grid $m \times n$ and uses p -th part of a window ($p = 1, \dots, m \times n$) for the graphic output.

To get the additional information about language and opportunities of MATLAB one may type at the command line, for example: `intro`, `demo`, `help arith`, `help slash`, `help punct`, `help elmat`, `help elfun`, `help lang`.

Appendix B

Statistical methods and transforms

B.1 Examples of numerical simulation of random values

There are numerical procedures for simulation of random values which can be realized by computers. The next script produces the *samples*¹ belonging to uniform (at the interval $(0, 1)$) and normal ($N(0, 1)$) distributions.

```
%Script r0a.m
%Examples of random values
%indd=1 : uniform distribution at interval (0,1)
%indd=2 : Gaussian distribution (N(0,1))
clear
% INPUT
indd=1;
m=25;      % sample size
% END INPUT
n=1:m;      % n=1,2,...,m
if indd==1
    s=rand(1,m);      % s : row vector
else
    s=randn(length(n),1);      % s : column vector; length(n)=m
    s=s';      % s : row vector
end
rect=[0.1 0.1 .8 0.5];      % size of figure
axes('position',rect)
stem(n,s);      % graphic presentation
```

¹The word *sample* is often used in the case that the results of measurements are independently and identically distributed with some distribution function. Sometimes the numbers produced by the experiment is called *a batch of numbers*, but usually we shall use a notion *sample* for both these cases.

B.1.1 Exercises

1. With the use of script `r0a.m` to implement graphic presentation of samplings belonging to the *uniform* and *normal* distributions.
2. By using the command line to implement graphic presentation of samplings belonging to the *uniform* and *normal* distributions with the use of `plot` command.

B.2 Creation of histogram

The histogram characterizes the probability properties of a random variable. The next script creates the histogram for a sampling belonging to the normal distribution ($N(0, 1)$).

```
%Script p0hist.m
% creation of histogram
% nr : sample size
clear
% INPUT
nr=500;      %sample size
del=0.5;      %input of bin at x-coordinate
% END INPUT
x=-5:del:5; % creation of bins at x-coordinate
d=randn(1,nr);           % d : a batch of numbers (row vector)
%hist(d,x);
[n,x]=hist(d,x);
bar(x,n./(nr.*del)); %presentation of the normalized histogram
hold on
d2=d_gauss(x,0,1); %Gaussian density function (row vector)
plot(x,d2,'g') %graph presentation of Gaussian density function
```

B.2.1 Exercises

1. To compare the histogram from Gaussian distribution for the different sample sizes (f.e. 100 – 500).
2. To compare the histogram from Gaussian distribution for different number of the bins (f.e. 4 – 20).

B.3 Description of a random variable

The complete description of random variable is reached by the use of *probability function*, *density function*, *(cumulative) distribution function* (see 1.2). The scripts for calculation the density functions, probability functions and the generators for simulation of a sampling belonging to these distributions are presented below.

B.3.1 Beta distribution

Density function

$$f_{\xi}(x|\alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{if } 0 < x < 1, \\ 0 & \text{if } x \leq 0 \text{ or } x \geq 1 \end{cases} \quad (\text{B.1})$$

where $\alpha > 0, \beta > 0$.

$$M\xi = \frac{\alpha}{(\alpha + \beta)}, \quad D\xi = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Graphic presentation of the *beta* density function – script `d0beta.m`. Function `d=d_beta(x,al,bt)` returns row vector `d` – values of the *beta* density function.

B.3.2 Binomial distribution

See 1.6.1. Graphic presentation of the *binomial* probability function – script `d0bin.m`. Function `d=d_bin(n,p)` returns row vector `d` – values of the *binomial* density function.

```
% Script d0bin.m
% density function for binomial distribution
% call function d=d_bin(n,p)
% n : n=1,2,...; - maximum value of argument x
%      x=0,1,...,n
% p : 0 < p < 1
% d : row vector, returns the probability function
%
clear
%INPUT
n=50;
p=0.7;
%END INPUT
d=d_bin(n,p);
x=0:n;
% graphic presentation
rect=[0.1 0.1 .8 0.5]; % size of figure
axes('position',rect)
stem(x,d);
comp_prob=sum(d) %must be equal 1
math_exp= n .* p %mathematical expectation
stand_dev=sqrt(n .* p .* (1-p)) %standard deviation
m1=sum(x.*d) %mathematical expectation (by definition)
s1=sqrt(sum((x-m1).^2.*d)) %standard deviation (by definition)
pause
hold on
dn=d_gauss(x,math_exp,stand_dev);
plot(x,dn,'g');
```

Function `d_bin(n,p)`.

```

function d=d_bin(n,p);
%Binomial distribution d=d_bin(n,p)
% n : n=1,2,...; - maximum value of argument x
%      x=0,1,...,n
% p : 0 < p < 1
%
n1=prod(1:n);
d=zeros(1,n+1);      % preparation of array d
for m=0:n
    if m==0
        m1=1;
    else
        m1=prod(1:m);
    end
    nm=n-m;
    if nm==0
        nm1=1;
    else
        nm1=prod(1:nm);
    end
    d(m+1)=(n1/(m1*nm1))*p^ m * (1-p)^ nm;
end

```

B.3.3 Cauchy distribution

See 1.7.9. Graphic presentation of the *Cauchy* density function – script `d0cauch.m`. Function `d=d_cauch(x)` returns row vector `d` – values of the *Cauchy* density function.

B.3.4 Chi-square distribution

See 1.7.4. Graphic presentation of the *chi-square* density function – script `d0chisq.m`. Function `d=d_chisq(x,n)` returns row vector `d` – values of the *chi-square* density function.

```

%Script d0chisq.m
%density function for chi-square distribution
% call d=d_chisq(x,n)
% n : number of degree of freedom
% x : random variable, a < x < b
% d : row vector, returns the density function
clear
%INPUT
n=50
% determination of interval at x axis

```

```

a=0;      % (a < x < b)
b=100.0;
nx=500;    %number of samples
%END INPUT
del=(b-a)/(nx-1);    % determination of a bin
x=a:del:b;
d=d_chisq(x,n);
plot(x,d);    % graphic presentation
comp_prob=trapz(x,d)    %must be equal 1
math_exp=n    %mathematical expectation
stand_dev=sqrt(2 .* n)    %standard deviation
m1=trapz(x,x.*d)    %mathematical expectation (by definition)
s1=sqrt(trapz(x,(x-m1).^ 2.*d))    %standard deviation
%           (by definition)
pause
% comparison with Gaussian distribution
hold on
dn=d_gauss(x,math_exp,stand_dev);
plot(x,dn,'g');

```

Function `d=d_chisq(x,n)`

```

function d=d_chisq(x,n)
%Chi-square distribution d=d_chisq(x,n)
% n : number of degree of freedom
% x : 0 < x < inf
% d : row vector, returns the density function
%
d=((x./2).^( (n./2)-1).*exp(-x ./ 2))/(2.*gamma(n./2));

```

B.3.5 Exponential distribution

See 1.7.7. Graphic presentation of the *exponential* density function – script `d0expon.m`. Function `d=d_expon(x,1)` returns row vector `d` – values of the *exponential* density function.

B.3.6 Fisher distribution

See 1.7.6. Graphic presentation of the *Fisher* density function – script `d0fish.m`. Function `d=d_fish(x,n1,n2)` returns row vector `d` – values of the *Fisher* density function.

B.3.7 Gamma distribution

Density function

$$f_\xi(x|\alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0 \end{cases} \quad (\text{B.2})$$

where $\alpha > 0, \beta > 0$.

$$M\xi = \frac{\alpha}{\beta}, \quad D\xi = \frac{\alpha}{\beta^2}.$$

Graphic presentation of the *gamma* density function – script `d0gam.m`. Function `d=d_gam(x,al,bt)` returns row vector `d` – values of the *gamma* density function.

B.3.8 Gaussian distribution

See 1.7.1. Graphic presentation of the *Gaussian* density function – script `d0gauss.m`. Function `d=d_gauss(x,xm,s)` returns row vector `d` – values of the *Gaussian* density function.

B.3.9 Geometrical distribution

See 1.6.3. Graphic presentation of the *geometrical* probability function – script `d0geom.m`. Function `d=d_geom(n,p)` returns row vector `d` – values of the *geometrical* probability function.

B.3.10 Hypergeometrical distribution

Probability function

$$f_\xi(x|n, a, b) = \begin{cases} \frac{C_x^a C_{n-x}^b}{C_n^{a+b}} & x = 0, 1, 2, \dots, n \\ 0 & x \neq 0, 1, 2, \dots, n \end{cases} \quad (\text{B.3})$$

where $a = 0, 1, 2, \dots, b = 0, 1, 2, \dots, n = 0, 1, 2, \dots, a + b \geq n$.

$$M\xi = \frac{nab}{(a+b)^2}, \quad D\xi = \frac{a+b-n}{a+b-1}.$$

Graphic presentation of the *hypergeometrical* probability function – script `d0hypg.m`. Function `d=d_hypg(n,a,b)` returns row vector `d` – values of the *hypergeometrical* probability function.

B.3.11 Laplace distribution

See 1.7.8. Graphic presentation of the *Laplace* density function – script `d0lapl.m`. Function `d=d_lapl(x,l)` returns row vector `d` – values of the *Laplace* density function.

B.3.12 Logarithmic normal distribution

See 1.7.10. Graphic presentation of the *logarithmic normal* density function – script `d0lnorm.m`. Function `d=d_lnorm(x,xm,s)` returns row vector `d` – values of the *logarithmic normal* density function.

B.3.13 Poisson distribution

See 1.6.2. Graphic presentation of the *Poisson* probability function – script `d0poiss.m`. Function `d=d_poiss(n,a)` returns row vector `d` – values of the *Poisson* probability function.

B.3.14 Snedecor distribution

Density function

$$f_{\xi}(x|n, m) = \frac{\Gamma((n+m)/2)(n/m)^{n/2}}{\Gamma(n/2)\Gamma(m/2)} x^{(n/2)-1} \left(1 + \frac{nx}{m}\right)^{-(n+m)/2}, \quad x > 0, \quad (\text{B.4})$$

where $n > 0, m > 0$.

$$M\xi = \frac{m}{m-2}, \quad D\xi = \frac{2m^2(n+m-2)}{n(m-4)(m-2)^2}, \quad m > 4$$

Graphic presentation of the *Snedecor* density function – script `d0sned.m`. Function `d=d_sned(x,n1,n2)` returns row vector `d` – values of the *Snedecor* density function.

B.3.15 Student distribution

See 1.7.5. Graphic presentation of the *Student* density function – script `d0stu.m`. Function `d=d_stu(x,n)` returns row vector `d` – values of the *Student* density function.

B.3.16 Uniform distribution

See 1.7.3. Graphic presentation of the *uniform* density function – script `d0unif.m`. Function `d=d_unif(x,al,bt)` returns row vector `d` – values of the *uniform* density function.

B.3.17 Two dimensional normal distribution

See 1.7.2. Script `d0gauss2.m` returns density function for 2-d normal distribution.

B.3.18 Exercises

1. To compare the binomial probability function and Gaussian density function for different values of the parameters n and p .

2. To compare the Poisson probability function and Gaussian density function for different values of the parameter a .
3. To compare the chi-square density function and Gaussian density function for different values of the parameter n .
4. To compare the Student density function and Gaussian density function for different values of the parameter n .
5. Using script `d0gauss2.m` to show a graphic presentation of the 2-d normal density function for different values of the mean values (zero and non zero) and covariance matrix.

B.4 Computer simulation of the random values

There are the simple algorithms for a numerical modeling of the random values which belong to the discrete and continuous type with known probability function and the density function correspondingly. For the both cases we need a generator for a *uniform random distribution* on interval $[0, 1]$ with the following density function

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1, \\ 0, & x < 0 \text{ or } x > 0. \end{cases} \quad (\text{B.5})$$

For the discrete case the general method for the modeling of the discrete random variable ξ with the probability function $p_m = P(\xi = x_m)$ is based on the next relation

$$P\left(\sum_{n=0}^{m-1} p_n \leq \alpha < \sum_{n=0}^m p_n\right) = p_m, \quad (\text{B.6})$$

where a value α belongs to uniform distribution from (B.5).

For the continuous case the formula for modeling of random variable with cumulative distribution function $F(x)$ can be written as

$$\xi = F^{-1}(\alpha), \quad (\text{B.7})$$

where $F^{-1}(\alpha)$ is an inverse (*cumulative distribution function*), α belongs to uniform distribution from (B.5) and a value ξ belongs to distribution with cumulative distribution function $F(x)$. Because $F(x)$ is a strongly monotone continuous function, we can write the next equalities

$$P(F^{-1}(\alpha) < x) = P(\alpha < F(x)) = F(x). \quad (\text{B.8})$$

The scripts `rnd0alc.m` and `rnd0ald.m` return random values correspondingly for continuous and discrete distributions which are introduced in B.3.

The scripts `rnd0gau.m` and `rnd0uni.m` return random values belonging to Gaussian and normal distributions correspondingly. Script `rnd0n2.m` returns a sampling belonging to the 2-D Gaussian distribution.

B.4.1 Kolmogorov criterion

Script `gc0kolm.m` gives an opportunity to realize the Kolmogorov criterion of belonging of sample cumulative distribution function to one or other type of continuous distribution from B.3.

The Kolmogorov criterion gives probability of the random event, that a difference between cumulative distribution function $F_\xi(x)$ and its sample image $F_N^*(x)$ is smaller than given quantity γ . In explicit form this criterion reads as

$$\lim_{N \rightarrow \infty} P \left(\sup_{-\infty < x < \infty} |F_\xi(x) - F_N^*(x)| < \frac{\gamma}{\sqrt{N}} \right) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2\gamma^2), \quad (\text{B.9})$$

where N is sample size.

B.4.2 Exercises

1. Using scripts `rnd0alc.m` and `rnd0ald.m` from B.4 to compare mean values and standard deviation with their population values for different sample sizes (f.e. 10 – 500) for the cases of chi-square and Student distributions.
2. Using scripts `rnd0gau.m` and `rnd0uni.m` to investigate an influence of the sample size (f.e. 10 – 500) on the sample mean value and sample standard deviation.
3. Using script `rnd0n2.m` to investigate an influence of the sample size (f.e. 100 – 5000) on the sample mean and sample covariance matrix (correlation and covariance coefficients).
4. With the use of script `gc0kolm.m` to calculate the probability for the Kolmogorov criterion for different types of the continuous distributions (Gaussian, χ^2 , Student, from B.3). To pay the main attention to a behavior of the sample (cumulative) distribution function in dependence of sample size (f.e. 10 – 1000).

B.5 Confidence intervals

See 1.7.12. Script `g0conf.m` returns the value of confidence interval for the sample mean of the random value distributed with Gaussian distribution. Graphic output of script `g0conf.m` is presented at Figure B.1

B.5.1 Exercises

1. Using script `g0conf.m` to find the confidence intervals for the sample mean of the random values distributed with the Gaussian distribution for different values of the confidence probability β (f.e. $\beta=0.2, 0.5, 0.9$) and sample size (10 – 100).

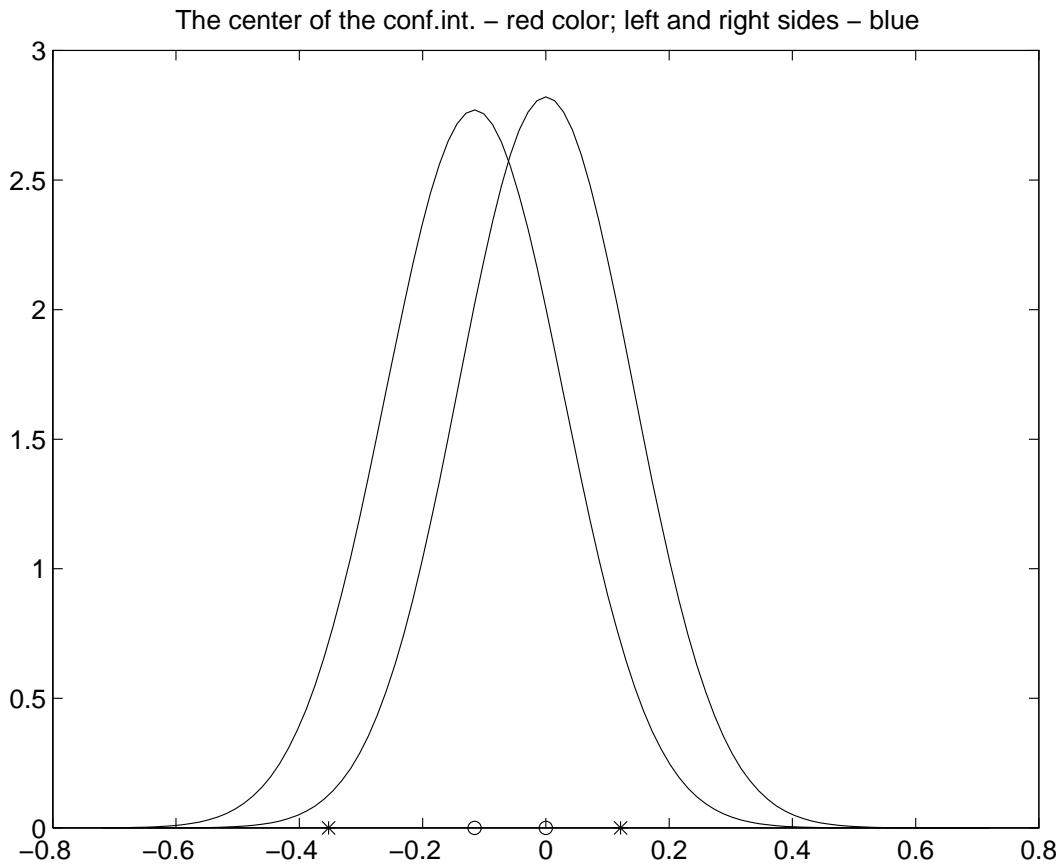


Figure B.1: Graphic presentation of the confidence interval Stars are the left and right sides of the confidence interval. The circle in the middle position between the stars is an estimated value of mathematical expectation. The curves are the density functions for the sample mean (right one) and its estimation using the sampling (left one).

B.6 The time series

B.6.1 Correlated and non-correlating processes

Script `p0covrp.m` returns non-correlating time series and Markov chain, creates sample covariance matrix. Script `r0mark1.m` return the Markov chain.

B.6.2 Auto-correlation and cross-correlation functions

Script `p0corf.m` returns auto-correlation function or cross-correlation function produced by the correlated and non-correlated processes.

B.6.3 Exercises

1. Using script `p0covrp.m` to compare the covariance and correlation matrices for correlated and non-correlated processes. To use different length of time series (f.e. 100, 1000, 10000)
2. Using script `p0corf.m` to compare auto-correlation functions and cross-correlation functions for correlating and non-correlating processes.

B.7 Transforms

B.7.1 Fourier transform

The functions `fft` and `ifft` are realized the discrete Fourier transform and inverse discrete Fourier transform respectively (see Matlab Help). These transforms are determined by the next relations

$$y_k = \sum_{r=0}^{N-1} x_r \exp \left[-i \frac{2\pi kr}{N} \right], \quad k = 0, \dots, N-1 \quad (\text{B.10})$$

(Fourier transform),

$$x_r = \frac{1}{N} \sum_{k=0}^{N-1} y_k \exp \left[i \frac{2\pi kr}{N} \right], \quad r = 0, \dots, N-1 \quad (\text{B.11})$$

(inverse Fourier transform).

B.7.2 Wavelets

See MATLAB 3.5, 3.6, 3.7. Script `p0wlet.m` creates a simple seismic trace: wavelet and Gaussian noise. The spectral characteristics, autocorrelation function, power spectrum of that trace are calculated. The signals created by function `p_wlet.m`, that allows to use the simple signals in time domain:

1. Discrete “delta-function”:

$$x_r = \begin{cases} 1, & \text{if } r = j, \\ 0, & \text{if } r \neq j. \end{cases} \quad (\text{B.12})$$

2. Like delta-function:

$$x_r = \exp \left[-\pi(r - t_p)/t_c \right]^2, \quad r = 0, \dots, N-1 \quad (\text{B.13})$$

where t_c determines the width of the signal and t_p is determined the time shift in the time axis.

3. Ricker wavelet:

$$x_r = \exp \left[-\left(\pi * (r - (t_c + t_p)) / t_c \right)^2 \right] \left(1 - 2 \left(\pi(r - (t_c + t_p)) / t_c \right)^2 \right), \\ r = 0, \dots, N - 1, \quad (\text{B.14})$$

where t_c determines the width of the signal and t_p is determined the time shift in the time axis.

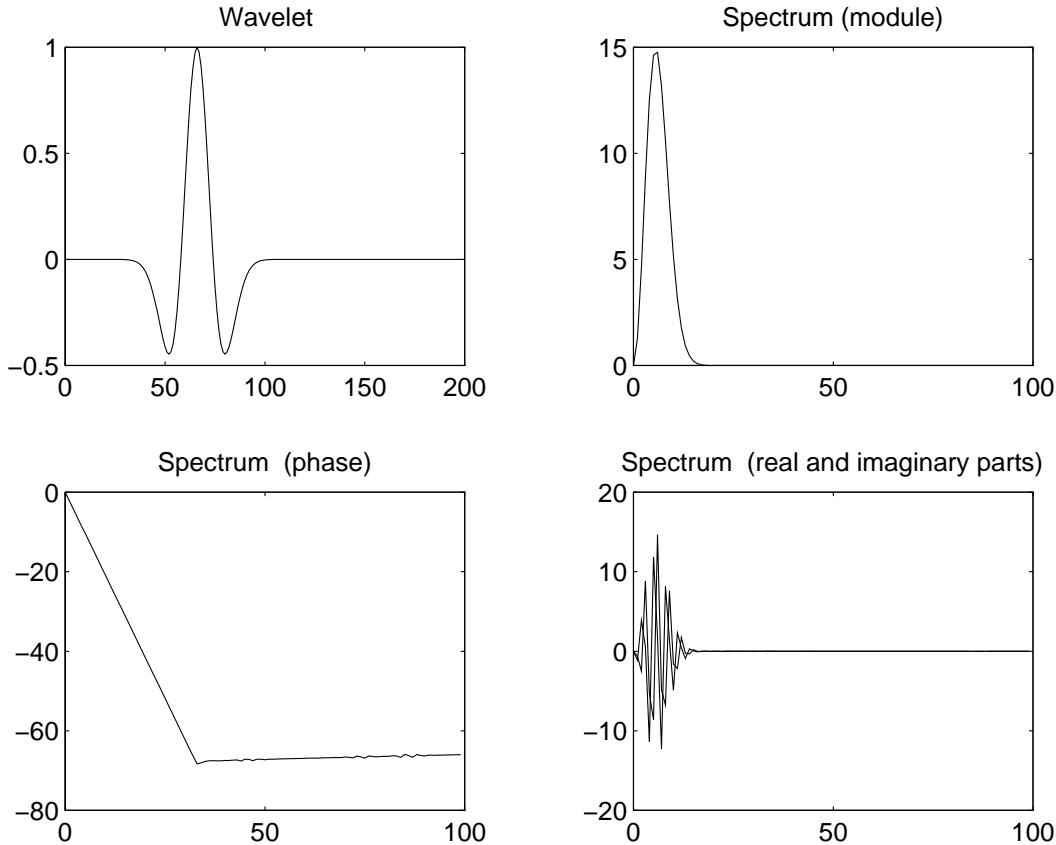


Figure B.2: An example of the graphic output of script `p0wlet.m`.

B.7.3 Multifactor analysis

See 11. Script `p0cater1.m` is realized the multifactor analysis.

Let us consider an example of application of the script `p0cater1.m` to processing of the simple trace “signal with noise” – Figure B.3. The input data are the next:

```
np=500; %the length of trace (sample)
jm=20; % maximum number of principal components
unzer=2; % number of principal components
% (nonzero eigenvalues) for recovery
op=3; % a type of the wavelet
```

```

tp=100;           % the time shift of the wavelet (sample)
%
xm=0;            % expectation of the noise
s=0.05;          % st deviation of the noise

```

The covariance matrix, its eigenvalues and eigenvectors are presented at Figure B.4, Figure B.5 and Figure B.6 correspondingly. The diagrams of the pairs of the principal components and a full collection of the principal components is presented at Figure B.7 and Figure B.8 correspondingly (see *Example from 11*). The result of recovery of the noise using two principal components is presented at Figure B.9 together with input data.

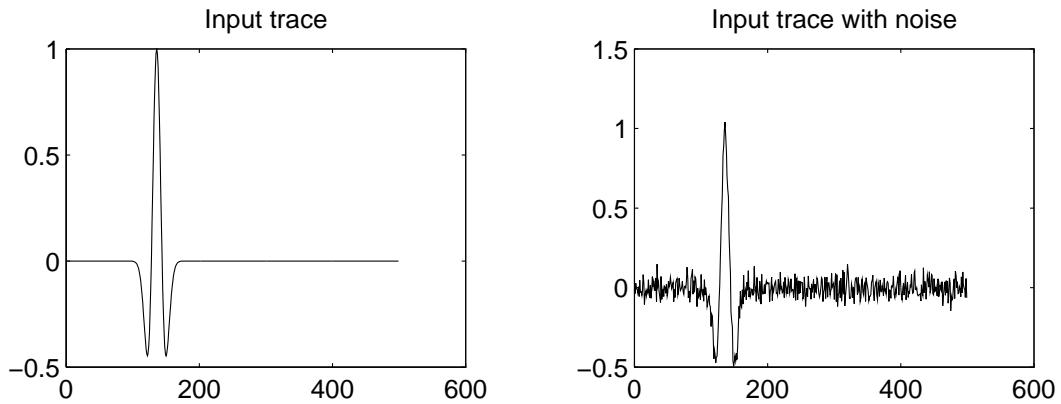


Figure B.3: Input traces.

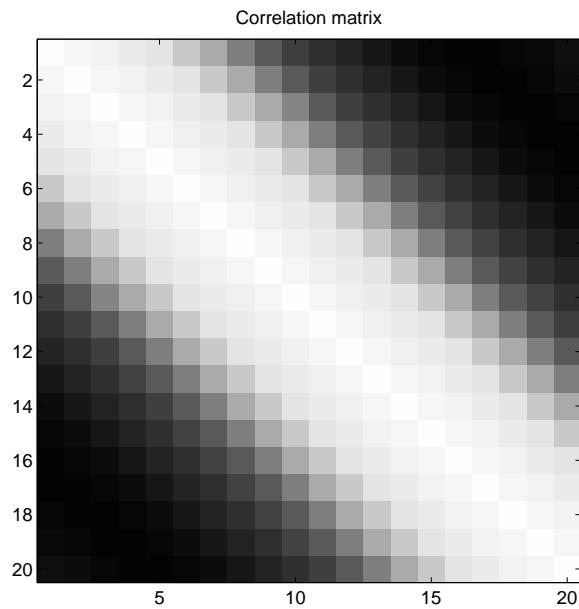


Figure B.4: Correlation matrix $M = 20$.

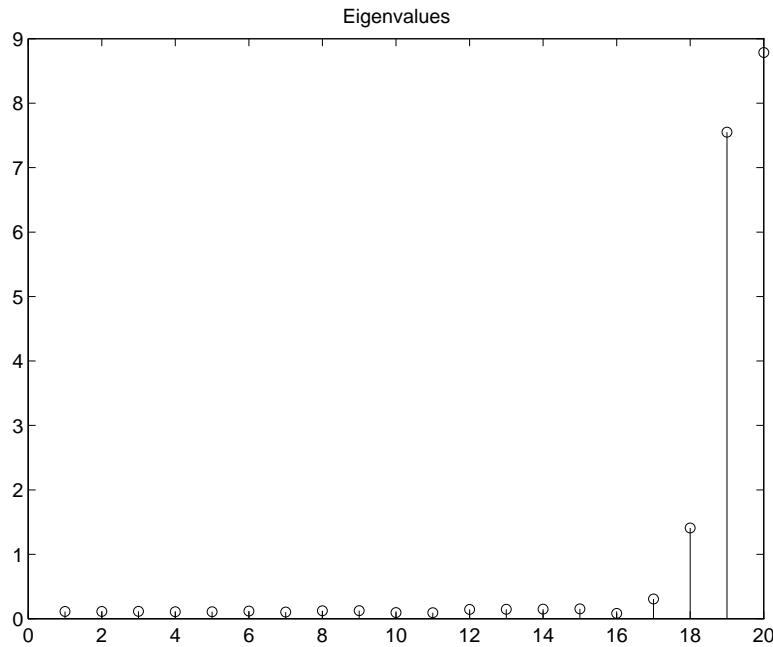


Figure B.5: Eigenvalues of the correlation matrix.

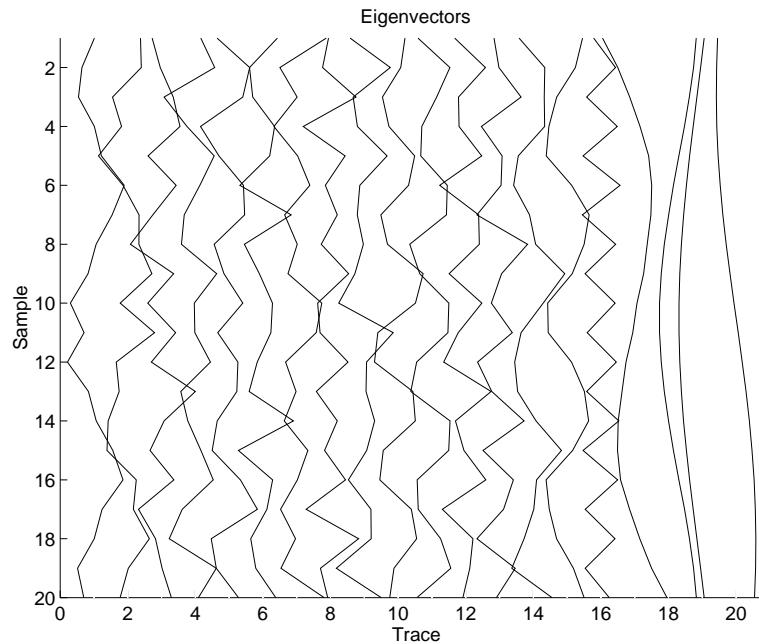


Figure B.6: Eigenvectors of the correlation matrix.

B.7.4 Cepstral transform

Script p0cep1.m is realized the cepstral transform. The Fourier transform from time domain to frequency domain allows to make operation, equivalent to convolution, with

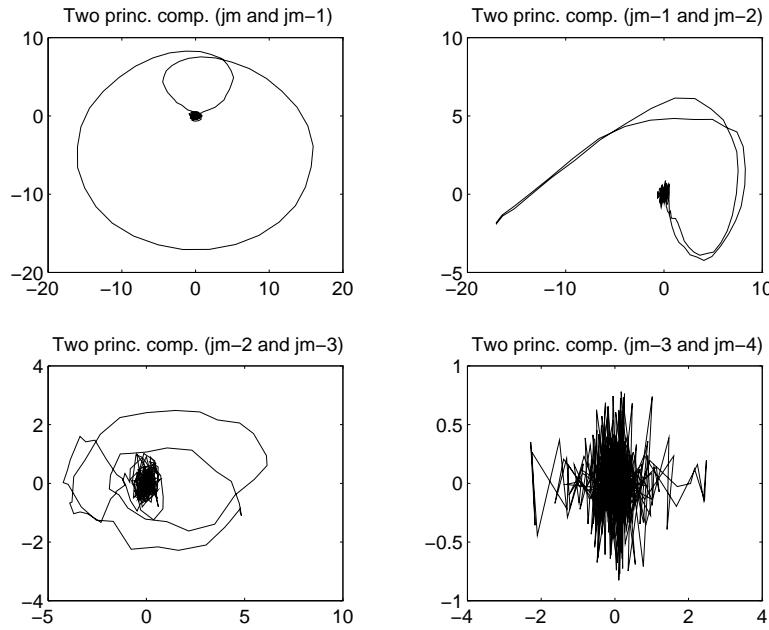


Figure B.7: Diagrams of pairs of the principal components ($jm = 20$).

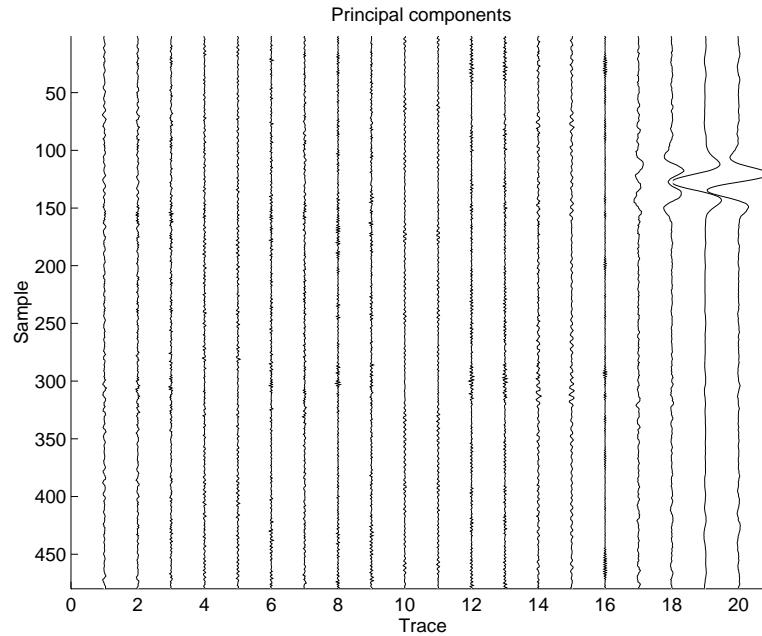


Figure B.8: A full system of the principal components.

the help of simple multiplication. The cepstral transform from frequency domain to cepstral domain enables to carry out such operations by means of more simple action - addition. Besides in some cases the frequencies, which are superposed in frequency domain are rather divided in cepstral domain, so filtering can be more effective.

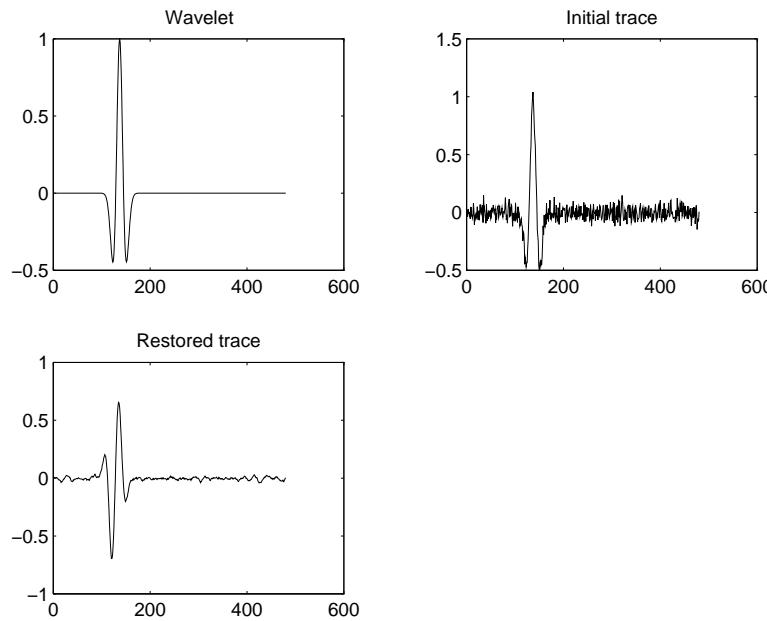


Figure B.9: Input traces (without and with noise) and the trace after processing (using of two principal components).

The cepstral transform in `p0cep1.m` is realized as following:

```
d=fft(trace); % Fourier transform of the vector "trace"
dabs=abs(d); % the absolute value of the spectrum
dang=unwrap(angle(d)); % the phase of the spectrum
%about "unwrap" and "angle" - see MATLAB HELP
dcep=fft(log(dabs)+i*dang); % Fourier transform of logorithm
% of the spectrum (cepstrum)
```

The inverse cepstral transform has the next realization:

```
dback=ifft(dcep); % the inverse Fourier transform
dexp=exp(dback); % "exp" - means exponent
dback0=ifft(dexp); % the inverse Fourier transform
```

At the Figure B.10 and Figure B.11 are presented the two wavelets and their cepstral characteristics correspondingly.

B.7.5 Exercises

1. Using script `p0wlet.m` to calculate the spectra of signals (Ricker wavelet, δ -function) for different values of t_p (the time shift). To implement this with non-correlated Gaussian noise for different values of signal/noise ratio.
2. Using script `p0cater1.m` to implement processing of the trace (wavelet with Gaussian noise) for the cases of different values of caterpillar length (5 – 20),

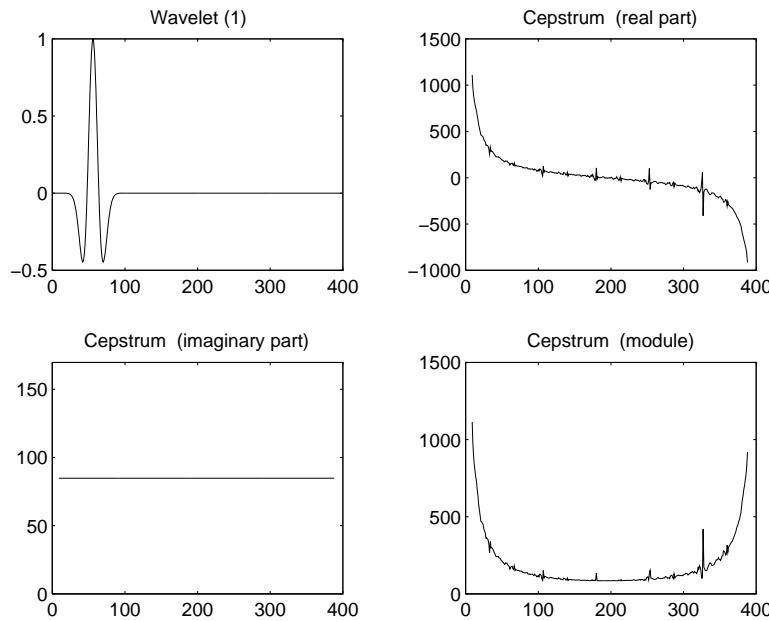


Figure B.10: The first wavelet and its cepstrum characteristics.

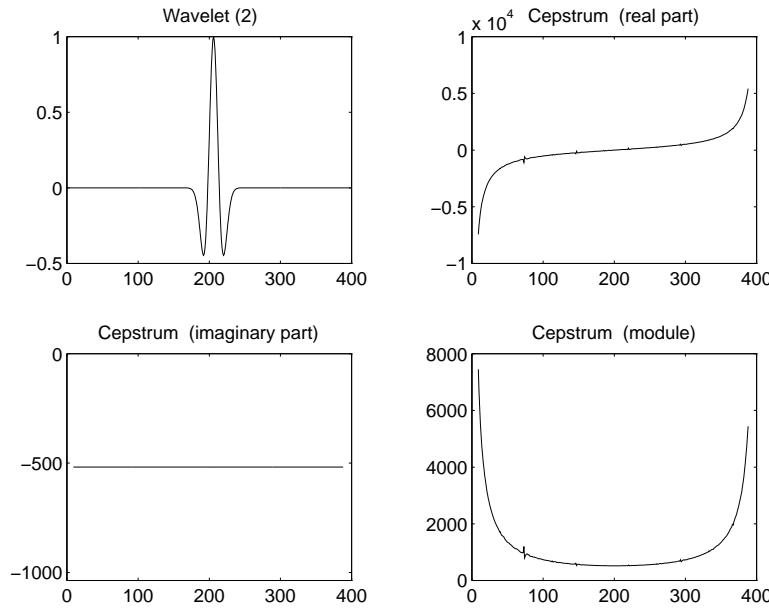


Figure B.11: The first wavelet and its cepstrum characteristics.

different number of nonzero eigenvectors (2 – 10), different values of signal/noise ratio.

3. Using script p0cep1.m to verify that convolution in time domain is the sum in

The result of the convolution of these wavelets is compared with the inverse cepstral transform of the sum of their cepstrums – see Figure B.12.

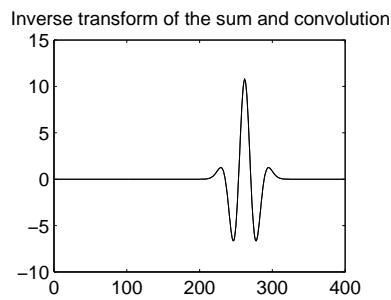


Figure B.12: Comparison of the convolution and summation in the cepstral domain.

cepstral domain.

Appendix C

Direct and inverse problem solution

C.1 Computer simulation of gravitational attraction

C.1.1 Gravitational attraction of a sphere

Script `f0sph1.m` is realized the gravitational attraction of uniform sphere.

Gravitational potential of a sphere is given by

$$V = \gamma \frac{4}{3} \frac{\pi a^3 \rho}{r}, \quad (\text{C.1})$$

where

γ is the gravitational constant;

a is a radius of the sphere;

ρ is mass density;

r is a distance between a center of the sphere and an observation point.

Thus gravitational attraction reads as

$$\vec{g} = -\gamma \frac{4}{3} \frac{\pi a^3 \rho}{r^2} \vec{n}, \quad (\text{C.2})$$

where \vec{n} is a unit vector defined direction from a center of the sphere to an observation point.

C.1.2 Gravitational attraction of a horizontal cylinder

Script `f0cyl1.m` is realized the gravitational attraction of uniform cylinder.

Gravitational potential of a cylinder is given by

$$V = 2\pi a^2 \rho \gamma \log \frac{1}{r}, \quad (\text{C.3})$$

where

γ is the gravitational constant;

a is a radius of the cylinder;

ρ is mass density;

r is a perpendicular distance between a center of the cylinder and an observation point.

Thus gravitational attraction reads as

$$\vec{g} = -\frac{2\pi a^2 \rho \gamma}{r} \vec{n}, \quad (\text{C.4})$$

where \vec{n} is a unit vector defined direction from a center of the cylinder to an observation point

C.2 Computer simulation of magnetic induction

C.2.1 Magnetic induction of a dipole

Script `f0dip1.m` is realized the magnetic induction a dipole.

Magnetic potential of a dipole is given by

$$V = C_m \frac{(\vec{m} \cdot \vec{r})}{r^3}, \quad (\text{C.5})$$

where

C_m is equal $\mu_0/4\pi$ in SI units;

\vec{m} is the dipole moment;

r is adistance between a dipole and an observation point.

Thus magnetic induction reads as

$$\vec{B} = C_m \frac{m}{r^3} [3(\vec{n}_m \cdot \vec{n}_r) \vec{n}_r - \vec{n}_m], \quad (\text{C.6})$$

where $m = |\vec{m}|$, $\vec{n}_m = \vec{m}/|\vec{m}|$, $\vec{n}_r = \vec{r}/|\vec{r}|$.

C.2.2 Magnetic induction of a horizontal cylinder

Script `f0mcyl1.m` is realized the magnetic induction of a magnetic cylinder.

Magnetic potential of a cylinder is given by

$$V = 2C_m \frac{(\vec{m} \cdot \vec{n}_r)}{r}, \quad (\text{C.7})$$

where

C_m is equal $\mu_0/4\pi$ in SI units;

\vec{m} is the dipole moment per unit length;

r is a distance between a center of the cylinder and an observation point;

$\vec{n}_r = \vec{r}/|\vec{r}|$.

Thus magnetic induction reads as

$$\vec{B} = \frac{2C_m m}{r^2} [2(\vec{n}_m \cdot \vec{n}_r) \vec{n}_r - \vec{n}_m] \quad (\text{C.8})$$

where $m = |\vec{m}|$, $\vec{n}_m = \vec{m}/|\vec{m}|$.

C.3 Computer simulation of seismic field

In the script `fd0mn1.m` the numerical simulation of the propagation of the elastic waves in the piecewise half-space (1-D case) is realized by the finite difference method. The field is satisfied the partial equation

$$(\lambda + 2\mu) \frac{\partial^2 \varphi}{\partial z^2} + \frac{\partial(\lambda + 2\mu)}{\partial z} \frac{\partial \varphi}{\partial z} + f = \rho \frac{\partial^2 \varphi}{\partial t^2}, \quad (\text{C.9})$$

boundary conditions at the free surface ($z = 0$)

$$\left. \frac{\partial \varphi}{\partial z} \right|_{z=0} = 0 \quad (\text{C.10})$$

and at welded interfaces ($z = z_i$)

$$\varphi|_{z=z_i-0} = \varphi|_{z=z_i+0}, \quad (\lambda + 2\mu) \left. \frac{\partial \varphi}{\partial z} \right|_{z=z_i-0} = (\lambda + 2\mu) \left. \frac{\partial \varphi}{\partial z} \right|_{z=z_i+0}, \quad (\text{C.11})$$

where ρ is mass density, λ and μ are Lame parameters. The piecewise homogeneous medium can be perturbed by a smooth inhomogeneity of $E = \lambda + 2\mu$ parameter, which is given by the formula

$$\delta E(z) = \begin{cases} \delta E = 0 & \text{if } |z - \tilde{z}| > \Delta, \\ \delta E = E_m 0.5[1 + \cos(\pi(z - \tilde{z})/\Delta)] & \text{if } |z - \tilde{z}| < \Delta \end{cases} \quad (\text{C.12})$$

or smooth inhomogeneity of ρ parameter

$$\delta\rho(z) = \begin{cases} \delta\rho = 0 & \text{if } |z - \tilde{z}| > \Delta, \\ \delta\rho = \rho_m 0.5[1 + \cos(\pi(z - \tilde{z})/\Delta)] & \text{if } |z - \tilde{z}| < \Delta, \end{cases} \quad (\text{C.13})$$

where \tilde{z} is inhomogeneity location, Δ is its half-size, E_m and ρ_m are the maximum values of a perturbation of E and ρ respectively. The source function f from (C.9) is the point source

$$f = \delta(z)\hat{f}(t). \quad (\text{C.14})$$

The time function $\hat{f}(t)$ can be $\delta(t)$ or, for example, Ricker wavelet with the spectral function (C.36).

C.3.1 Exercises

To calculate the seismic wave field for piecewise homogeneous half-space.

Let us consider the next model of the half-space:

the source location – $z_s = 0.003$ km;

the receiver locations – $z_r = 0.003, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5$ km;

location of the interfaces $z=0.3, 0.5, 0.7, 0.9$ km;

Lame parameters – $E = \lambda + 2\mu = 1$;

mass density $\rho = 1.0, 1.2, 0.8, 1.2, 1.0$;

time sample $dt = 0.001$ s;

as the source time dependence the Ricker wavelet of 25 Hz frequency is used.

At Figure C.1 the model, wavelet and the first trace are presented. The seismogram is presented at Figure C.2.

C.4 Deconvolution by the Wiener filter

Script p0decon2.m gives an example of application of the method of least squares (see 6.1). At the first step the Wiener filter is created, then this filter is used for processing of seismogram.

The discrete linear convolution of two vectors \vec{h} and \vec{x} (see 3.4) is defined as

$$y_k = \sum_{j=0}^{L_h} h_j x_{k-j}, \quad k = 0, \dots, L_y \quad (L_y = L_h + L_x - 1), \quad (\text{C.15})$$

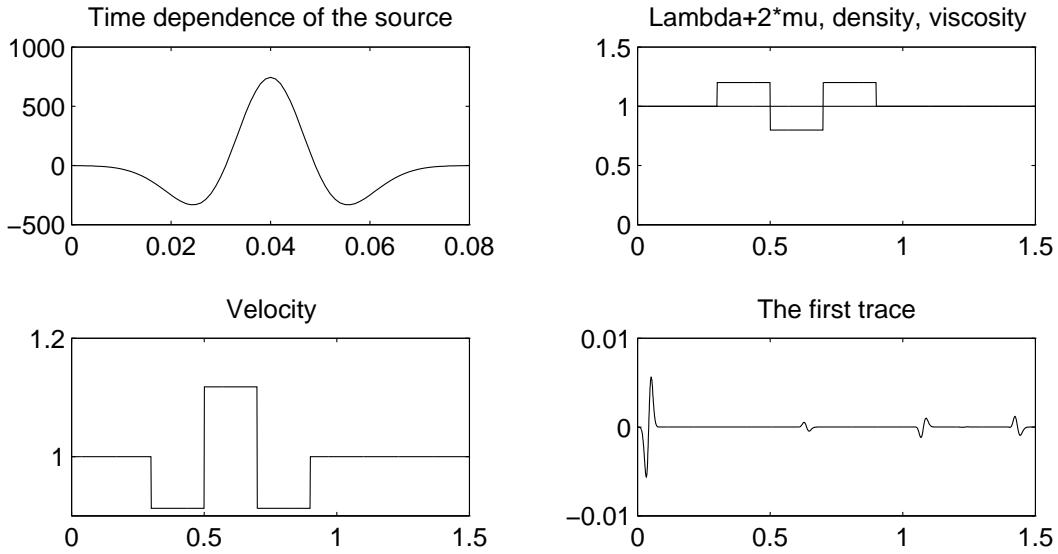


Figure C.1: Graphic presentation of the model and the first trace.

where L_h is the length of the filter \vec{h} , L_x is the length of the input vector \vec{x} . The discrete convolution can be presented in a matrix form

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{L_y} \end{bmatrix} = \begin{bmatrix} x_0 & 0 & 0 & \dots & 0 \\ x_1 & x_0 & 0 & \dots & 0 \\ x_2 & x_1 & x_0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & x_{L_x} \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ \vdots \\ h_{L_h} \end{bmatrix} \quad (\text{C.16})$$

or

$$\vec{y} = X\vec{h}. \quad (\text{C.17})$$

If \vec{y} and X are known quantities, then using the MLS procedure unknown quantity \vec{h} appears as the solution of the next system of the linear equations

$$X^T X \vec{h} = X^T \vec{y} \quad (\text{C.18})$$

or, introducing a regularization term,

$$(X^T X + \alpha I) \vec{h} = X^T \vec{y}, \quad (\text{C.19})$$

where I is identity matrix and α is regularizing coefficient.

Let only one element of vector \vec{y} is equal 1 and others are equal 0, then filter \vec{h} is named the Wiener filter. In this case \vec{x} is transformed to $\vec{y} = [\dots, 1, \dots]$ by the filter \vec{h} .

Let us we have the seismic field from Exercise C.3.1 (Figure C.2) with additional noise. We shall suppress the noise and contract the signals together with their transformation to more simple (like δ -function) shape.

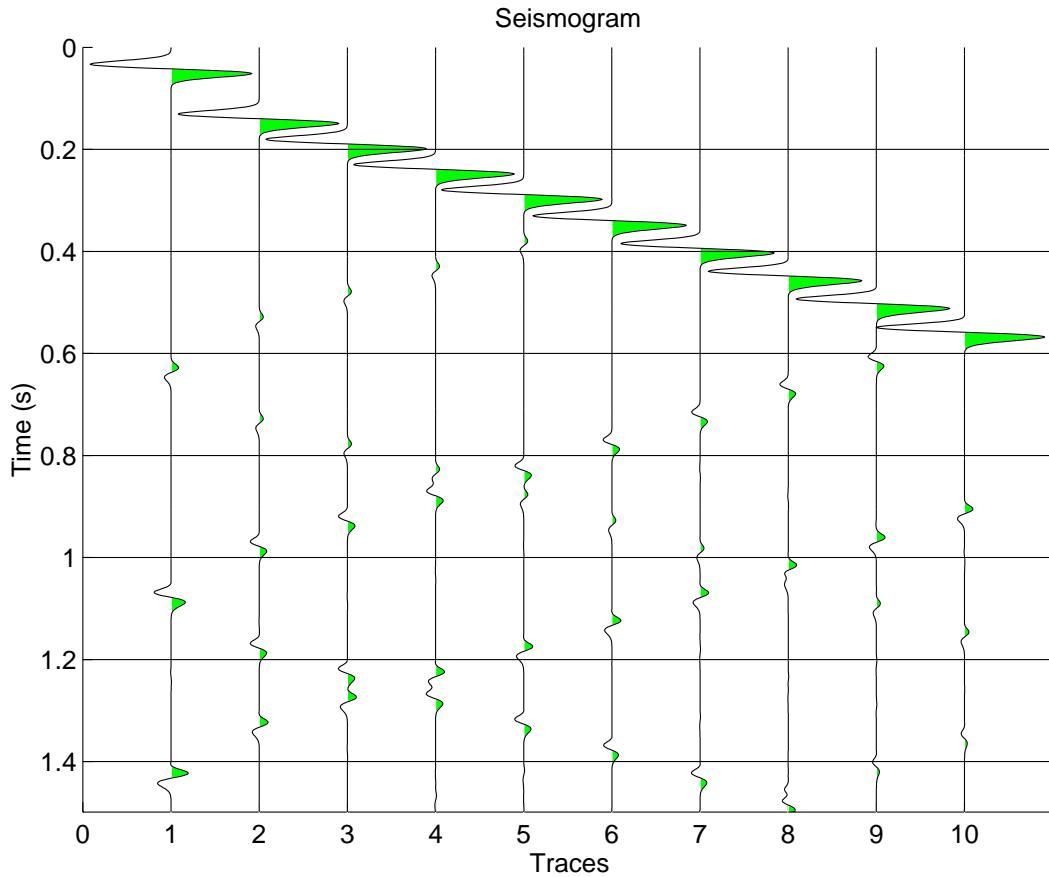


Figure C.2: Seismogram.

Using script `p0decon2.m` we find vector \vec{h} . At the Figure C.3 the seismic signal (curve 1), “ δ -function” (curve 2) and the desired filter (curve 3) are presented. At the Figure C.4 “ δ -function” (curve 1) and the result of convolution of the seismic signal with obtained filter (curve 2). At Figure C.5 eigenvalues of matrices $X^T X$ from (C.18) (curve 1) and $(X^T X + \alpha I)$ from (C.19) (curve 2) are shown. A fragment of the seismogram shown at Figure C.2 with Gaussian noise ($N(0, \sigma A_m)$, where $\sigma=0.1$, A_m is maximum value of seismic signal at seismogram) is presented at Figure C.6. The result of convolution with the obtained filter is shown at Figure C.7.

C.4.1 Exercises

To implement the deconvolution of the seismic field:

without the noise (variables `sig1=sig2=0`);

for different values of regularizing coefficient α (from relation (C.19)) (variable `eps1=2.0, 1.0, 0.5`);

for different values of the noise level `sig1=sig2=0.5, 0.1`.

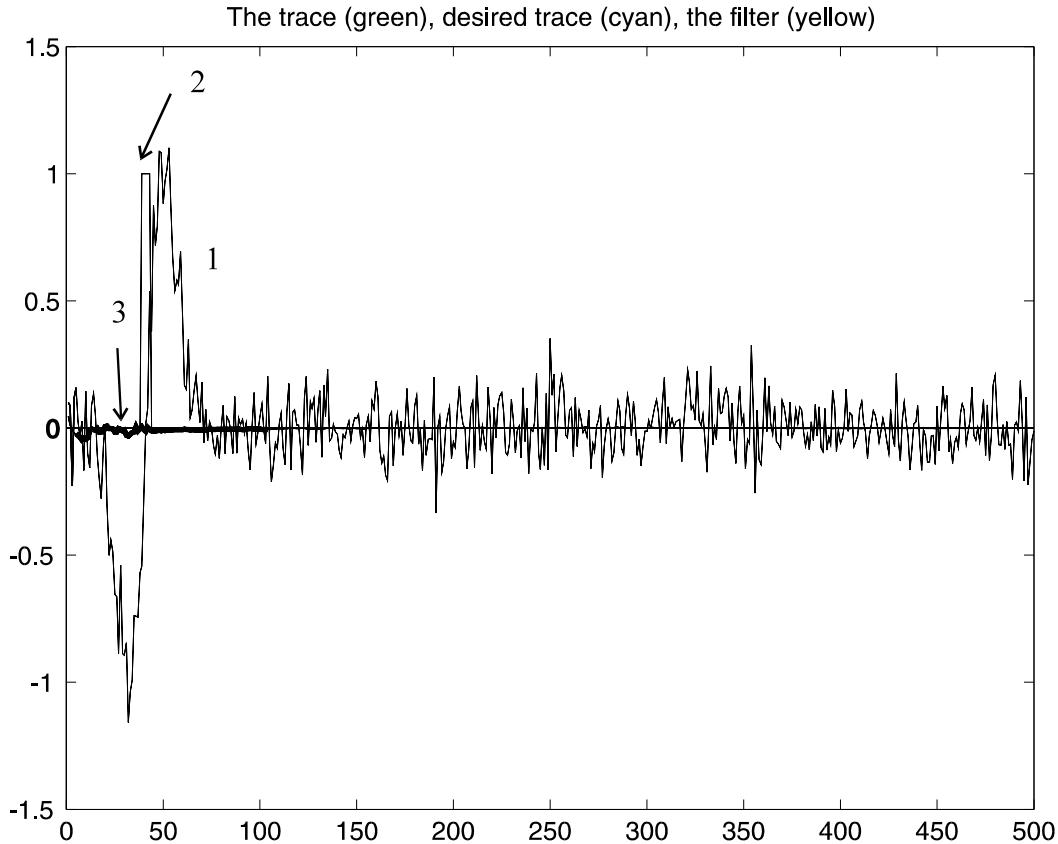


Figure C.3: Seismic signal (1) with Gaussian noise ($N(0, \sigma = 0.1)$), “ δ -function” (2) and obtained filter (3).

C.5 Quantitative interpretation

See 5.2, 6.5, 7.1. Let us consider an application of the Newton-Le Came method to the point estimation of the parameters $\vec{\theta}$ of the geophysical objects (script `p0mg1.m`). As an examples the magnetic bodies with vertical magnetization are considered:

thread of poles

$$f_k(\vec{\theta}) = \frac{2Mh}{h^2 + (x_k - \xi)^2} \quad \vec{\theta} = \{M, h, \xi\}; \quad (\text{C.20})$$

thread of dipoles

$$f_k(\vec{\theta}) = \frac{2M[h^2 - (x_k - \xi^2)^2]}{[h^2 + (x_k - \xi)^2]^2}, \quad \vec{\theta} = \{M, h, \xi\}; \quad (\text{C.21})$$

point magnetic pole

$$f_k(\vec{\theta}) = \frac{M(2h^2 - x_k^2)}{(h^2 + x_k^2)^{5/2}}, \quad \vec{\theta} = \{M, h\}; \quad (\text{C.22})$$

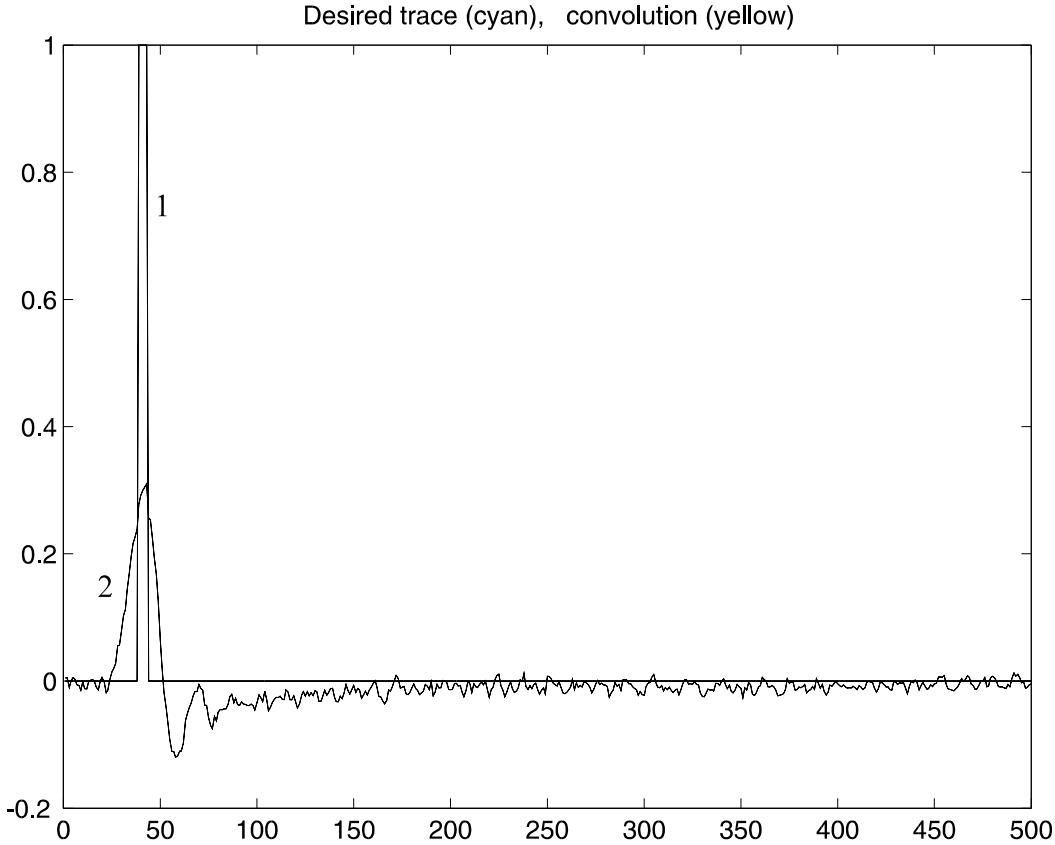


Figure C.4: “ δ -function” (1) and convolution of the seismic signal with obtained filter (2).

point magnetic dipole

$$f_k(\vec{\theta}) = \frac{Mh}{(h^2 + x_k^2)^{3/2}}, \quad \vec{\theta} = \{M, h\}; \quad (\text{C.23})$$

where M is magnetic moment, h is a depth, ξ is a coordinate of the function $f_k(\vec{\theta})$ maximum.

Applying the maximum likelihood method to the additive model of the data

$$u_k = f_k(\vec{\theta}) + \varepsilon_k \quad (\vec{u} = \vec{f}(\vec{\theta}) + \vec{\varepsilon}) \quad (\text{C.24})$$

with uncorrelated Gaussian noise ε_k ($N(0, \sigma)$) we can get the next procedure for estimation of $\vec{\theta} = (\theta_1, \dots, \theta_S)$. At initial point to choose an initial value $\vec{\theta}_0$ of $\vec{\theta}$. After that, to solve a system of the linear equations (see 7.1)

$$\begin{aligned} \tilde{C} \Delta \vec{\theta}^{(1)} &= \vec{d}, \\ d_s &= \frac{1}{\sigma^2} \left[(\vec{u} - \vec{f}(\vec{\theta}))^T \frac{\partial \vec{f}(\vec{\theta})}{\partial \theta_s} \right] \Big|_{\vec{\theta}=\vec{\theta}^{(0)}}, \end{aligned}$$

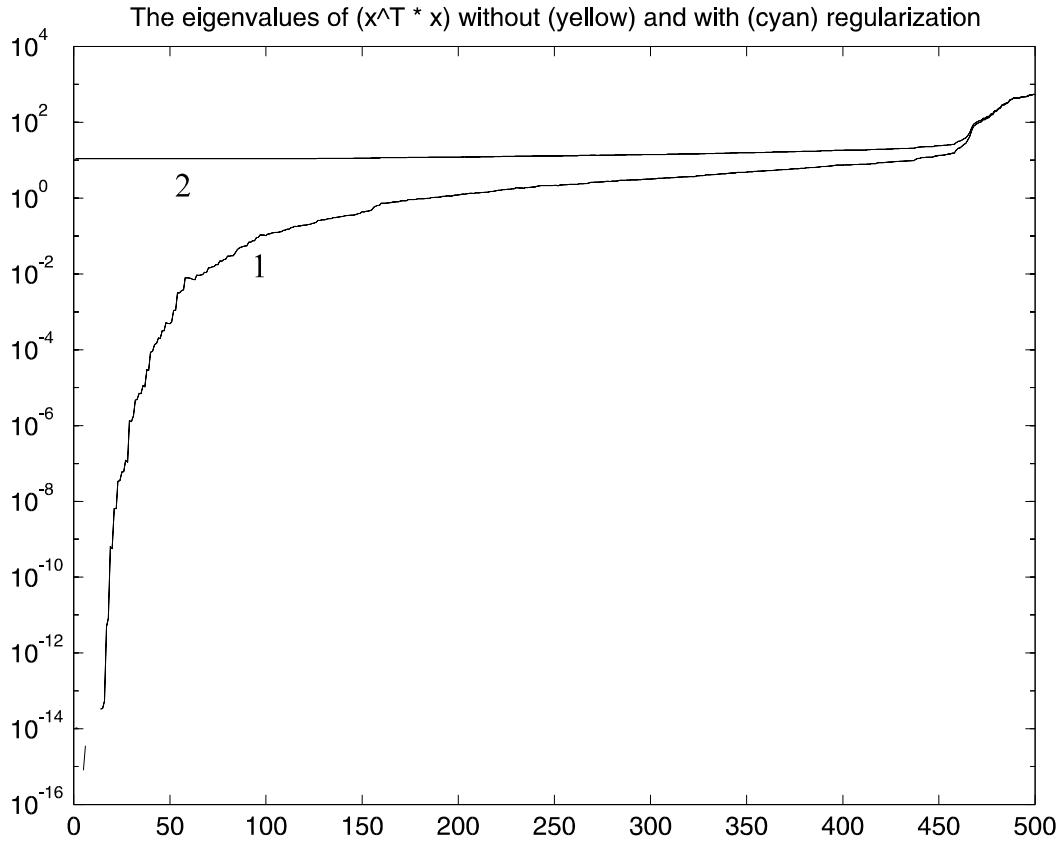


Figure C.5: Eigenvalues of matrices $X^T X$ (curve 1) from (C.18) and $(X^T X + \alpha I)$ (curve 2) from (C.19).

$$\tilde{c}_{ss'} = \frac{1}{\sigma^2} \left. \frac{\partial \vec{f}^T(\vec{\theta})}{\partial \theta_s} \right|_{\vec{\theta}=\vec{\theta}^{(0)}} \left. \frac{\partial \vec{f}(\vec{\theta})}{\partial \theta_{s'}} \right|_{\vec{\theta}=\vec{\theta}^{(0)}},$$

$$s = 1, \dots, S \quad (C.25)$$

and to calculate the first $\hat{\vec{\theta}}^{(1)}$ approximation of $\vec{\theta}$

$$\hat{\vec{\theta}}^{(1)} = \vec{\theta}^{(0)} + \Delta \vec{\theta}^{(1)}. \quad (C.26)$$

The iteration process is finished if the threshold condition is satisfied, for example,

$$|\Delta \hat{\theta}_s^{(n)} / \hat{\theta}_s^{(n)}| \approx 10^{-2} \div 10^{-3}. \quad (C.27)$$

An example of similar numerical simulation for field (C.20) is presented at Figure C.8-C.10 for the next input data

`x=-4:0.1:4; % observation points`

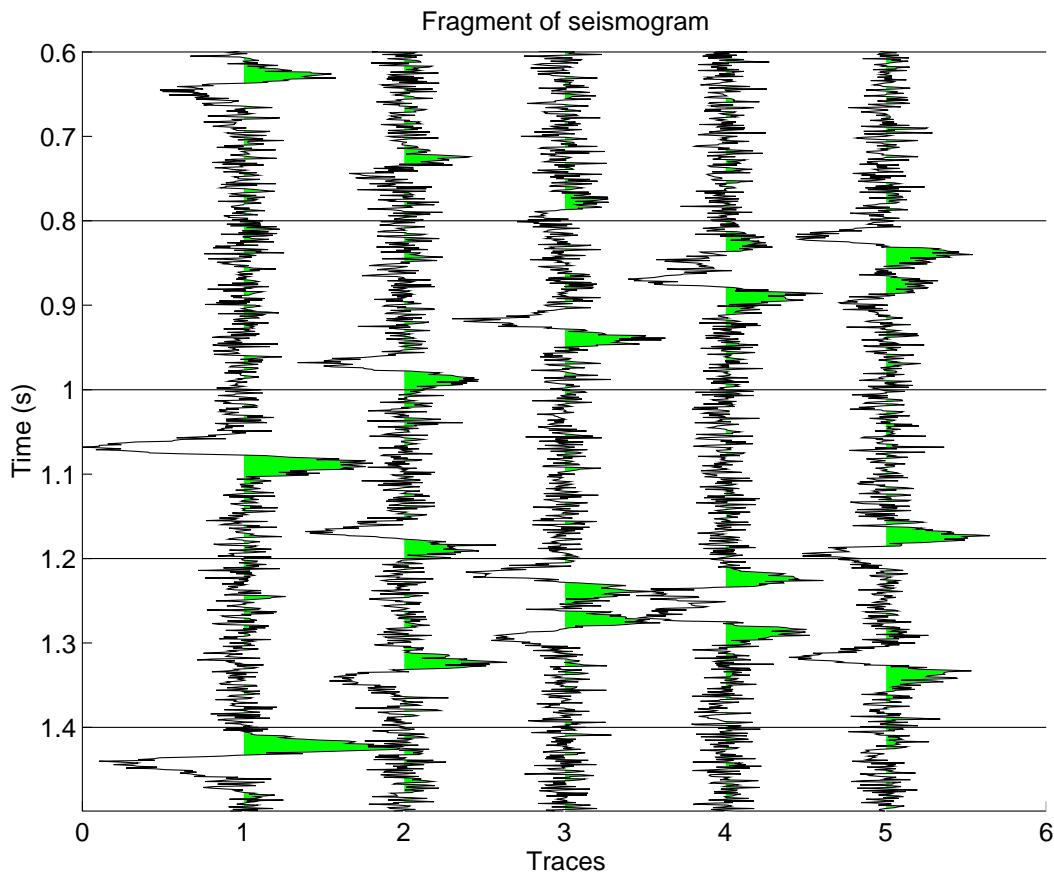


Figure C.6: A fragment of seismogram from Figure C.2 with Gaussian noise.

```

m=1;      %intensity
h=1;      %depth
xi=0.0;    %location of the object
xm=0.0;    %expectation of the noise
s=0.1;     %standard deviation of the noise
m0=0.5;    %initial value of m
h0=1.5;    %initial value of h
xi=0.;     %initial value of xi
eps_m=1.0e-5;    %threshold for end of iterations (m parameter)
eps_h=1.0e-5;    %threshold for end of iterations (h parameter)
The values of  $M$  and  $h$  are estimated. As the result of numerical simulation the
Fisher matrix and its inverse matrix are calculated.

```

C.5.1 Exercises

Using script p0mg1.m to implement the next exercises for models (C.20) – (C.23) (to pay attention to the Fisher matrix and its inverse matrix):

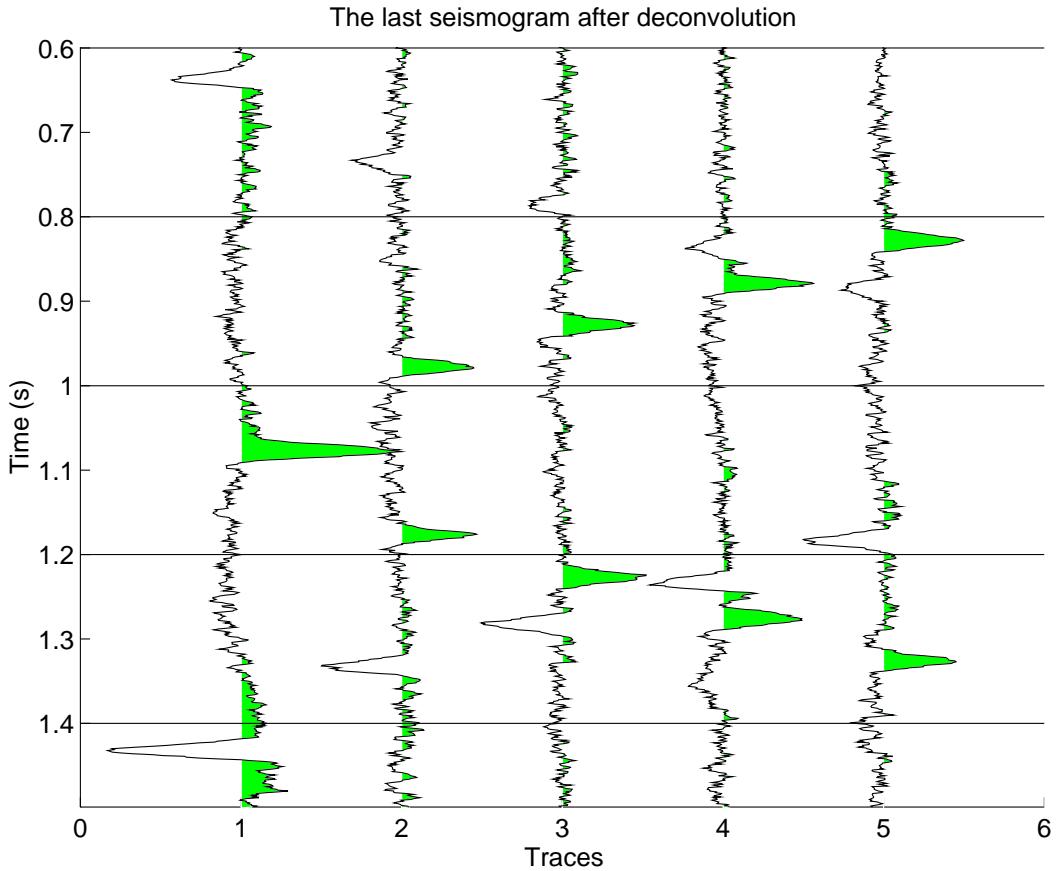


Figure C.7: Deconvolution of seismogram from Figure C.6.

1. To investigate restoration of M and h for different values of standard deviation of Gaussian noise ($\sigma = 0.1, 0.2, 0.3$).
2. To investigate a convergence of the iteration process in dependence on initial values of parameters M and h .
3. To implement the numerical simulation for the different sample sizes (i.e. different numbers of the observation points).

C.6 Qualitative interpretation

Using script `p0mg2.m` to implement the qualitative interpretation (see 5.4 “signal and noise (H_0) or noise (H_1)” with the help of a posteriori probability ratio (see 8, 8.2) for the signal models from (C.20) – (C.23). For the case of uncorrelated Gaussian noise this criterion reads as

$$\alpha = \ln \frac{P(1)}{P(0)} + \frac{1}{2\sigma^2} \left[(\vec{u} - \vec{f}_1)^T (\vec{u} - \vec{f}_1) - \vec{u}^T \vec{u} \right]. \quad (\text{C.28})$$

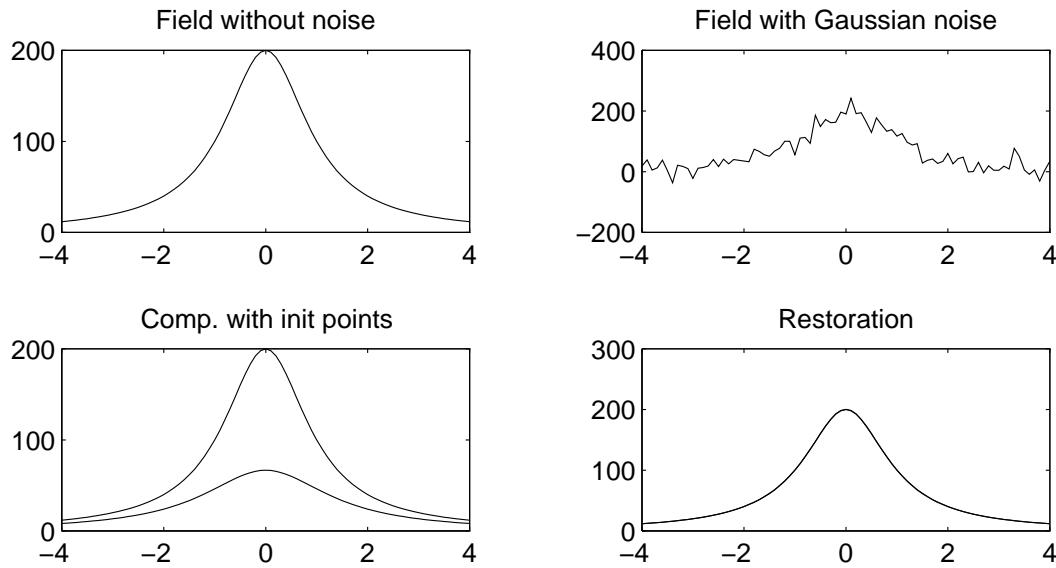


Figure C.8: Model field without noise, model field with noise, model field and field for initial parameters, comparison of the model field and field with restored parameters.

If $\alpha \geq 0$, then hypothesis H_1 (“noise”). If $\alpha < 0$ then hypothesis H_0 (“signal and noise”). We assume that a model function $f_1(\vec{\theta})$ differs from real function $f_1(\vec{\theta} + \Delta\vec{\theta})$ due to $\Delta\vec{\theta}$.

An example of similar numerical simulation for field (C.20) is presented at Figure C.11 for the next input data:

```

x=-4:0.1:4; % observation points
m=1; %intensity
h=1; %depth
xi=0.0; %location of the object
ddm=0.0; %deviation of m
ddh=0.0; %deviation of h
ddxi=0.2; %deviation of xi
xm=0.0; %expectation of the noise
p0=0.5; %a priori probability of existence of the object
s=0.1; %initial standard deviation of the noise
s1=s:0.4:4; % standard deviations of the noise for the numerical simulation

```

C.6.1 Exercises

Using script p0mg2.m to implement the next exercises for models (C.20) – (C.23):

1. To implement numerical simulation for a wide range of a noise/signal ratio (variable s_1).
2. To implement simulation for the different deviations of $\Delta\vec{\theta}$.

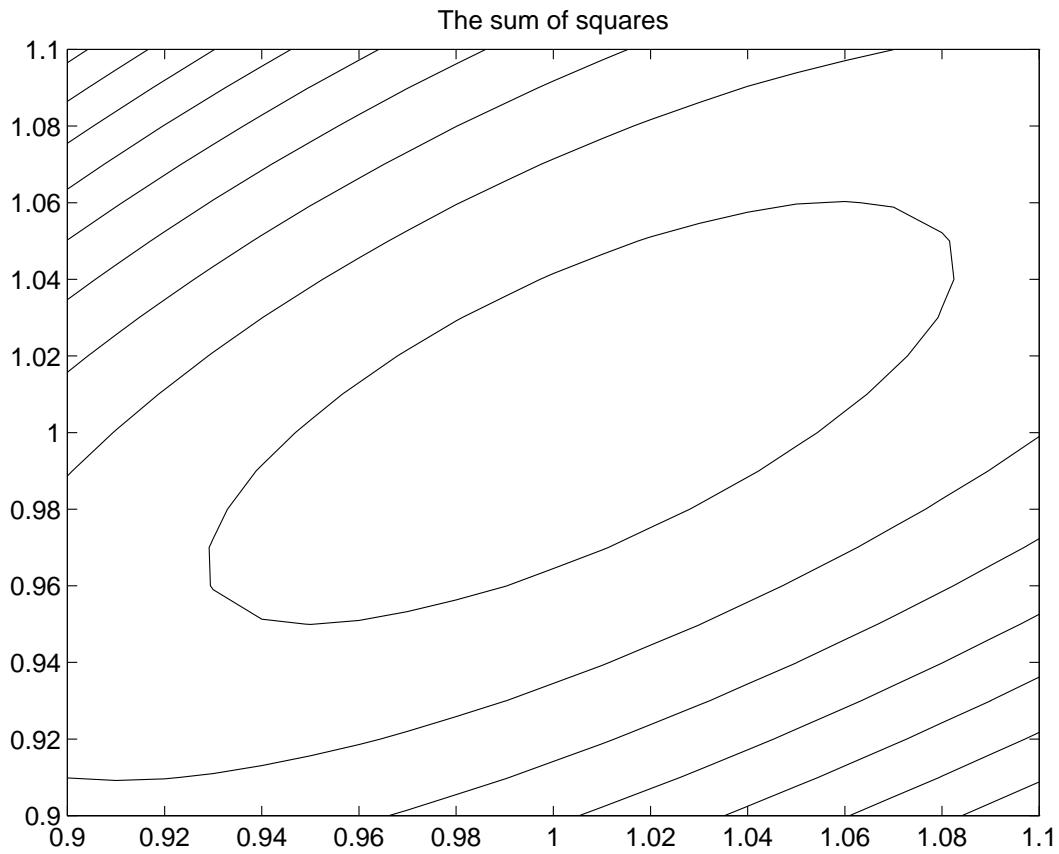


Figure C.9: The values of $(\vec{u} - \vec{f}(\vec{\theta}))^T(\vec{u} - \vec{f}(\vec{\theta}))$ as a function of M and h .

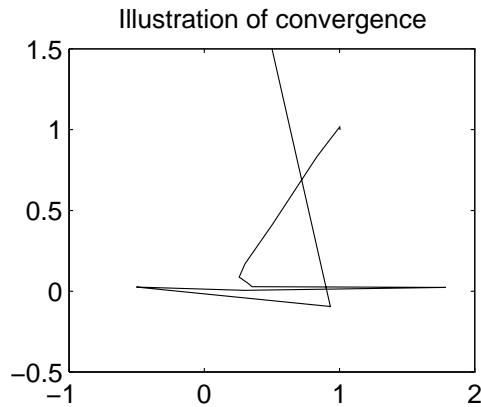


Figure C.10: An illustration of the convergence of the iteration process in the (M, h) plane.

3. To implement simulation for different values of a priori probability (for example, $p_0=0.25, 0.5, 0.75$).

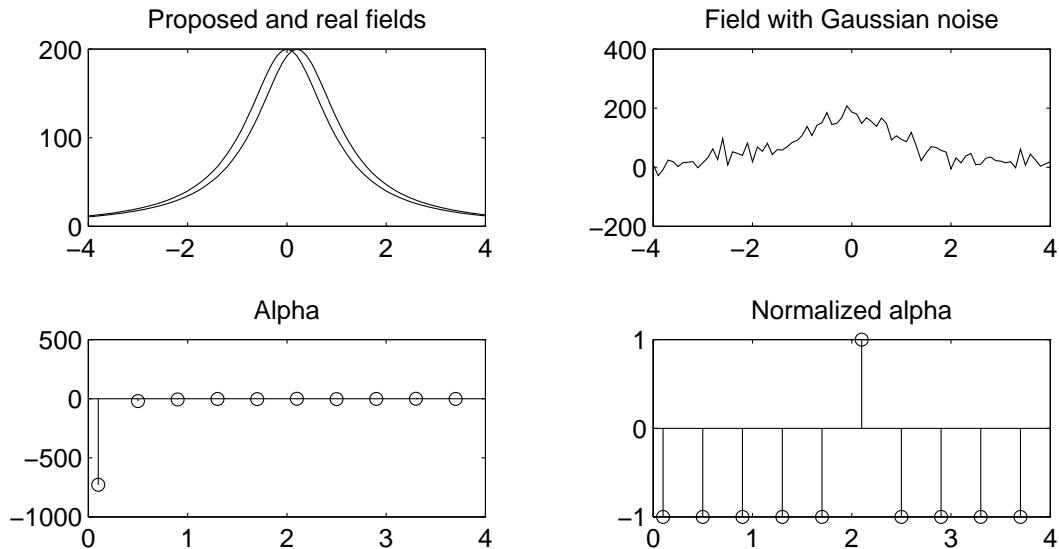


Figure C.11: The model field $f_1(\vec{\theta})$ and real field $f_1(\vec{\theta} + \Delta\vec{\theta})$ with the deviations of the parameters, model field with noise, the parameter α , normalizes parameter α .

4. To implement the numerical simulation for the different sample sizes (i.e. different numbers of the observation points).

C.7 Diffraction tomography

See 10. As an example of application of the *diffraction tomography method* for restoration of the elastic local inhomogeneity, comparable in size with the wavelength, is considered with the help of script `dt0mn1.m`.

In this script the direct problem solution for the propagation of the elastic wave in the uniform half-space (1-D case) containing a target object is solved by the finite difference method. The field is satisfied the partial equation

$$(\lambda + 2\mu) \frac{\partial^2 \varphi}{\partial z^2} + \frac{\partial(\lambda + 2\mu)}{\partial z} \frac{\partial \varphi}{\partial z} + f = \rho \frac{\partial^2 \varphi}{\partial t^2}, \quad f = \delta(z) \tilde{f}(t) \quad (\text{C.29})$$

and boundary condition at the free surface

$$\left. \frac{\partial \varphi}{\partial z} \right|_{z=0} = 0, \quad (\text{C.30})$$

where ρ is mass density, λ and μ are Lame parameters. The restoration of the parameters λ , μ and ρ is realized by the diffraction tomography method as follows.

The scattered (or *difference*) field $\Delta\varphi$ produced by inhomogeneity, using a notion of the tomography functionals p_ρ and p_E , connected with ρ and $E = \lambda + 2\mu$

correspondingly,

$$\begin{aligned} p_\rho(z, z_s, z_r, t) &= - \int_0^\infty \varphi_{out}(z, z_r, t - \tau) \frac{\partial^2}{\partial t^2} \varphi_{in}(z, z_s, \tau) d\tau, \\ p_E(z, z_s, z_r, t) &= - \int_0^\infty \frac{\partial}{\partial z} \varphi_{out}(z, z_r, t - \tau) \frac{\partial}{\partial z} \varphi_{in}(z, z_s, \tau) d\tau, \end{aligned} \quad (\text{C.31})$$

with the use of the first-order Born approximation, can be presented approximately by the next formula

$$\Delta\varphi \approx \int_L (p_E \delta E + p_\rho \delta \rho) dl, \quad (\text{C.32})$$

where

z_s and z_r are the source and receiver locations;

φ_{in} is the wave-field produced by the source with its time dependence;

φ_{out} is the wave-field produced by the “artificial” source, located at the observation point with time dependence as $\delta(t)$ -function;

L is the reconstructed region;

δE and $\delta \rho$ are desired parameters of inhomogeneity.

If, for example, a value of $\delta \rho \approx 0$ and we shall find only a value of δE , then the integral equation (C.32) after digitization can be presented as a system of the linear equations

$$P\vec{\delta}_E = \vec{\delta}_\varphi, \quad (\text{C.33})$$

where $\vec{\delta}_E$ is the desired vector and $\vec{\delta}_\varphi$ is a vector of the sampled difference field.

The simplest regularization of the solution of (C.33) consists of the solution of the next system of the linear equations

$$(P'P + \varepsilon D'D)\vec{\delta}_E = P'\vec{\delta}_\varphi \quad (\text{C.34})$$

where ε is regularizing parameter and D , for example, is identity matrix.

Let us consider an example of numerical simulation for the next model:

$z_s = z_r = 0.003$ km;

$\rho = 1$, $E = 1$, so $v = \sqrt{E/\rho} = 1$;

the smooth inhomogeneity is given by the formula

$$\delta E(z) = \begin{cases} \delta E = 0 & \text{if } |z - \tilde{z}| > \Delta, \\ \delta E = E_m 0.5[1 + \cos(\pi(z - \tilde{z})/\Delta)] & \text{if } |z - \tilde{z}| < \Delta, \end{cases} \quad (\text{C.35})$$

where \tilde{z} is inhomogeneity location, Δ is its half-size, E_m is a maximum value of a perturbation of E .

$\tilde{z} = 0.3$ km, $\Delta = 0.03$ km, $E_m = 0.3$;

the Ricker wavelet

$$F(\omega) = \left(\frac{\omega}{\omega_0}\right)^2 \exp\left\{-\left(\frac{\omega}{\omega_0}\right)^2\right\} e^{-2\pi i \omega / \omega_0} \quad (\text{C.36})$$

with apparent frequency $\omega_0 = 25$ Hz is used;

the interval L of restoration is $L = 0.21 - 0.39$ km;

regularizing parameter $\varepsilon = 10^{-5}$.

Graphic presentation of the model and the results of the direct problem solution are presented at Figure C.12 and Figure C.13. Figure C.14 gives a graphic presentation of

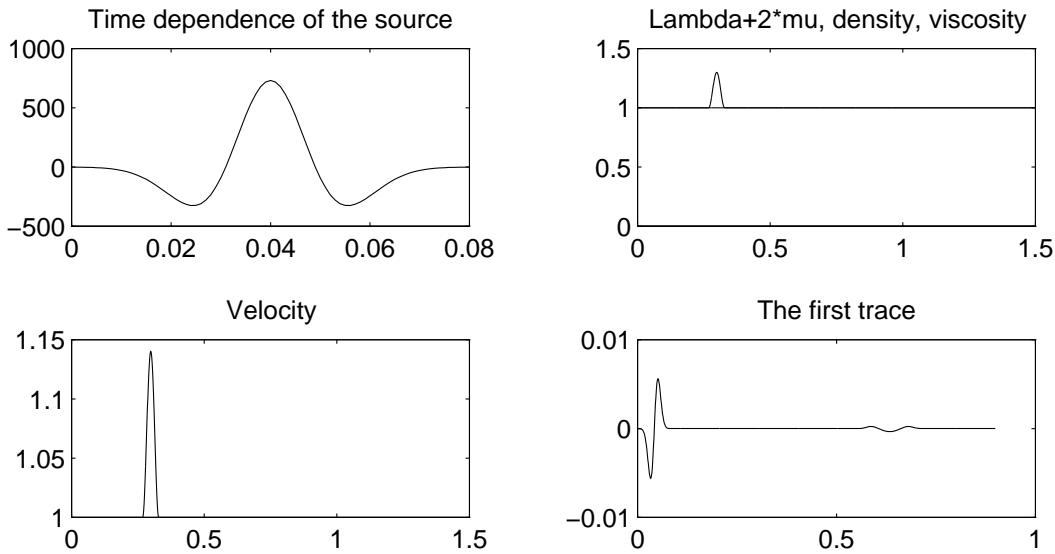


Figure C.12: Graphic presentation of the model and a result of the direct problem solution.

a system of the linear equations (C.33). The last trace is the right side of the equation (C.33) (see also Figure C.15). Others traces correspond to matrix P (actually it's a graphic presentation of the matrix P with the opposite sign). The first trace of Figure C.14 is presented at the Figure C.15 too. The result of reconstruction and the model of inhomogeneity is presented at Figure C.16.

C.7.1 Exercises

To investigate a quality of the restoration of perturbation of E parameter by diffraction tomography method in dependence on contrast and size of inhomogeneity (f.e. $E = 0.05, 0.1, 0.2, 0.5$ $\Delta = 0.01, 0.02, 0.03, 0.04$).

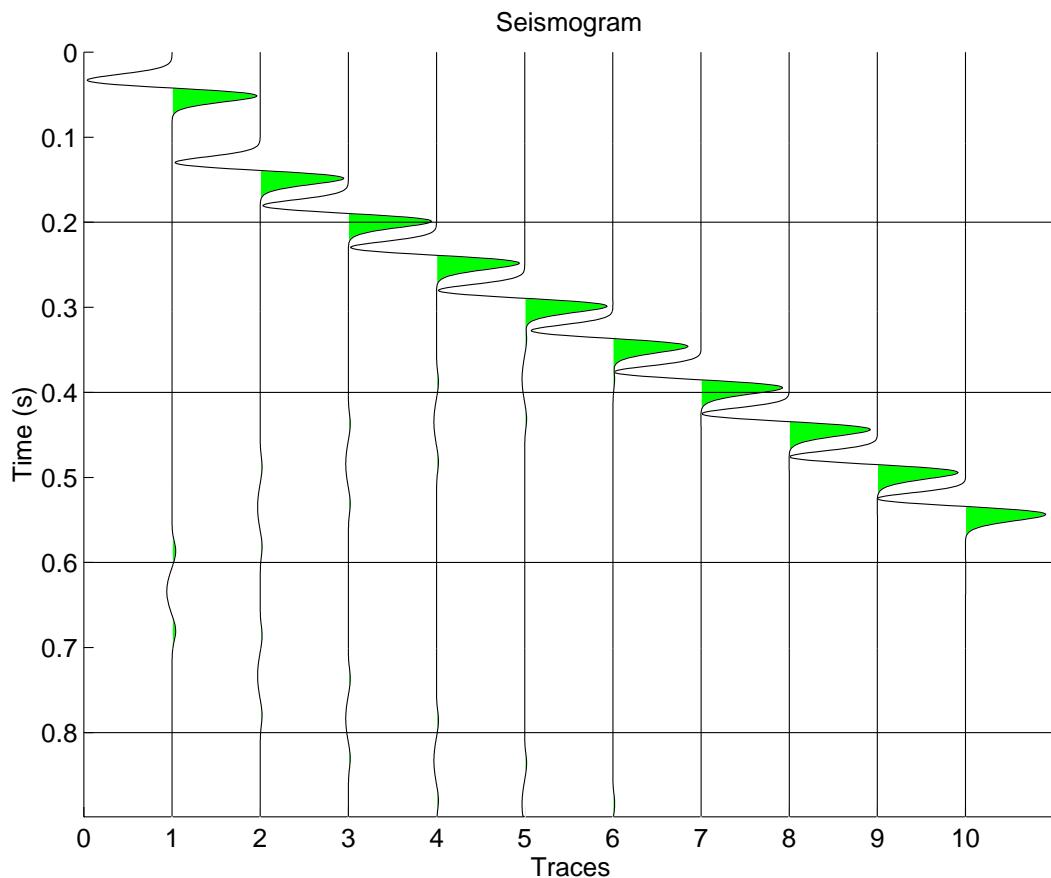


Figure C.13: The seismogram for the observation points (km): 0.003, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5.

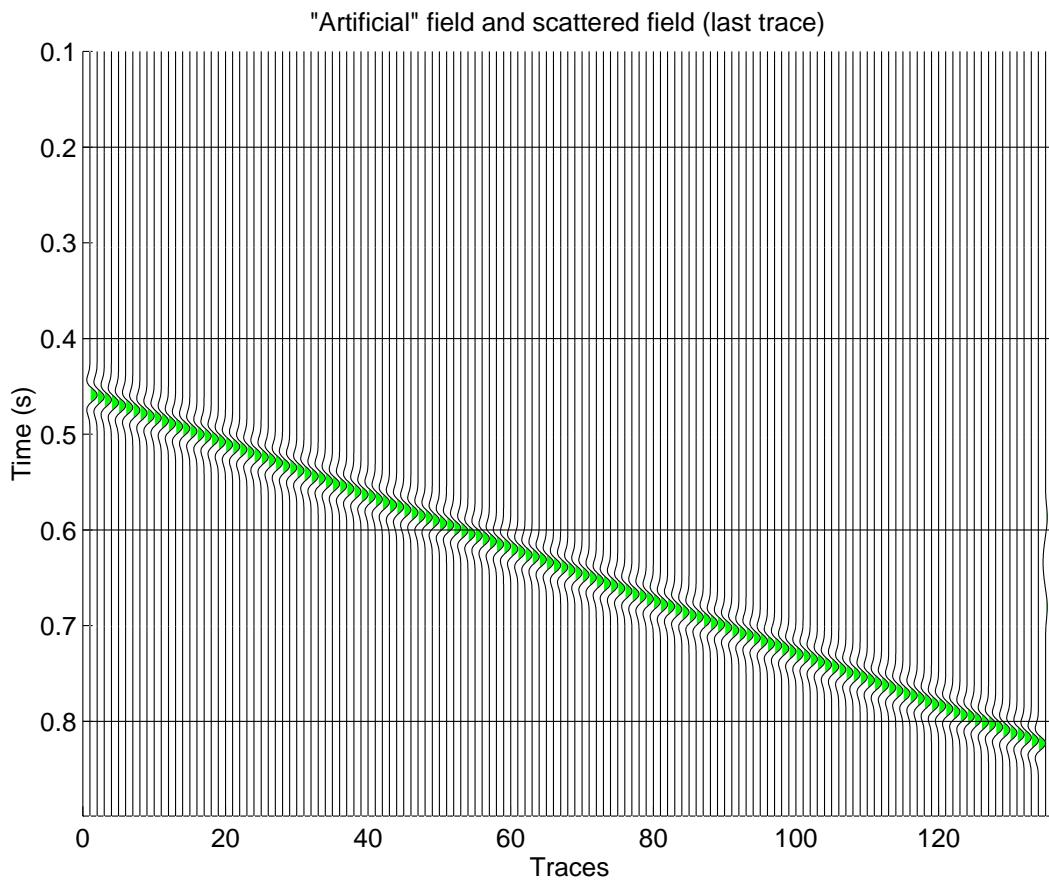


Figure C.14: The “artificial” field and the scattered field (last trace).

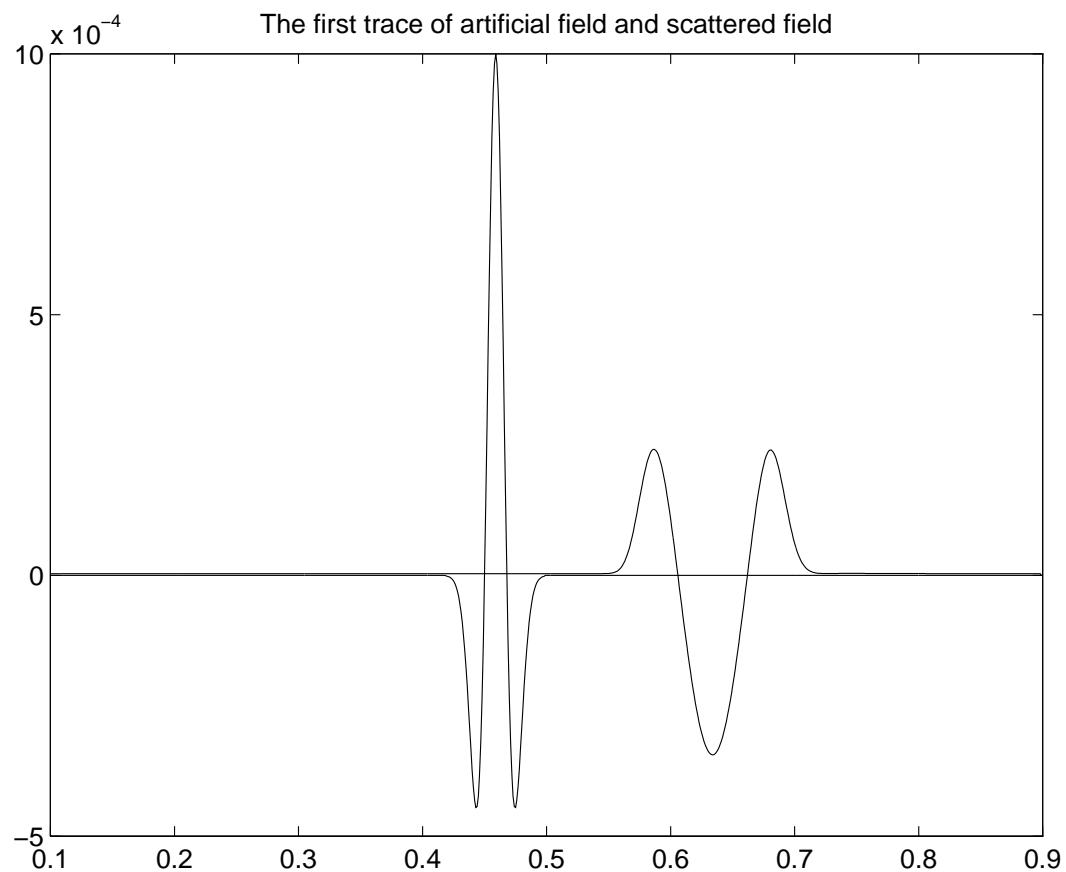


Figure C.15: The first trace of the “artificial” field and scattered field.

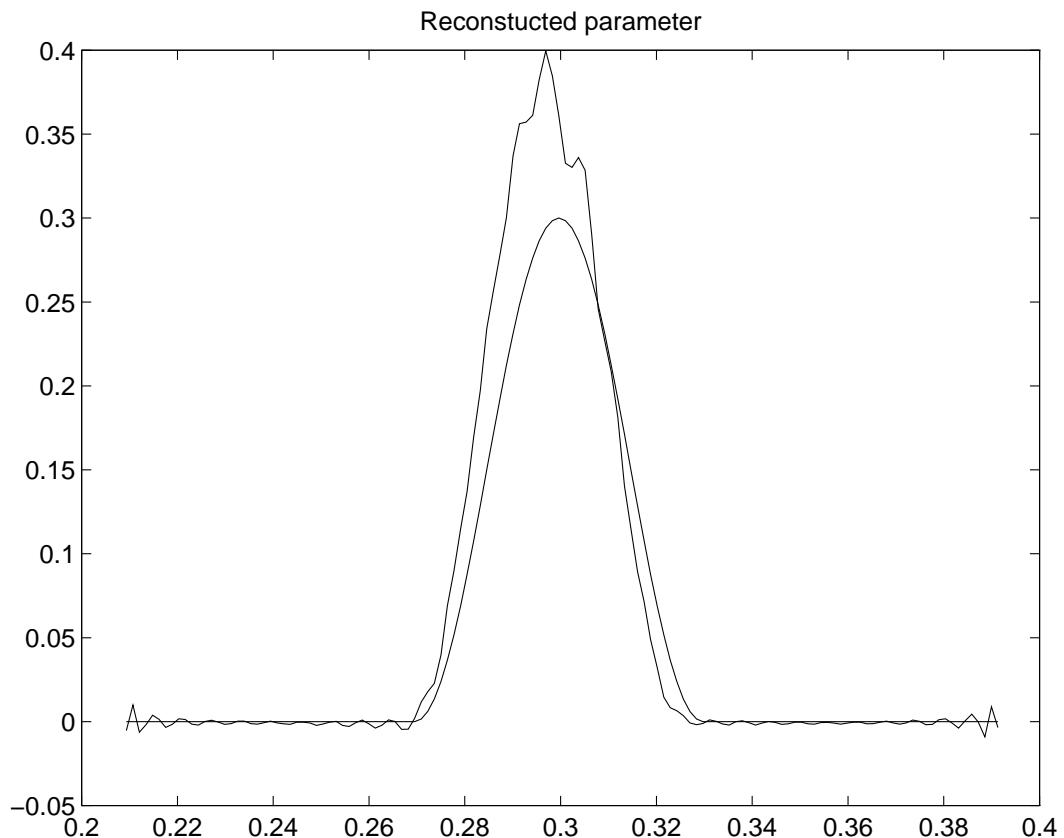


Figure C.16: The model and result of restoration.

Index

- p*-percentage point
 - lower, 20
 - upper, 20
- Estimation*
 - maximum a posteriori*, 98
- Absolute deviation*
 - mean, 21
- Autocorrelation*
 - function, 56
- Coefficient*
 - asymmetry, 23
 - covariation, 19
- Condition number*, 102
- Correlation*
 - function, 56
- Data*, 65
- Deciles*, 20
- Density*
 - conditional, 14
 - marginal, 14
- Density function*
 - joint, 13
 - normal
 - standard, 33
- Dispersion*, 18
- Distribution*
 - normal, 77, 80
 - Poisson, 30
 - uniform, 80
- Distribution function*, 9
- Distribution function*
 - normal
 - standard, 33
- Entropy*
 - conditional
- complete (average), 48
 - particular, 48
- Estimation*
 - maximum likelihood*, 91
 - maximum likelihood*
 - generalized, 97
 - point, 66
- Estimator*, 66
- Estimator*
 - biased, 67
 - consistent, 66
 - efficient, 69
 - sufficient, 70
- Event*
 - certain, 4
 - elementary, 1
 - impossible, 4
 - random, 1
- Events*
 - algebra, 5
- Excess*, 23
- Fisher*
 - matrix, 52
- iif*, 2
- Information*
 - Fisher, 69
 - mutual
 - complete, 47, 50
- Likelihood*, 90
- Likelihood function*, 90
- Matrix*
 - Fisher information, 69
 - resolution, 106
- Maximum-likelihood method*
 - linear model, 92

- Measurement, 65
- Median, 20
- Mode (*of the posterior density*), 97
- Model
 - arbitrary additive, 91
- Moments, 22
- Moments
 - central, 23
- Observations, 65
- Perturbation method, 154
- Probability, 5
- Probability
 - a posteriori*, 8
 - a priori*, 8
 - distribution, 9
- Process
 - stochastic, 55
- Radon transform
 - back projection, 143
 - differential operator, 142
 - Hilbert transform, 142
- Range of the probability series, 21
- Regularization
 - statistical, 98
 - Tikhonov, 99
- Resolution length, 172
- Set
 - empty (null), 1
 - of elementary events, 1
 - subset, 1
- Set operations
 - compliment, 3
 - difference, 4
 - equality, 2
 - product (intersection), 2
 - sum, 3
 - union, 3
- Sets
 - mutually exclusive (*disjoint*), 2
- Singular analysis, 104
- Singular value, 103
- Space
- probability, 5
- Standard deviation, 18
- Stationarity
 - strict sense, 56
 - wide sense, 57
- Time series, 55
- Variable
 - random
 - expectation, 17
- Variation coefficient, 21
- Vector
 - random, 13
- Width
 - interquartile, 21

Bibliography

- [1] Aki K., Richards P.G. *Quantitative seismology*. Vol. 2, San-Francisco: Freeman and Company, 1980.
- [2] Afifi A.A., Azen S.P. *Statistical analysis a computer oriented approach*. New York: Academic Press, 1979.
- [3] Anderson T.W. *The statistical analysis of time series*. New York: John Wiley & Sons, 1971.
- [4] Brandt S. *Statistical and computational methods in data analysis*. Heidelberg: Heidelberg University, 1970.
- [5] Carrion Ph. *Inverse problems and tomography in acoustics and seismology*. Atlanta: University Press Atlanta, 1988.
- [6] Claerbout J.F. *Fundamental of geophysical data processing*. New York: McGraw Hill, 1976.
- [7] Claerbout J.F. *Imaging the Earth interior*. Oxford: Blackwell Scientific Publications, 1985.
- [8] Cramer H. *Mathematical methods of statistics*. Princeton: Univ. Press, 1946.
- [9] Johnson N., Lion F. *Statistics and the experiment design in techniques and science*. New York: Academic Press, 1976.
- [10] Hatton L., Worthington M.N., Makin J. *Seismic data processing*. Oxford: Blackwell Scientific Publications, 1986.
- [11] Herman G.T. *Image reconstruction from projections*. New York: Acad. Press, 1980.
- [12] *Seismic tomography*. Ed. G. Nolet. Boston: Reidel Publishing Company, 1987.
- [13] Parker R. *Inverse geophysical problems*. New York: D. Reidel Publishing Company, 1994.
- [14] Rao C.R. *Linear statistical inference and its application*. New York: Willey, 1965.

- [15] Sheriff R.E., Geldart L.P. *Exploration seismology*. Vol. 2. Cambridge: Cambridge University Press, 1983.
- [16] Tarantola A. *Inverse problem theory*. Amsterdam: Elsevier Scientific Publishing Company, 1987.
- [17] Yilmaz O. *Seismic data processing*. Tulsa: Press of Geophysical Exploration Society, 1988.

Contents

1	Introduction to the probability theory	1
1.1	The definition of probability	1
1.1.1	Set of elementary events	1
1.1.2	Axiomatic definition, probability space	5
1.1.3	Basic properties of probability	6
1.1.4	Relative-frequency definition	6
1.1.5	Classical definition	7
1.1.6	Geometrical definition	7
1.1.7	Conditional probabilities and independent events	7
1.1.8	Total probability	8
1.1.9	Bayes theorem	8
1.2	Distribution functions	9
1.2.1	Random variable	9
1.2.2	Definition of distribution function	9
1.2.3	Properties of the distribution function	10
1.2.4	The density function	11
1.2.5	The distribution and density of function of one random variable	12
1.2.6	Random vectors	13
1.2.7	Marginal and conditional distributions	14
1.2.8	Joint distribution and density functions	14
1.3	Expected values and moments of random variables	17
1.3.1	Mathematical expectation	17
1.3.2	Variance	18
1.3.3	Properties of variance and correlation coefficients	18
1.3.4	Quantiles	20
1.3.5	Characteristics of a distribution density function	21
1.4	Characteristic and generating functions	23
1.4.1	Characteristic functions	23
1.4.2	Moment generating function	25
1.4.3	Semi-invariants or cumulants	25
1.5	Limit theorems	26
1.5.1	Convergence in probability	26
1.5.2	Chebyshev inequality	26

1.5.3	Chebyshev theorem	26
1.5.4	Central-limit theorem	27
1.6	Discrete distribution functions	28
1.6.1	Binomial distribution	29
1.6.2	Poisson distribution	30
1.6.3	Geometrical distribution	31
1.7	Continuous distribution functions	31
1.7.1	Univariate normal distribution	32
1.7.2	Multivariate normal distribution	34
1.7.3	Uniform distribution	35
1.7.4	χ^2 — distribution	36
1.7.5	Student's t-distribution	37
1.7.6	Fisher distribution	38
1.7.7	Exponential distribution	39
1.7.8	Laplace distribution	40
1.7.9	Cauchy distribution	40
1.7.10	Logarithmic normal distribution	41
1.7.11	Significance of the normal distribution	41
1.7.12	Confidence interval	42
2	Information and entropy	45
2.1	Entropy of the set of discrete states of system	45
2.2	Entropy of the complex system	45
2.3	Shannon information (discrete case)	47
2.4	Entropy and information (continuous case)	48
2.5	Fisher information	52
3	Random functions and its properties	55
3.1	The expected value and variance	56
3.2	Properties of random functions	56
3.2.1	Stationarity	56
3.2.2	Ergodicity	57
3.3	Properties of autocorrelation functions	58
3.4	Action of a linear operator on a random function	58
3.5	Cross-correlation function	60
3.6	Wiener-Khinchin theorem and power spectrum	60
3.7	Examples of correlation functions and power spectrum	61
3.8	Estimation of numerical characteristics of random function	63
4	Elements of mathematical statistics	65
4.1	Sampling. Estimation.	65
4.2	Consistent estimator	66
4.3	Unbiased estimator	67
4.4	Rao-Kramer inequality. Efficient estimator	68

4.5 Sufficient estimate	70
4.6 Robust estimator	70
5 Models of measurement data	71
5.1 Additive models	71
5.2 Models of the quantitative interpretation	71
5.3 Regression model	73
5.4 Models of qualitative interpretation	74
5.5 Models of qualitative-quantitative interpretation	76
5.6 Random component models and its properties	77
5.6.1 The normal distribution	77
5.6.2 The random components is a stationary process	77
5.6.3 The random component is a non-correlated process	77
5.6.4 The random component is Markov process	78
5.6.5 The Laplace distribution	79
5.7 Model with random parameter	79
5.8 A priori information about field of parameters	79
6 Inverse problems: linear model	83
6.1 Method of least squares (MLS)	83
6.1.1 Properties of the MLS-estimation	84
6.2 Method of least squares: non-linear case	85
6.3 Method of least squares: orthogonal polynomials	86
6.4 Method of least squares with constraints	88
6.5 Maximum likelihood method (MLM)	90
6.6 Method of least modulus	93
6.7 Huber robust method	94
6.8 Andrews robust method	95
6.9 Bayes method and statistical regularization	97
6.9.1 The recursion algorithm of the maximum a posteriori probability method	99
6.9.2 Kalman filter	100
6.10 Singular analysis and method of least squares	102
6.10.1 Resolution matrix	106
6.11 The method of Backus and Gilbert	107
7 Inverse problems: nonlinear model	111
7.1 Newton-Le Came method	111
7.2 Method of least squares: nonlinear case	113
7.3 Method of least squares for non-stationary model	114
7.4 Interval estimation method	116
7.5 Genetic algorithms	117
7.5.1 Coding	118
7.5.2 Selection	118
7.5.3 Crossover	119

7.5.4	Mutation	119
7.5.5	Choice	120
8	Statistical criteria for choice of models	123
8.1	Parametric hypothesis test	123
8.2	Criterion of a posteriori probability ratio	124
8.3	The signal resolution problem	127
9	Ray geophysical tomography method	141
9.1	Definition of ray tomography	141
9.2	Radon transform	142
9.3	Algebraic methods	143
9.4	Slowness and absorption coefficient reconstruction	146
9.5	Radon transform for seismogram processing	148
9.6	Hilbert transform and analytical signal	150
10	Diffraction tomography method	153
10.1	Diffraction tomography in the scalar case	153
10.2	Born approximation for seismic trace model	157
10.3	Diffraction tomography for vector field	159
10.3.1	Model of measurement	159
10.3.2	Definition of tomography functionals	161
10.3.3	Tomography functional for scalar wave equation	162
10.3.4	Tomography functional for Lame equation	163
10.4	The reconstruction algorithm	168
10.5	Reconstruction tomography, Backus-Gilbert method	171
11	Multifactor analysis and processing of time series	177
A	About MATLAB	183
B	Statistical methods and transforms	195
B.1	Examples of numerical simulation of random values	195
B.1.1	Exercises	196
B.2	Creation of histogram	196
B.2.1	Exercises	196
B.3	Description of a random variable	196
B.3.1	Beta distribution	197
B.3.2	Binomial distribution	197
B.3.3	Cauchy distribution	198
B.3.4	Chi-square distribution	198
B.3.5	Exponential distribution	199
B.3.6	Fisher distribution	199
B.3.7	Gamma distribution	200
B.3.8	Gaussian distribution	200

B.3.9 Geometrical distribution	200
B.3.10 Hypergeometrical distribution	200
B.3.11 Laplace distribution	200
B.3.12 Logarithmic normal distribution	201
B.3.13 Poisson distribution	201
B.3.14 Snedecor distribution	201
B.3.15 Student distribution	201
B.3.16 Uniform distribution	201
B.3.17 Two dimensional normal distribution	201
B.3.18 Exercises	201
B.4 Computer simulation of the random values	202
B.4.1 Kolmogorov criterion	203
B.4.2 Exercises	203
B.5 Confidence intervals	203
B.5.1 Exercises	203
B.6 The time series	204
B.6.1 Correlated and non-correlating processes	204
B.6.2 Auto-correlation and cross-correlation functions	204
B.6.3 Exercises	205
B.7 Transforms	205
B.7.1 Fourier transform	205
B.7.2 Wavelets	205
B.7.3 Multifactor analysis	206
B.7.4 Cepstral transform	208
B.7.5 Exercises	210
C Direct and inverse problem solution	213
C.1 Computer simulation of gravitational attraction	213
C.1.1 Gravitational attraction of a sphere	213
C.1.2 Gravitational attraction of a horizontal cylinder	213
C.2 Computer simulation of magnetic induction	214
C.2.1 Magnetic induction of a dipole	214
C.2.2 Magnetic induction of a horizontal cylinder	215
C.3 Computer simulation of seismic field	215
C.3.1 Exercises	216
C.4 Deconvolution by the Wiener filter	216
C.4.1 Exercises	218
C.5 Quantitative interpretation	219
C.5.1 Exercises	222
C.6 Qualitative interpretation	223
C.6.1 Exercises	224
C.7 Diffraction tomography	226
C.7.1 Exercises	228