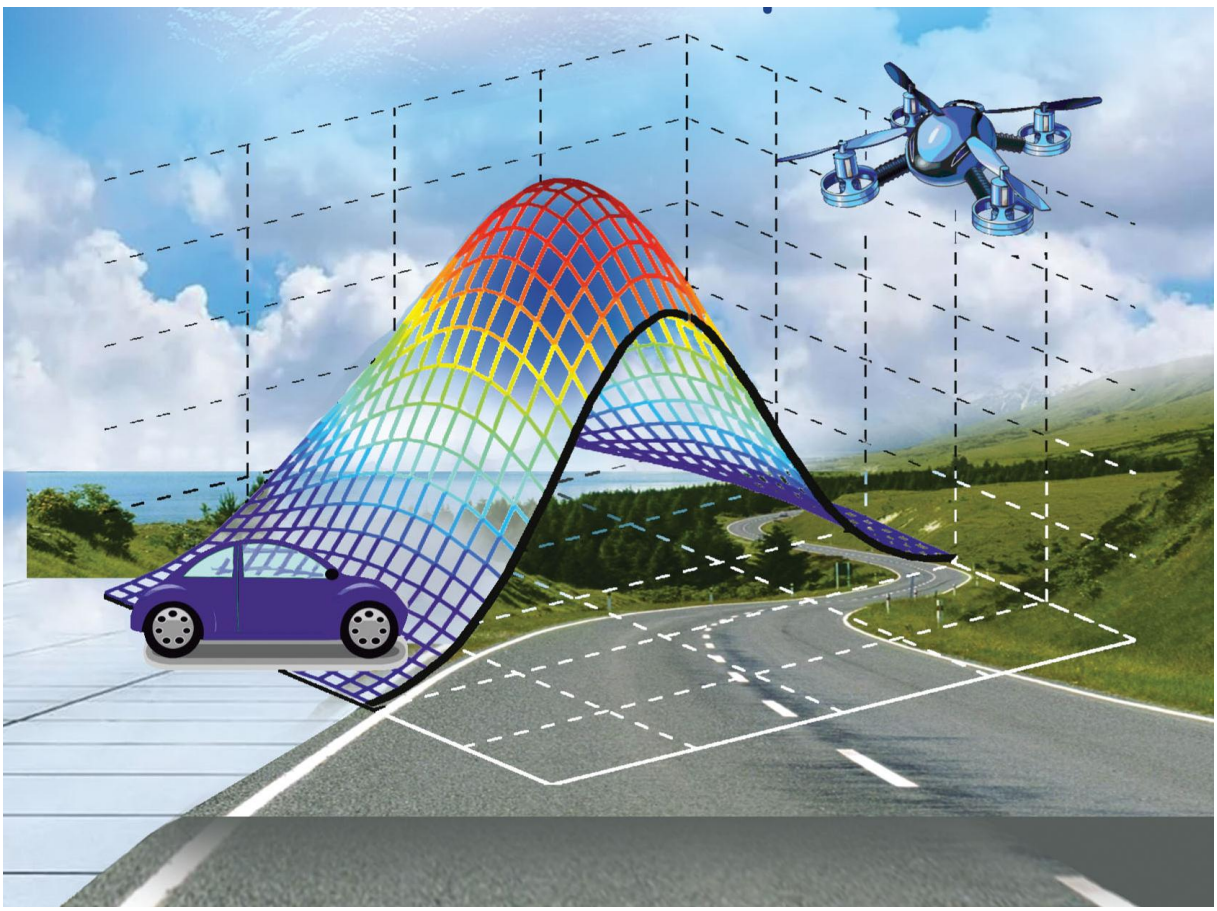
ITMO UNIVERSITY

O.A. Stepanov

ESTIMATION ALGORITHMS FOR NAVIGATION DATA PROCESSING



St. Petersburg
2021

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

О.А. Степанов

ESTIMATION ALGORITHMS FOR NAVIGATION DATA PROCESSING

УЧЕБНОЕ ПОСОБИЕ

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО
по направлению подготовки 24.03.02 Системы управления движением и навигация
в качестве учебного пособия для реализации основных профессиональных образовательных программ высшего образования бакалавриата

Санкт-Петербург
2021

Stepanov O.A., Estimation algorithms for navigation data processing –
St. Petersburg: ITMO University, 2021. – 174 pages.

Reviewer:
A.V. Nebylov. Doctor of Technical Sciences, Professor of Saint Petersburg State University of Aerospace Instrumentation, Director of the International Institute for Advanced Aerospace Technologies (IIAAT).

The general principles used in the measurement data processing aimed at estimation of unknown parameters and random sequences are described. Different approaches and methods for designing estimation algorithms are discussed. Considerable attention is given to the recursive discrete Kalman filter and its analog for continuous time. The tutorial is intended for students and graduate students involved in processing of redundant measurement data. The theory is supposed to help them solve problems related to navigation data processing.

УНИВЕРСИТЕТ ИТМО

ITMO University is the leading Russian university in the field of information and photonic technologies, one of the few Russian universities with the status of the national research university granted in 2009. Since 2013 ITMO University has been a participant of the Russian universities' competitiveness raising program among the world's leading academic centers known as «5 to 100». The objective of ITMO University is the establishment of a world-class research university being entrepreneurial in nature aimed at the internationalization of all fields of activity.

© ITMO University, 2021
© Concern CSRI Elektropribor, 2021
© O.A. Stepanov, 2021

CONTENTS

Introduction

The tutorial outlines the general principles used in measurement data processing aimed at estimation of unknown parameters. Depending on the level of a priori information of a statistical nature about the parameters being estimated and measurement errors, different approaches to the design of estimation algorithms are considered, including the least squares method, maximum likelihood method, and the Bayesian method. The relationship between the estimation algorithms designed within the framework of the considered approaches is analyzed. The methods and algorithms obtained for a constant vector are generalized with regard to the estimation of random sequences, the most important of which is the recursive discrete Kalman filter. In addition, the Kalman filter for continuous time is considered, and the transition from continuous algorithms to their discrete analogs is discussed. While presenting the material, we discuss not only the problem of algorithm design, but also the problem of accuracy analysis, which is very important for practical applications. In this regard, preference is given to algorithms resulting from the Bayesian approach. In addition to the estimate itself, such algorithms provide the possibility of generating current accuracy characteristics in the form of an appropriate covariance matrix of the estimation errors. The theory presented in this tutorial can be applied to solve problems of navigation data processing. It is in this field that the estimation algorithms considered in the tutorial are extensively used.

The tutorial is intended for students who take courses related to processing of redundant measurement data. The tutorial is based on the author's monographs [24, 25]. The main aim is to introduce the fundamentals of the theory and principles of designing algorithms for estimation of unknown parameters and random sequences.

The material is presented in eight chapters. Each chapter contains a theoretical part and illustrating examples, which include problem statements and algorithms used for their solutions.

At the end of each chapter, exercises are given for the learners to practice the solution of estimation problems and test questions, to revise the material.

The material is presented with the use of the mathematical apparatus which is actively used in the courses delivered at technical universities to first- and second-year students. The tutorial requires that the students have a good knowledge of the fundamentals of probability theory, matrix calculus, and the theory of ordinary differential equations. In this regard, for the readers' convenience, the appendices to the tutorial provide basic information on each of the above-mentioned subjects.

The author is grateful to his colleagues from ITMO University, the staff and postgraduate students of the Scientific and Educational Center of the Concern CSRI Elektropribor for reading the manuscript and their helpful remarks. Special thanks to T.P. Utkina and N.T. Zhigunova for the final editing of the text.

List of abbreviations

CDF	–	cumulative distribution function
CRLB	–	Cramer-Rao lower bound
GLSM	–	generalized least squares method
GNSS	–	global navigation satellite system
INS	–	inertial navigation system
LSM	–	least squares method
MLSM	–	modified least squares method
PDF	–	probability density function
RMS	–	root-mean-square
WLSM	–	weighted least squares method

Terms and definitions

Measurements are outputs of different sensors. With regard to navigation problems, the latter are accelerometers, gyroscopes. In this tutorial, the term ‘measurements’ is also used to mean coordinates and/or components of the vehicle velocity and/or parameters of its attitude generated by inertial navigation systems, dead-reckoning system, and GNSS equipment. In addition, measurements can be generated by various kinds of additional aids (correcting equipment), for example, barometers, rangefinders, beacons, sensors of different geophysical fields, etc.

For measurements, we usually use notation $y_i = y(t_i)\ \ i = 1, 2....$ for discrete time or $y(t)$ for continuous time.

Like all parameters defined below, measurements can be both scalar and vector.

Measurement errors are differences between true and measured values.

For measurement errors, we usually use notation $v_i,\ \ i = 1, 2....$ for discrete time or $v(t)$ for continuous time.

Parameters to be estimated are unknown parameters that we need to find (estimate, calculate) using the accumulated measurements.

For the parameters to be estimated, we usually use notation $x_i,\ \ i = 1, 2....$ for discrete time or $x(t)$ for continuous time.

A measurement model is a mathematical relation $y_i = f(x_i, v_i)$ that determines how the measurements depend on the parameters to be estimated and measurement errors.

In what follows, we use linear and nonlinear models or functions $\varphi(x)$, i.e. the ones that satisfy or do not satisfy the superposition principle:

$$\varphi(ax_1 + bx_2) = a\varphi(x_1) + b\varphi(x_2),$$

where *a* and *b* are arbitrary constants .

A model of parameters to be estimated is mathematical relations in the form of difference or differential equations that determine how these parameters vary in time.

Further, models given by linear equations are used, as a rule.

An estimation problem is a problem, which is aimed to find unknown parameters (obtain an estimate) using measurements and information about models for measurements and the parameters to be estimated.

A parameter estimate is the parameter value obtained from the solution of the estimation problem.

For the parameter estimates, we usually use notation $\hat{x}_i,\ i=1,2....$ for discrete time or $\hat{x}(t)$ for continuous time.

An estimation algorithm is a sequence of calculations and logical operations performed with measurements aimed to obtain estimates of the sought parameters.

Estimation errors are the differences between the true values of the parameters and their estimates.

For the parameter estimate errors, we usually use notation

$\varepsilon_i = x_i - \hat{x}_i,\quad i=1,2....$ for discrete time or $\varepsilon(t) = x(t) - \hat{x}(t)$ for continuous time.

Estimation criterion is a function or functional which must be minimized or maximized when finding the parameters to be estimated.

1. Problems of estimating constant parameters in navigation data processing: Statements and examples

Consider some examples of typical linear and nonlinear estimation problems for constant (time-invariant) unknown parameters that are often solved in navigation data processing.

1.1. Estimation of polynomial coefficients

Let the aircraft height be unknown and measurements be made at times t_i, $i=\overline{1.m}$ using an onboard sensor. Assume that the unknown height is constant during the measurement interval and is equal to h. Then, introducing $x=h$ and assuming that the measurements include errors v_i, $i=\overline{1.m}$, h, the problem of the height determination can be reduced to estimation of unknown constant value x by a set of noisy measurements:

$$y_i = x + v_i,\ i=\overline{1.m}. \tag{1.1}$$

Introducing unity column H, $H^{\mathrm{T}}=[1,1....1]$, and m-dimensional vectors $y=(y_1,...y_m)^{\mathrm{T}}$ and $v=(v_1,...v_m)^{\mathrm{T}}$, we can write measurements (1.1) as

$$y = Hx + v. \tag{1.2}$$

In this equation, the measurements linearly depend on the parameter being estimated. Thus we can speak of the linear character of measurements. A more complicated model can be introduced to describe the height variations during the measurement interval. Particularly, if the height changes by the linear law (representing **a linear trend**), the measurements can be presented as

$$y_i = x_0 + Vt_i + v_i,\ i=\overline{1.m}, \tag{1.3}$$

where x_0, V are the initial height and vertical velocity being estimated, considered to be constant; $t_i=(i-1)\Delta t$ are the time instants since the beginning of the observation; Δt is the interval between measurements.

Measurements (1.3) can also be presented using (1.2) by introducing the vector being estimated and matrix H:

$$x=(x_1,x_2)^{\mathrm{T}}=(x_0,V)^{\mathrm{T}};\ H^{\mathrm{T}}=\begin{bmatrix}1 & 1 & . & . & 1\\ t_1 & t_2 & . & . & t_m\end{bmatrix}. \tag{1.4}$$

Representation in the form of (1.2) can be derived also for a more general case, wherein the height variation is described by the $(n-1)$-th order polynomial, then the problem can be mathematically reduced to estimation of the polynomial coefficients

$x = (x_1, x_2, ..x_n)^{\mathrm{T}}$ by measurements

$$y_i = x_1 + x_2 t_i + x_3 t_i^2 + .. + x_n t_i^{n-1} + v_i, \; i = \overline{1.m}. \qquad (1.5)$$

The problem of preprocessing the sensor measurements to decrease the noise level is often reduced to this formulation.

The need to estimate polynomial coefficients often arises in the so-called **problem of sensor calibration or calibration problem**. It consists in comparing the sensor outputs with the reference value of the parameter being measured or more accurate measurements. The model of sensor errors should be constructed using the difference between sensor outputs and the reference value of parameter being measured or more accurate measurements. In the absence of the reference, this model makes it possible to improve the accuracy of measurements. Examples of measurement errors containing a constant component (a) and **a quadratic trend** (second-order polynomial) (b) are shown in Fig. 1.1.

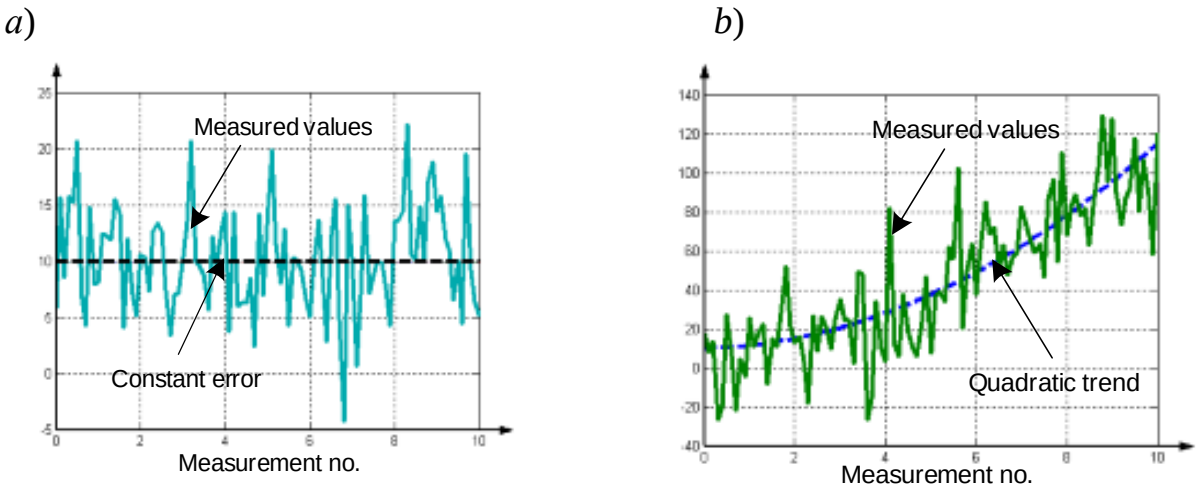


Fig. 1.1. Examples of measurement errors containing a constant component (*a*) and a quadratic trend, $t = 10\ s$, $\Delta t = 0.1\ s$ (*b*).

Other arguments can be used in (1.5) instead of time. They can be other physical quantities, such as temperature. As is known, the errors of high-accuracy sensors significantly depend on temperature. A special thermal stabilization system is used to reduce these errors. Preliminary description of the temperature-error dependence, for example, by using a polynomial model, sometimes makes it possible to lower requirements on expensive thermal stabilization systems.

1.2. Initial alignment of an inertial navigation system. A simple case study

As is known, neglecting the errors of the inertial sensor (accelerometers and gyroscopes), in the simplest case, the velocity error for the eastern channel of an inertial navigation system (INS) can be approximately described with the following formula [7, 15]:

$$\Delta V_E(t_i) = -\alpha(0)\sqrt{gR}\sin\omega_{sch}t_i + \Delta V_E(0)\cos\omega_{sch}t_i,\ i=\overline{1.m},$$

where $\alpha(0)$, $\Delta V_E(0)$ are the initial vertical error (angle between the true and INS horizon planes) and the initial velocity error; g is the gravity acceleration; R is the Earth's radius; $\omega_{sch} = \sqrt{\frac{g}{R}}$ is the frequency of Schuler period: $T_{sch} = 2\pi\sqrt{\frac{R}{g}} \approx 84$ min.

The INS carrier is assumed motionless, i.e., its velocity is zero. Then, the velocity measured by the INS is actually the INS error. Taking these measurements at discrete times, we can write:

$$y(t_i) = -\alpha(0)\sqrt{gR}\sin\omega_{sch}t_i + \Delta V_E(0)\cos\omega_{sch}t_i + v_i, \qquad (1.6)$$

where v_i is the readout error.

Using the measurement set $y(t_i)$, $i=\overline{1.m}$, we can estimate the initial vertical error and the velocity error. Actually, this is the problem of initial alignment of INS in its simplest interpretation. Introducing the vector of the parameters being estimated and matrix *H*:

$$x^{\mathrm{T}} = (\alpha(0),\ \Delta V_E(0));\qquad (1.7)$$

$$H^{\mathrm{T}} = \begin{bmatrix} -\sqrt{gR}\sin\omega_{sch}t_1, & -\sqrt{gR}\sin\omega_{sch}t_2, & . & . & -\sqrt{gR}\sin\omega_{sch}t_m \\ \cos\omega_{sch}t_1, & \cos\omega_{sch}t_2, & . & . & \cos\omega_{sch}t_m \end{bmatrix}, \qquad (1.8)$$

it is also easy to present these measurements in the form similar to (1.2).

Assume that the problem is solved over a short time interval, as compared with the Schuler period. Expanding functions $\sin\omega_{sch}t_i$ and $\cos\omega_{sch}t_i$ in Taylor series, keeping only the first-order members, this problem can be easily reduced to estimation of polynomial coefficients. In particular, if we assume that $\sin\omega_{sch}t_i \approx \frac{2\pi}{T_{sch}}t_i$, $\cos\omega_{sch}t_i \approx 1$ with $t_i << T_{sch}$, the measurements can be approximately written using the first-order polynomial:

$$y(t_i) \approx -\alpha(0)\sqrt{gR}\frac{2\pi}{T_{sch}}t_i + \Delta V_E(0) + v_i. \qquad (1.9)$$

We also deal with estimation of the second- and third-order polynomials in the problem when the INS errors in the velocity and displacement components over interval ΔT are approximately described by polynomial models.

1.3. Statement of a linear estimation problem

All the above problems can be reduced to the following **common statement of a linear estimation problem**.

We have an unknown constant n-dimensional vector $x=(x_1,...x_n)^{\mathrm{T}}$

$$\dot{x}=0 \tag{1.10}$$

and an m-dimensional measurement vector $y=(y_1,...y_m)^{\mathrm{T}}$

$$y=Hx+v, \tag{1.11}$$

where H is an $m\times n$-dimensional matrix, and $v=(v_1,...v_m)^{\mathrm{T}}$ is an m-dimensional vector describing the measurement errors.

It is required to determine estimate $\hat{x}(y)$ of unknown vector x using measurement (1.11). In estimation problems, this vector is referred to as the **state vector**. Notation $\dot{x}=\frac{dx}{dt}=0$ means that vector x is constant (time-invariant), that is, the solution to this simplest differential equation is $x=\mathrm{const}$. This notation is used to relate the problem of constant vector estimation considered in this chapter to more general problems of estimating the variable vector whose behavior with time can be described using differential or difference equations (see Chapters 7 and 8).

In solving the formulated problem, not only the estimate calculation algorithm using the available measurements proves to be important but it is also very important to be able to quantify **the estimation error** defined as

$$\varepsilon(y)=x-\hat{x}(y), \tag{1.12}$$

i.e., quantify the accuracy of the estimate generated by the proposed algorithm. This is especially relevant in solving problems of navigation data processing. Thus, two important subproblems can be distinguished within the considered problem: **algorithm design**, i.e., obtaining a detailed procedure for calculating estimates $\hat{x}(y)$, and **accuracy analysis,** which consists in studying the properties of estimation error (1.12), the error levels, in particular.

1.4. Estimation of time delay

Further, consider some examples of nonlinear estimation problems. In practice, we often deal with the problem of estimating a time delay or shift of a measured sample of any signal relative to another reference sample of this signal. Let us explain the essence of this problem. Assume that we have a known nonlinear function $s(t)$ of scalar argument and measurements of the type

$$y_i=s(t_i+\tau)+v_i,\ i=\overline{1.m}, \tag{1.13}$$

where v_i, $i=\overline{1.m}$ are the measurement errors at points $t_i+\tau$, and τ is an unknown constant.

It is required to estimate τ, knowing $s(t)$ and having the values y_i, t_i, $i=\overline{1.m}$, i.e., to determine the time delay (shift) of the measured sample y_i, $i=\overline{1.m}$ with respect to

the other reference $s(t_i+\tau)$, $i=\overline{1.m}$.

If $s(t)$ is a harmonic oscillation, the measurements similar to (1.13) can be specified as

$$y_i = A\sin(\omega t_i+\phi_0)+v_i, i=\overline{1.m}, \qquad (1.14)$$

where A is the amplitude, $\omega=2\pi f$ is the circular frequency, and ϕ_0 is the phase.

If we know the amplitudes and the frequencies, we have the **phase estimation problem**.

In formulas (1.13), (1.14), the measurements depend nonlinearly on the parameter being estimated. Thus, we can speak about nonlinear measurements and a nonlinear estimation problem.

In general form, in **nonlinear estimation problems**, measurements can be written as

$$y = s(x)+v, \qquad (1.15)$$

where x, v are n- and m-dimensional vectors; $s(x)=(s_1(x),..s_m(x))^{\mathrm{T}}$ is an m-dimensional vector-function.

For harmonic oscillation, with known amplitudes and frequencies, formula (1.14) is reduced to (1.15) if $x=\phi_0$ and $s(x)=(A\sin(\omega t_1+x),....A\sin(\omega t_m+x))^{\mathrm{T}}$. The essence of this problem is clear from Fig. 1.2.

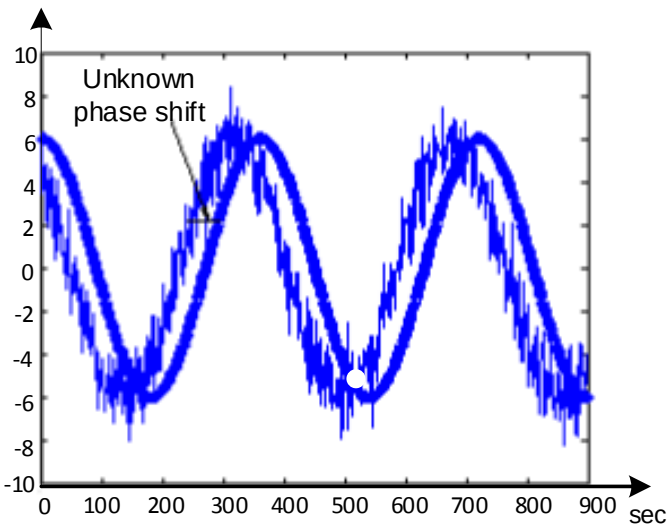


Fig. 1.2. Phase determination problem

If it is required to solve **a frequency estimation problem** in the case when the phase and the amplitude are known, then $x=\omega$ and $s(\bullet)=(A\sin(xt_1+\varphi_0),....A\sin(xt_m+\varphi_0))^{\mathrm{T}}$. If both the signal phase and the frequency need to be estimated, we should introduce $x^{\mathrm{T}}=(\phi_0,\omega)$ and $s(\bullet)=(A\sin(x_2t_1+x_1),....A\sin(x_2t_m+x_1))^{\mathrm{T}}$ into formula (1.15). Thus, we have the **phase**

lock problem, which is often solved in designing various measuring devices. In general case, in the **problem of harmonic signal parameter estimation,** all three parameters—amplitude, frequency, and phase—can be unknown, then the vector of parameters being estimated will be three-dimensional $x^{\mathrm{T}} = (A, \phi_0, \omega)$. It can be easily seen that with the known frequency and phase, the measurements linearly depend on the amplitude, as the result, we have a linear estimation problem (see Exercise 1.2.).

The majority of radionavigation systems are also based on time delay estimation, i.e., estimation of the shift between the measured sample and the reference one. In particular, the possibility of determining coordinates in global navigation satellite systems (GNSS) is provided by simultaneous measurements of several delays conditioned by the finite time of radio wave propagation between the satellites and the user. These delays are estimated by comparing the envelopes extracted from satellite signals with their copies generated in the user receivers [10, 14].

The so-called map-aided navigation problem is also reduced to a similar formulation [3, 23, 26]. The idea of the method is to determine the vehicle position by comparing the measured sample of some geophysical parameter, such as terrain features, with the reference sample of these parameters computed using the earlier constructed map. Then measurements (1.13) will correspond to the one-dimensional version of the problem, if the time argument is substituted with the spatial one, and function $s(\bullet)$ agrees with the map terrain variation from this spatial coordinate (Fig. 1.3). Thus, it is necessary to estimate the spatial shift of the measured sample with respect to the map.

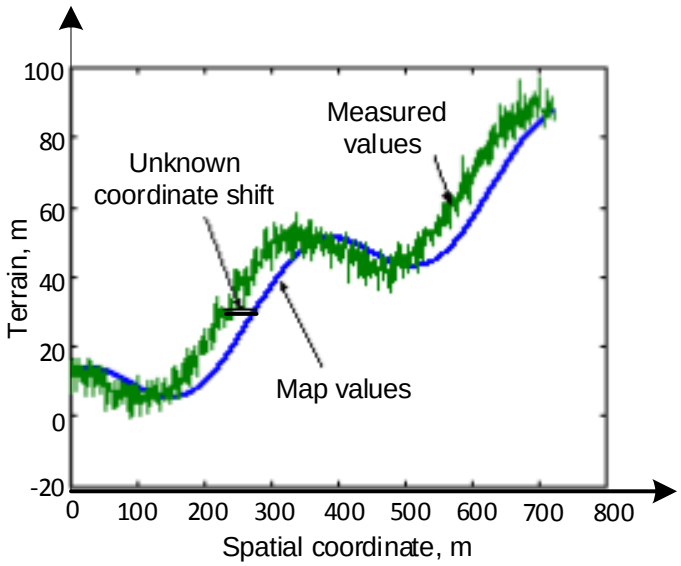


Fig. 1.3. One-dimensional problem of map-aided navigation

1.5. Position determination by ranging to beacons

As an example of a nonlinear estimation problem, we can consider the problem of

two-dimensional (2D) position determination of a vehicle using ranges (distances) to the beacons with the known coordinates (Fig. 1.4). In this case, $x=(x_1,x_2)^{\mathrm{T}}$, and the measurements are

$$y_i = s_i(x) + v_i = \sqrt{\left(x_1^i - x_1\right)^2 + \left(x_2^i - x_2\right)^2} + v_i,\ i=\overline{1.m}, \quad (1.16)$$

where x_1, x_2 are the coordinates of the vehicle; x_1^i, x_2^i, $i=\overline{1.m}$ are the coordinates of the beacons

$$s_i(x) = D_i(x) = \sqrt{\left(x_1^i - x_1\right)^2 + \left(x_2^i - x_2\right)^2}.$$

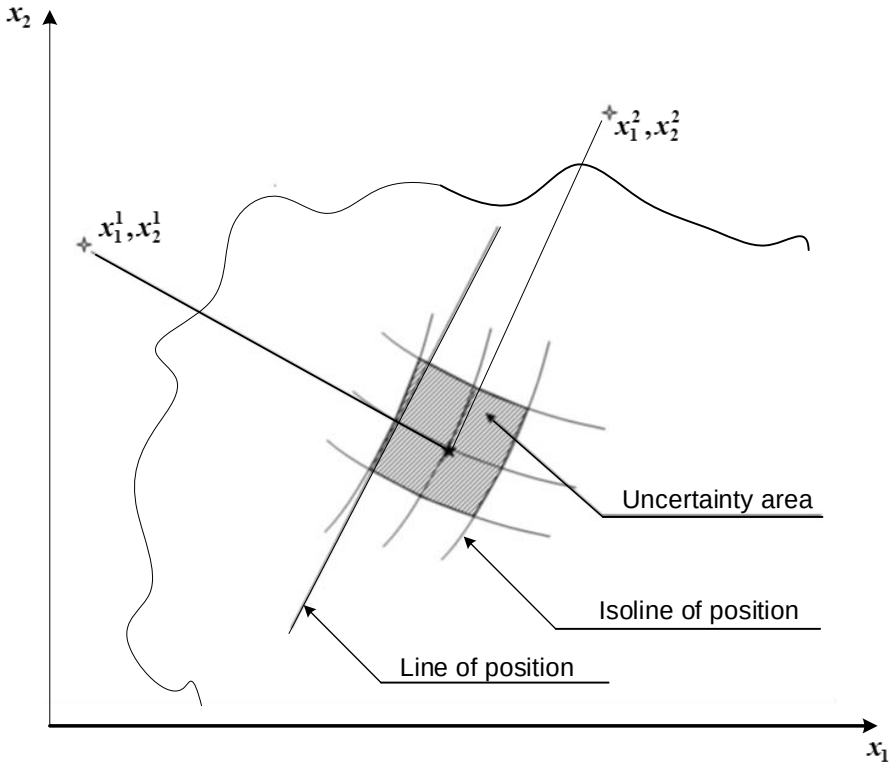


Fig. 1.4. 2D position determination using range measurements to two beacons with known coordinates

The idea of position determination by a ranging method is rather simple. With a correct range measurement to a beacon, the vehicle position can be estimated to the accuracy of its location on a circle with the radius equal to the measured range. The lines connecting the points of equal values of the parameters being measured are called **isolines (contour lines) of position** in navigation. In this case, they are circles. Having two exact ranges, vehicle coordinates can be determined as one of the possible crossing points of these isolines. Since the measurements contain errors, bars will be formed instead of lines, enclosed between the circles equal to the maximum and minimum possible ranges. With two measurements available, the vehicle coordinates are likely to be arranged inside a figure formed by the crossing of two bars. If there are more

measurements, there arises the problem of position determination with maximum accuracy.

If all the measurements contain the same constant error component denoted by x_3, the following equation can be written:

$$y_i = s_i(x) + v_i = \sqrt{\left(x_1^i - x_1\right)^2 + \left(x_2^i - x_2\right)^2} + x_3 + v_i,\ i = \overline{1.m}. \qquad (1.17)$$

This error component is usually called a systematic error component or just **a systematic error**. Inclusion of unknown parameters into the state vector being estimated is called **state vector augmentation**.

1.6. Position and velocity determination by satellite data

The previous problem can be easily generalized to three-dimensional (3D) position determination. 3D measurements are used in GNSS position determination. Having measurements of the delays between the received signals and the signals generated in user equipment and knowing the radio wave propagation speed, it is possible to present the measured ranges to satellites as follows:

$$\rho_i = \sqrt{\left(x_1^i - x_1\right)^2 + \left(x_2^i - x_2\right)^2 + \left(x_3^i - x_3\right)^2} + c\Delta t + \varepsilon_i, \qquad (1.18)$$

where x_1, x_2, x_3 are the unknown user coordinates at the moment of signal reception in the Cartesian geocentric coordinate system; x_j^i, $j = 1,2,3$ are the coordinates of the i-th satellite in the same coordinate system delivered to the user in navigation message; Δt is the user clock error; ε_i is the total measurement error; c is the speed of light [10, 14, 15].

The orbits and the arrangement of satellites are selected so that almost at any point on the Earth at any time, measurements from at least four satellites are available. Along with satellites, GNSS include ground control stations which determine satellite motion parameters (position and velocity) (Fig. 1.5).

The components of the estimated vector $x = (x_1, x_2, x_3, \Delta t)^{\mathrm{T}}$ are the user position and the clock error.

GNSS is also used to determine the user velocity components. This is done by measuring the Doppler shifts of carrier frequency (f_j^{dop}) due to mutual displacements of the satellite and the user.

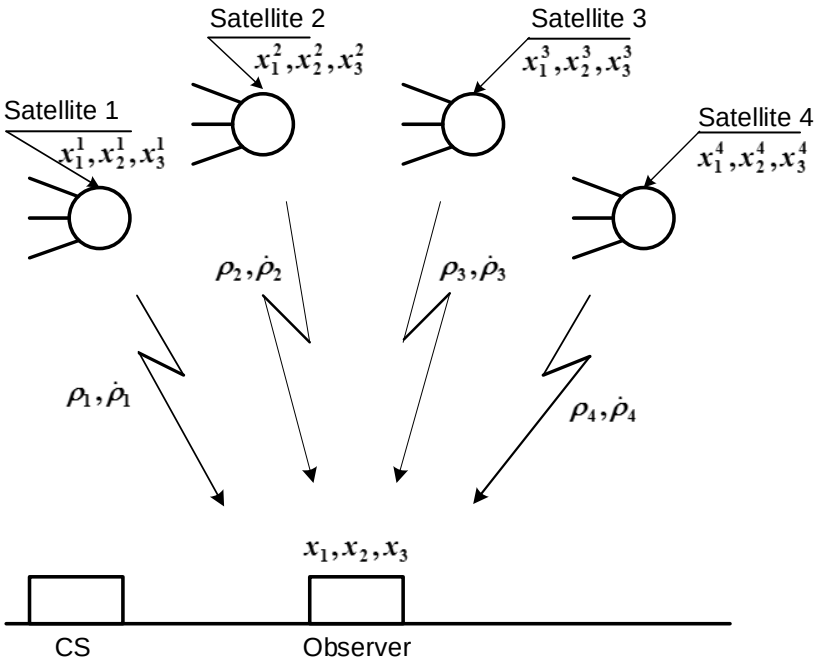


Fig.1.5. Determination of navigation parameters by GNSS data:
CS are control stations determining the satellite motion parameters

Using these measurements, we can write:

$$\dot{\rho}_i = \frac{(x_1^i - x_1)(\dot{x}_1^i - \dot{x}_1) + (x_2^i - x_2)(\dot{x}_2^i - \dot{x}_2) + (x_3^i - x_3)(\dot{x}_3^i - \dot{x}_3)}{\sqrt{\left(x_1^i - x_1\right)^2 + \left(x_2^i - x_2\right)^2 + \left(x_3^i - x_3\right)^2}} + c\Delta\dot{t} + \tilde{\varepsilon}_i \,, \quad (1.19)$$

where $\dot{x}_j; \dot{x}_j^i$, $j = 1,2,3$ are the velocity components of the user and the i-th satellite; $\Delta\dot{t}$ is the user clock drift error; $\tilde{\varepsilon}_i$ is the total error of Doppler measurements. Here, we estimate the vector including the user velocity components $\dot{x}_1, \dot{x}_2, \dot{x}_3$ and its clock drift error.

As seen from the given equations, errors in the range and the rate of its change caused by the errors of the user clock and its offset drift are systematic errors. The presence of these errors explain the terms ‘pseudorange’ and ‘pseudovelocity’ used for the parameters being measured (1.18), (1.19) and the fact that at least four satellites are needed simultaneously to get the navigation solution.

1.7. Statement of a nonlinear estimation problem and its linearization

All the problems considered above can be reduced to the following **common statement of the nonlinear estimation problem**.

Let the unknown constant n-dimensional vector $x = (x_1, \ldots x_n)^{\mathrm{T}}$ be given by

$$\dot{x} = 0, \quad (1.20)$$

and also, we have an m-dimensional measurement vector $y = (y_1, \ldots y_m)^{\mathrm{T}}$

$$y = s(x) + v, \tag{1.21}$$

where $s(\bullet) = (s_1(x),..s_m(x))^{\mathrm{T}}$ is a known m-dimensional nonlinear function, and $v = (v_1,...v_m)^{\mathrm{T}}$ is the m-dimensional vector.

It is required to find estimate $\hat{x}(y)$ of unknown vector x using measurement (1.21).

Here, as in the linear case, we should distinguish between the problems of algorithm design and accuracy analysis, the latter consisting in studying the properties of estimation errors (1.12). Clearly, this statement includes the linear case as well, since substituting $s(x)$ for Hx in (1.21), we come to the problem (1.10), (1.11).

Further, in the discussion of the approaches to algorithm design and accuracy analysis, we will consider both linear and nonlinear estimation problems. However, the emphasis will be put on linear problems. This is a consequence conditioned by the fact that many nonlinear problems can be reduced to linear statements without an essential loss in accuracy. It is made possible due to **linearization of function** $s(\bullet) = s(.) = (s_1(x),..s_m(x))^{\mathrm{T}}$, i.e., by its approximate presentation as Taylor series, keeping only the first-order members:

$$s(x) \approx s(x^l) + \left.\frac{ds}{dx^{\mathrm{T}}}\right|_{x=x^l} (x - x^l) = s(x^l) + H(x^l)(x - x^l), \tag{1.22}$$

where x^l is the linearization point,

$$H(x^l)\left.\frac{ds}{dx^{\mathrm{T}}}\right|_{x=x^l}. \tag{1.23}$$

In designing estimation algorithms based on linearization, it is convenient to use a new vector given by

$$\delta x = (x - x^l) \tag{1.24}$$

as the vector to be estimated.

Rearrange the known summands to the left-hand part of the equation and introduce the notation

$$\tilde{y}(x^l) \stackrel{\Delta}{=} y - s(x^l). \tag{1.25}$$

Then, the following approximate equation can be written:

$$\tilde{y}(x^l) \approx \left.\frac{ds}{dx^{\mathrm{T}}}\right|_{x=x^l} (x^l)\delta x + v = H(x^l)\delta x + v, \tag{1.26}$$

where $\tilde{y}(x^l)$ is measurement (1.25) calculated from the initial measurement by subtracting the known values.

Thus, the initial nonlinear problem (1.20), (1.21) is approximately reduced to the linear problem of δx estimation by measurements (1.26).

Clearly, the linearized description will be permissible only in the vicinity of the linearization point. The accuracy of this representation for the scalar case can be

approximately estimated by

$$\delta = \frac{d^2 s}{dx^2}(x - x^l)^2, \tag{1.27}$$

which determines the level of the second-order summands in Taylor expansion and depends on the second derivative $\frac{d^2 s}{dx^2}$, on the one hand, and on the expectable possible deviations of real unknown values of the estimated parameter from the linearization point, i.e., from the difference $(x - x^l)$, on the other hand.

The feasibility of linearized description in estimation problems can be assessed by comparing the expected measurement errors v with the expected values

$$\delta = \frac{d^2 s}{dx^2}(x - x^l)^2.$$

Let us illustrate this by the example of the 2D position determination problem by ranges to beacons using measurements (1.16) (see section 1.5). Using the described linearization procedure, the following can be written:

$$\tilde{y}(x^l) \approx H(x^l)\delta x + v, \tag{1.28}$$

where

$$\tilde{y}_i(x^l) \stackrel{\Delta}{=} y_i - D_i(x^l),\ i = \overline{1.m}; \tag{1.29}$$

$$D_i(x_l) = \sqrt{\left(x_1^i - x_1^l\right)^2 + \left(x_2^i - x_2^l\right)^2}; \tag{1.30}$$

$$\begin{aligned} H(x^l) &= \begin{bmatrix} (x_1^l - x_1^1)/D_1(x_л) & (x_2^l - x_2^1)/D_1(x^l) \\ (x_1^l - x_1^2)/D_2(x_л) & (x_2^l - x_2^2)/D_2(x^l) \\ \cdot & \cdot \\ (x_1^l - x_1^m)/D_m(x_л) & (x_2^l - x_2^m)/D_m(x^l) \end{bmatrix} = \\ &= -\begin{bmatrix} \sin B_1(x^l) & \cos B_1(x^l) \\ \sin B_2(x^l) & \cos B_2(x^l) \\ \cdot & \cdot \\ \sin B_m(x^l) & \cos B_m(x^l) \end{bmatrix} \end{aligned} \tag{1.31}$$

is an $m \times 2$-dimensional matrix.

In these equations, $B_i(x^l)$ is the angle measured from axis ox_2, which determines the vector orientation from $x^l = (x_1^l, x_2^l)^{\text{т}}$ to the beacon. The vector determines the direction of the gradient for the navigation parameter being measured (here, range), i.e., the direction of its greatest change. The components of $H_i^{\text{т}}(x^l)$ correspond to the derivatives following the directions along coordinates ox_1 and ox_2. In this example, linearization consists in substituting the isolines in the form of circles (see above) by

straight lines called the **lines of position** (see Fig. 1.4).

For the parameter (1.27), the following is true in this example:

$$\delta = \frac{d^2D}{dx_1^{\,2}}(x_1 - x_1^l)^2 + \frac{d^2D}{dx_2^{\,2}}(x_2 - x_2^l)^2 + \frac{d^2D}{dx_1 dx_2}(x_1 - x_1^l)(x_2 - x_2^l),$$

which can be written as follows with $x_1 - x_1^l \approx x_2 - x_2^l \approx \Delta$:

$$\delta = \left(\frac{d^2D}{dx_1^{\,2}} + \frac{d^2D}{dx_2^{\,2}} + \frac{d^2D}{dx_1 dx_2} \right) \Delta^2.$$

Assuming, for simplicity, that one coordinate is known and the direction of the straight line connecting the beacon position with the linearization point agrees with the direction of the unknown coordinate, we can get the following for δ:

$$\delta = \frac{d^2D}{dx_1^{\,2}}\Delta^2 = \frac{\Delta^2}{D} = \frac{\Delta}{D}\Delta. \qquad (1.32)$$

From this it follows that the error in the linearized representation depends on the ratio between the possible coordinate error in selecting the linearization point and the range to the beacon. The value of δ then should be compared to the range measurement error. Estimation of the linearization error for the positioning problem by satellite data with the orbit altitude of 20,000 km yields that with the linearization point set accurate to 1–10 km, $\delta \approx (0.5-5)\,m$. This value is comparable with the range measurement error in real satellite systems.

Exercises

Exercise 1.1. We have a set of measurements

$$y_i = x_1 + x_2 + v_i,\ i = \overline{1.m}. \qquad (1)$$

Formulate the problem in the form (1.10), (1.11) for the case when only x_1 should be found using measurements (1), and the sum $\varepsilon_i = x_2 + v_i$ is treated as an error. Repeat the procedure to estimate x_1 and x_2.

Exercise 1.2. Let the following measurements be set at discrete times t_i, $i = \overline{1.m}$

$$y_i = A\sin(\omega t_i + \phi_0) + v_i, i = \overline{1.m},$$

where A is the amplitude; $\omega = 2\pi f$ is the circular frequency; ϕ_0 is the phase.

Considering the frequency and the phase to be known, formulate the amplitude estimation problem in the form (1.10), (1.11).

Exercise 1.3. Let the measurements $y_i = x_1 + A\sin\omega t_i + v_i$ be given at discrete times t_i, $i = \overline{1.m}$, where circular frequency $\omega = 2\pi f$ is considered known, and x_1 and A are unknown magnitudes.

Formulate the estimation problem in the form (1.10), (1.11) for the case when only

the amplitude needs to be found, and the sum $\varepsilon_i = x_1 + v_i$ is treated as an error. Repeat the procedure to estimate x_1 and A.

Exercise 1.4. Let the measurements be given as those in Exercise 1.2, and the phase should be found with the known frequency and amplitude. Write the linearized problem statement, taking $\overline{\phi}_0$ as the linearization point. Repeat the procedure if the frequency should be found with the known phase and amplitude.

Exercise 1.5. Tracking problem (trajectory measurement problem) (Fig. 1.6). Let there be a point with known 2D coordinate x_1^*, x_2^* from which ranges ρ_i and bearings θ_i to the vehicle moving rectilinearly with a constant speed can be measured with errors $\delta\rho_i$ and $\delta\theta_i$, $i = \overline{1.m}$ at discrete times. Formulate the estimation problem of the vehicle position and velocity components assuming the availability of ranges (*a*); bearings (*b*); or both, for integrated processing (*c*).

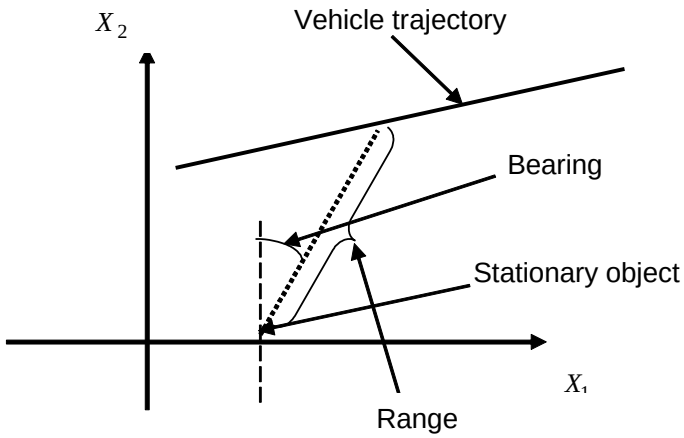


Fig. 1.6. Problem of vehicle coordinates determination by ranges and bearings

Exercise 1.6. Formulate the vehicle tracking problem assuming that the conditions of Exercise 1.5 are met, and the range and bearing errors also include constant components to be estimated.

Exercise 1.7. Formulate two previous problems in linearized form as an integrated problem of processing ranges and bearings to estimate velocity components and coordinates.

Test questions

1. Explain the essence and provide a mathematical statement of the following linear estimation problems: a constant scalar value; coefficients of linear and quadratic trends; one-dimensional position and speed with the vehicle uniform motion;

polynomial coefficients.

2. Explain the essence and provide a mathematical statement of the estimation problem for the simplest version of INS alignment.
3. Explain the essence and formulate a mathematical problem statement of estimating the shift of a measured sample of any signal relative to another reference sample of this signal. Illustrate it by the example of estimating the phase of a harmonic signal. Why is it a nonlinear problem?
4. Explain the essence and provide a mathematical statement of estimating the frequency of a harmonic signal and phase lock problems.
5. Explain the essence and provide mathematical statements of nonlinear problems of two-dimensional and three-dimensional vehicle position determination by ranges to beacons with and without constant components of measurement errors. What does the state vector augmentation mean?
6. Formulate the problem of constant vector estimation by noisy measurements and illustrate it by examples. Discuss the features of algorithm design and accuracy analysis problems.
7. Explain the essence of the linearization procedure and discuss the corresponding linearized problem by the example of the vehicle tracking problem, i.e., the problem of its position and velocity components determination by using the measurements of ranges and bearings.

2. Solution of estimation problems using the deterministic approach. Least squares method

This section considers the so-called **deterministic approach** in which the solution of estimation problems does not involve an assumption that the unknown vectors x and ν of measurement errors are random vectors, and, therefore, there is no need to use any a priori statistical information.

2.1. Fundamentals and statement of the estimation problem using the least squares method

A feature of the approach under consideration is that the problem of the algorithm design, i.e., obtaining a procedure for calculating the estimate of unknown vector x using measurements y, is based on the choice of the values that minimize the criterion characterizing the measure of closeness between the measured and calculated values $s(x)$ or Hx. In its simplest version, such a criterion can be introduced as the function

$$J^{\mathrm{LSM}}(x) = (y - s(x))^{\mathrm{T}}(y - s(x)) = \sum_{i=1}^{m}(y_i - s_i(x))^2 . \qquad (2.1)$$

The differences $\mu_i = y_i - s_i(x)$ are usually called **measurement residuals.**

The method based on the minimization of the type (2.1) criterion is called **the least squares method (LSM)**. In the subsequent discussion, the algorithms based on minimization of the type (2.1) criteria are also called LSM algorithms.

The criterion (2.1) and the corresponding estimate

$$\hat{x}^{\mathrm{LSM}}(y) = \arg\min_{x}(y - s(x))^{\mathrm{T}}(y - s(x)) \qquad (2.2)$$

have a clear meaning: to select such a value of the parameter being estimated that minimizes the sum of the squared differences of the calculated values from their measured values, i.e., to minimize the sum of the squared measurement residuals.

Determining the derivative for $J^{\mathrm{LSM}}(x)$ in accordance with the rules (A1.63) given in Appendix and taking into consideration the necessary condition of minimum, it is possible to write the so-called **normal equations**:

$$\frac{dJ^{\mathrm{LSM}}(x)}{dx} = -2\frac{ds^{\mathrm{T}}(x)}{dx}(y - s(x)) = 0. \qquad (2.3)$$

Recall that (2.3) is only a necessary condition; to provide a local minimum, we need to check the validity of the sufficient condition:

$$\frac{\partial^2}{\partial x \partial x^{\mathrm{T}}} J^{\mathrm{LSM}}(x)\Big|_{\hat{x}^{\mathrm{LSM}}(y)} \geq 0. \qquad (2.4)$$

We speak about the **generalized least squares method** (**GLSM**) if the criterion (2.1) is replaced by the function:

$$J^{\text{GLMS}}(x) = (y - s(x))^{\text{T}} Q(y - s(x)), \tag{2.5}$$

in which Q is a symmetric nonnegatively definite matrix. So, if Q is considered a diagonal matrix with elements $q_i, i = \overline{1.m}$, instead of (2.1), we have

$$J^{\text{GLMS}}(x) = \sum_{i=1}^{m} q_i (y_i - s_i(x))^2 .$$

The reason for introducing the weighting matrix Q is to be able to take into account different contributions of the measured and calculated values corresponding to different components of the measurement vector. Sometimes this method is called the **weighted least squares method** (**WLSM**).

And, at last, we speak about **the modified least squares method (MLSM)** when the following criterion is used:

$$J^{MLSM}(x) = (y - s(x))^{\text{T}} Q(y - s(x)) + (x - \bar{x})^{\text{T}} D(x - \bar{x}) . \tag{2.6}$$

Here, $\bar{x} = \left(\bar{x}_1, ... \bar{x}_n\right)^{\text{T}}$ and $D \geq 0$ are a specified known vector and a symmetric nonnegatively definite matrix. For diagonal matrices Q and D, this criterion takes the following form:

$$J^{MLSM}(x) = \sum_{i=1}^{m} q_i (y_i - s_i(x))^2 + \sum_{j=1}^{n} d_j (x_j - \bar{x}_j)^2 .$$

The aim of introducing the second additional summand is that if the obtained estimates $\hat{x}_j(y)$ are different from some of the values $\bar{x}_j$, there is a certain penalty, the level of which is set by coefficients d_j, $j = \overline{1.n}$.

It follows that the determination of estimates corresponding to the LSM or one of its variants is reduced to finding the minimum of functions (2.1), (2.5), or (2.6).

♦ **Example 2.1.** Suppose we need to estimate amplitude A of the harmonic oscillations using measurements (1.14), i.e., $y_i = A\sin(\omega t_i + \phi_0) + v_i, i = \overline{1.m}$, assuming that the phase and the frequency are known.

Let us solve this problem using the LSM. In this problem, criterion (2.1) at $x = A$ is written as

$$J^{\text{LSM}}(x) = \sum_{i=1}^{m} (y_i - x\sin(\omega t_i + \phi_0))^2 .$$

According to this criterion, the estimate is sought rather simply. Indeed,

$$\frac{dJ^{\text{LSM}}(x)}{dx} = 2\sum_{i=1}^{m} (y_i - x\sin(\omega t_i + \phi_0))\sin(\omega t_i + \phi_0) = 0 ,$$

from where

$$\hat{x}^{\mathrm{LSM}} = \frac{1}{\sum_{i=1}^{m} \sin^2(\omega t_i + \phi_0)} \sum_{i=1}^{m} \sin(\omega t_i + \phi_0) y_i = \sum_{i=1}^{m} \tilde{q}_i y_i ,$$

where $\tilde{q}_i$ are the coefficients determined as

$$\tilde{q}_i = \frac{\sin(\omega t_i + \phi_0)}{\sum_{i=1}^{m} \sin^2(\omega t_i + \phi_0)}.$$

It is clear that the sufficient condition (2.4) is also satisfied since

$$\frac{d^2 J^{\mathrm{LSM}}(x)}{dx^2} = 2 \sum_{i=1}^{m} \sin^2(\omega t_i + \phi_0) \geq 0 .$$

Figure 2.1 shows the samples of the measured values of the harmonic oscillation with $\omega = 1$ rad/s, $\varphi_0 = \frac{\pi}{2}$ and the unit amplitude on a 2-second interval with a 0.02-second increment, and the sample of harmonic oscillation without error, calculated for the same parameters at three amplitude values: $A_1 = 0.5$, $A_2 = 1.0$, and $A_3 = 1.5$.

Figure 2.1 makes clear the geometric sense of the problem of finding estimates using the LSM: to select (find) the amplitude value at which the best coincidence of the measured and calculated samples of harmonic oscillation can be obtained.

For the example under consideration, criterion $J^{\mathrm{LSM}}(x) = \sum_{i=1}^{m} (y_i - x \sin(\omega t_i + \phi_0))^2$ being minimized is shown in Fig. 2.2, from which it is evident that function $J^{\mathrm{LSM}}(x)$ is a parabola with one minimum at the point that determines the value of the estimate corresponding to the LSM.

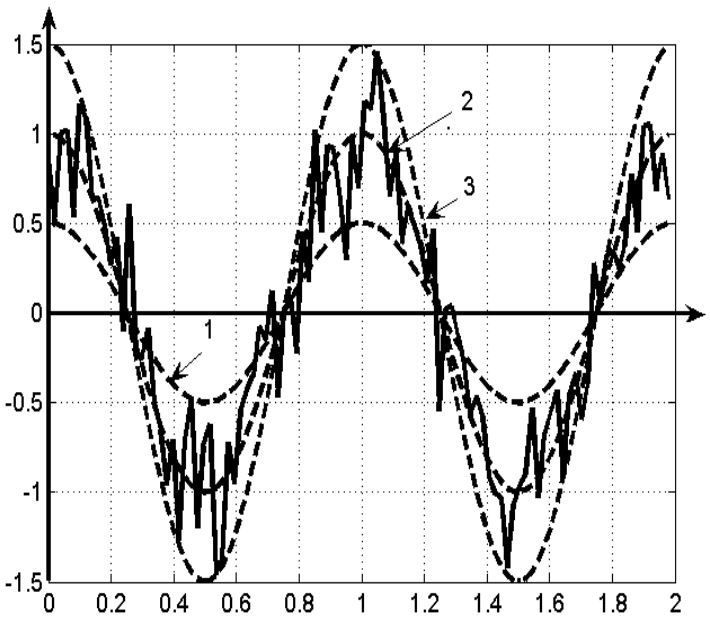


Fig. 2.1. Measured and calculated harmonic oscillations at different amplitude values: *1* – 0.5; *2* – 1.0; *3* – 1.5

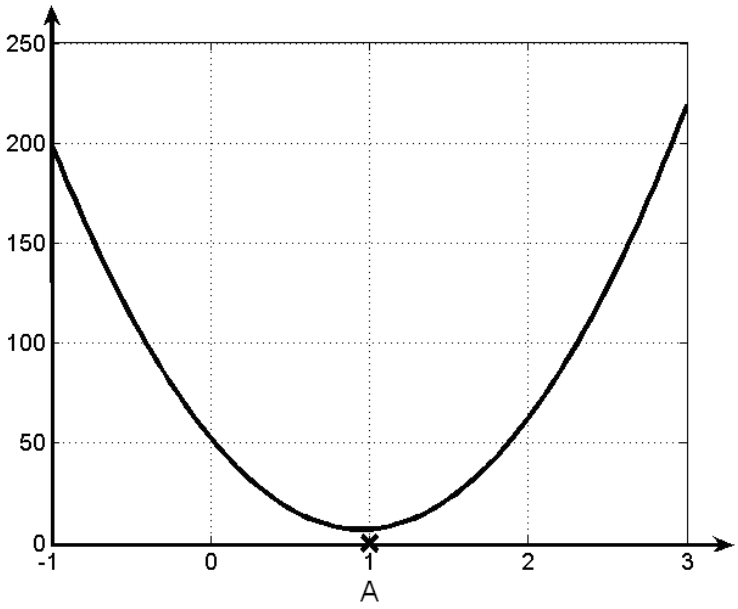


Fig. 2.2. Criterion $J^{\mathrm{LSM}}(x)$ as a function of x

The criterion being minimized for three values of the amplitudes— $A_1 = 0.5$, $A_2 = 1.0$, and $A_3 = 1.5$—and the minimum of the criterion at $\hat{x} = 0.9448\,0.9448$ in this example are given in Table 2.1.

Table 2.1

The criterion for different values of the amplitude

$x = A$	0.5	1.0	1.5	$\hat{x} = \hat{A} =$ 0.9448
$J^{\mathrm{LSM}}(x)$	18.8	9.0	24.4	8.86

From the table it follows that the criterion has a minimum at the point that does not coincide with the true value of the amplitude. This is due to the measurement errors generating the difference of the estimate from the true parameter being estimated, which is understandable since the equation for the estimate can be written as

$$\hat{x}^{\mathrm{LSM}} = \frac{\sum_{i=1}^{m} \sin(\omega t_i + \phi_0)\left(x \sin(\omega t_i + \phi_0) + v_i\right)}{\sum_{i=1}^{m} \sin^2(\omega t_i + \phi_0)} = x + \sum_{i=1}^{m} \tilde{q}_i v_i \,.$$

Hence, the following equation for the estimation error is valid:

$$\hat{x}^{\mathrm{LSM}} - x = \sum_{i=1}^{m} \tilde{q}_i v_i \,.$$

It is not difficult to obtain the solution to the problem under consideration for the GLSM as well. Thus, in the case of diagonal matrix Q and $x = A$, criterion (2.5) is written as

$$J^{\mathrm{GLSM}}(x) = \sum_{i=1}^{m} q_i (y_i - x\sin(\omega t_i + \phi_0))^2 .$$

It is quite easy to find the estimate corresponding to this criterion. Indeed,

$$\frac{dJ^{\mathrm{GLSM}}(x)}{dx} = 2\sum_{i=1}^{m} q_i (y_i - x\sin(\omega t_i + \phi_0))\sin(\omega t_i + \phi_0) = 0 ,$$

from where

$$\hat{x}^{\mathrm{GLSM}} = \frac{1}{\sum_{i=1}^{m} q_i \sin^2(\omega t_i + \phi_0)} \sum_{i=1}^{m} q_i \sin(\omega t_i + \phi_0) y_i = \sum_{i=1}^{m} \tilde{q}_i y_i ,$$

where $\tilde{q}_i$ are the coefficients calculated as

$$\tilde{q}_i = \frac{q_i \sin(\omega t_i + \phi_0)}{\sum_{i=1}^{m} q_i \sin^2(\omega t_i + \phi_0)} .$$

Finding the estimate for the MLSM does not present any problem either. In this case, too, the problem solution reduces to finding the location of the parabola extreme point:

$$J^{\mathrm{MLSM}}(x) = J^{\mathrm{GLSM}}(x) + d(x - \overline{x})^2 . \quad \blacklozenge$$

2.2. General solution of the linear estimation problem using the least squares method

The ease of obtaining the algorithm for the estimate calculation, i.e., the design problem in the above example, is a consequence of the fact that measurements are linearly dependent on the unknown amplitude and measurement errors. Consider to what the algorithm for solution of the estimation problem in the linear statement is reduced in the general linear case, i.e., when measurements may be represented as $y = Hx + v$. Having such measurements, for the LSM, the criterion to be minimized is written as

$$J^{\mathrm{LSM}}(x) = (y - Hx)^{\mathrm{T}}(y - Hx) . \qquad (2.7)$$

It is easy to see that with respect to x, this function is defined as quadratic form:

$$J^{\mathrm{LSM}}(x) = x^{\mathrm{T}} H^{\mathrm{T}} H x - 2x^{\mathrm{T}} H^{\mathrm{T}} y + y^{\mathrm{T}} y , \qquad (2.8)$$

which, at nonsingular $H^{\mathrm{T}}H$, has a one extremum; at the same time, the sufficient condition (2.4) is true.

Taking into consideration (A1.61), the normal equations corresponding to criterion (2.7) can be written as

$$H^{\mathrm{T}}(y - Hx) = 0 .$$

Here, the condition of $H^{\mathrm{T}}H$ nonsingularity is called **the observability condition**. The choice of this term is quite justified because in this case we have:

$$\hat{x}^{\mathrm{LSM}}(y) = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}y, \tag{2.9}$$

or

$$\hat{x}^{\mathrm{LSM}}(y) = K^{\mathrm{LSM}}y, \tag{2.10}$$

where

$$K^{\mathrm{LSM}} = \left(H^{\mathrm{T}}H\right)^{-1}H^{\mathrm{T}}. \tag{2.11}$$

Hence, it follows that in the absence of measurement errors and fulfillment of the observability condition,

$$\hat{x}^{\mathrm{LSM}}(y) = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}Hx \equiv x, \tag{2.12}$$

i.e., the estimate coincides with the true value of the vector.

Proceeding in a similar manner in solving the linear problem, in the case of the GLSM, and assuming $H^{\mathrm{T}}QH$ to be nonsingular, we can write:

$$J^{\mathrm{GLSM}}(x) = (y - Hx)^{\mathrm{T}}Q(y - Hx); \tag{2.13}$$

$$\hat{x}^{\mathrm{GLSM}}(y) = K^{\mathrm{GLSM}}y, \tag{2.14}$$

where

$$K^{\mathrm{GLSM}} = \left(H^{\mathrm{T}}QH\right)^{-1}H^{\mathrm{T}}Q. \tag{2.15}$$

Note that, generally, matrix Q is nonsingular, and, therefore, if the observability condition is met, matrix $H^{\mathrm{T}}QH$ is also nonsingular.

For the MLSM (see Exercise 2.1), we have the following equations:

$$J^{\mathrm{MLSM}}(x) = (y - Hx)^{\mathrm{T}}Q(y - Hx) + (x - \overline{x})^{\mathrm{T}}D(x - \overline{x}); \tag{2.16}$$

$$\hat{x}^{\mathrm{MLSM}}(y) = \overline{x} + K^{\mathrm{MLSM}}(y - H\overline{x}), \tag{2.17}$$

where

$$K^{\mathrm{MLSM}} = \left(D + H^{\mathrm{T}}QH\right)^{-1}H^{\mathrm{T}}Q. \tag{2.18}$$

The above equations are summarized in Table 2.2.

We draw your attention to a very important fact: all the resulting estimates in this table are linearly dependent on measurements, which is due to linearity of measurements and a quadratic nature of the criteria (2.7), (2.13), and (2.16) to be minimized.

It is clear that using the relationships from Table 2.2 for the above example, we can easily obtain the formulas for the estimates. For this purpose, we need to take into account matrix H, which is determined as

$H^{\mathrm{T}} = [\sin(\omega t_1), \sin(\omega t_2), \ldots . \sin(\omega t_m)]$.

Table 2.2

Criteria and algorithms in the linear estimation problem

Method	Criterion	Algorithm
LSM	$J^{\mathrm{LSM}}(x)=(y-Hx)^{\mathrm{T}}(y-Hx)$	$\hat{x}^{\mathrm{LSM}}(y)=K^{\mathrm{LSM}}y,$ $K^{\mathrm{LSM}}=\left(H^{\mathrm{T}}H\right)^{-1}H^{\mathrm{T}}$
GLSM	$J^{\mathrm{GLSM}}(x)=(y-Hx)^{\mathrm{T}}Q(y-Hx)$	$\hat{x}^{\mathrm{GLSM}}(y)=K^{\mathrm{GLSM}}y,$ $K^{\mathrm{GLSM}}=\left(H^{\mathrm{T}}QH\right)^{-1}H^{\mathrm{T}}Q$
MLSM	$J^{\mathrm{MLSM}}(x)=(y-Hx)^{\mathrm{T}}Q(y-Hx)+$ $+(x-\overline{x})^{\mathrm{T}}D(x-\overline{x})$	$\hat{x}^{\mathrm{MLSM}}(y)=\overline{x}+K^{\mathrm{MLSM}}(y-H\overline{x}),$ $K^{\mathrm{MLSM}}=\left(D+H^{\mathrm{T}}QH\right)^{-1}H^{\mathrm{T}}Q$

Consider two more examples.

♦ **Example 2.2.** Let us find the formulas for the estimates in the problem of estimating an unknown scalar value by scalar measurements $y_i=x+v_i$. In so doing, for simplicity, in criterion (2.13), matrix Q is chosen to be diagonal with elements $q_i>0, i=\overline{1.m}$, and in criterion (2.16), it is assumed that $D=d$.

Obviously, in this case too, the criterion to be minimized has the form of a parabola, and, since matrix $H^{\mathrm{T}}=[1,1....1]$, the equations for the estimates will be determined in accordance with the formulas in Table 2.3.

From these equations it follows that in the LSM, the estimate is an arithmetic mean of measurements y_i. For the GLSM, the estimate is determined by "weighting" measurements with normalized coefficients:

$$\sum_{i=1}^{m}\tilde{q}_i^{\mathrm{GLSM}}=1;\quad \tilde{q}_i^{\mathrm{GLSM}}=\frac{q_i}{\sum_{i=1}^{m}q_i}.$$

For the MLSM, normalization is changed due to the presence of a priori information and an additional term in the criterion.

Table 2.3

Estimation algorithms for three variants of the LSM for a simple example of finding x using measurements $y_i = x + v_i$, $i = \overline{1.m}$

Method	Criterion	Algorithm
LSM	$J^{\text{LSM}}(x) = \sum_{i=1}^{m} (y_i - x)^2$	$\hat{x}^{\text{LSM}} = \frac{1}{m} \sum_{i=1}^{m} y_i$
GLSM	$J^{\text{GLSM}}(x) = \sum_{i=1}^{m} q_i (y_i - x)^2$	$\hat{x}^{\text{GLSM}} = \sum_{i=1}^{m} \tilde{q}_i^{\text{GLSM}} y_i,$ $\tilde{q}_i^{\text{GLSM}} = \frac{q_i}{\sum_{i=1}^{m} q_i}$
MLSM	$J^{\text{MLSM}}(x) = d(x - \overline{x})^2 +$ $+ \sum_{i=1}^{m} q_i (y_i - x)^2$	$\hat{x}^{\text{MLSM}} = \beta \overline{x} + \sum_{i=1}^{m} \tilde{q}_i^{\text{MLSM}} y_i,$ $\beta = \frac{d}{d + \sum_{i=1}^{m} q_i},$ $\tilde{q}_i^{\text{MLSM}} = \frac{q_i}{d + \sum_{i=1}^{m} q_i}$

♦ **Example 2.3.** Assume that we need to estimate the initial value of coordinate x_0 and speed V using measurements $y_i = x_0 + Vt_i + v_i$.

In this problem, the LSM criterion will be determined as

$$J^{\text{LSM}}(x_0, V) = \sum_{i=1}^{m} (y_i - x_0 - Vt_i)^2 .$$

Taking into consideration the notation in (1.4), it is easy to see that this criterion is the quadratic form (2.8), which can be written as

$$J^{\text{LSM}}(x) = (x_1, x_2)^{\text{T}} \left(\begin{bmatrix} m & \sum_{i=1}^{m} t_i \\ \sum_{i=1}^{m} t_i & \sum_{i=1}^{m} t_i^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 2 \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} t_i y_i \end{bmatrix} \right) + \sum_{i=1}^{m} y_i^2$$

or

$$J^{\mathrm{LSM}}(x) = mx_1^2 + x_2^2\sum_{i=1}^{m} t_i^2 + 2x_1x_2\left(\sum_{i=1}^{m} t_i\right)^2 - 2\left(x_1\sum_{i=1}^{m} y_i + x_2\sum_{i=1}^{m} t_i y_i\right) + \sum_{i=1}^{m} y_i^2 .$$

Using (2.9), we obtain the following equation:

$$\begin{bmatrix} \hat{x}_1^{\mathrm{LSM}} \\ \hat{x}_2^{\mathrm{LSM}} \end{bmatrix} = \begin{bmatrix} m & \sum_{i=1}^{m} t_i \\ \sum_{i=1}^{m} t_i & \sum_{i=1}^{m} t_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} t_i y_i \end{bmatrix}. \qquad (2.19)$$

For the GLSM, with diagonal matrix Q, the criterion will include multipliers q_i, and the equations for the estimate are transformed into the form:

$$\hat{x}^{\mathrm{GLSM}} = \begin{bmatrix} \sum_{i=1}^{m} q_i & \sum_{i=1}^{m} q_i t_i \\ \sum_{i=1}^{m} q_i t_i & \sum_{i=1}^{m} q_i t_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{m} q_i y_i \\ \sum_{i=1}^{m} q_i t_i y_i \end{bmatrix}.$$

In this example, it is not difficult to define concretely the equation for the criterion and the estimate applied to the MLSM.

Figure 2.3 shows the sample of the measured coordinates against the background of the true coordinates for the case when measurements are taken on a 10-second interval every second at $x_0 = 1$m and $V = 1$ m/s. Figure 2.4 shows criterion $J^{\mathrm{LSM}}(x)$; it is a paraboloid, extended along one of the axes, corresponding to quadratic form.

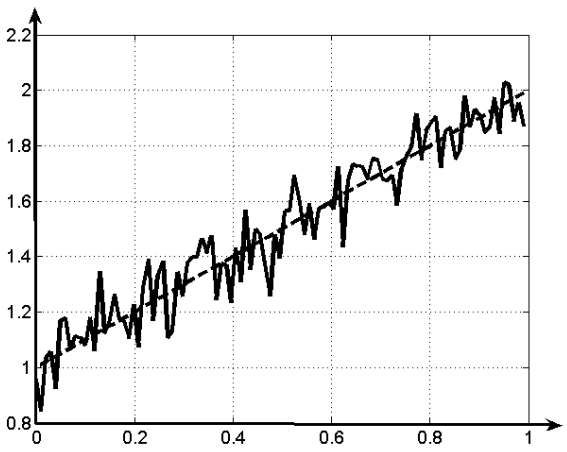


Fig. 2.3. Sample of measurements $y_{1i} = x_0 + Vt_i + v_i, i = \overline{1.m}$

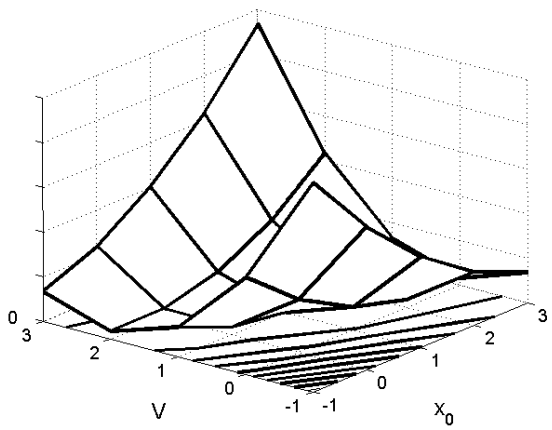


Fig. 2.4. Criterion $J^{LSM}(x_0, V)$ being minimized

♦

2.3. Accuracy analysis of the least squares method in the linear case

In the problem statement in section 1.1, it was noted that besides the solution of the algorithm design problem, the problem of accuracy analysis is of no less importance, i.e., studying properties of estimate errors. For the linear case, using (2.10), (2.14), for the LSM and GLSM, we can obtain the following formulas for the estimate errors:

$$\varepsilon^{\mathrm{LSM}}(y) = x - \hat{x}^{\mathrm{LSM}}(y) = x - K^{\mathrm{LSM}} y = (E - K^{\mathrm{LSM}} H)x - K^{\mathrm{LSM}} \nu;$$

$$\varepsilon^{\mathrm{GLSM}}(y) = x - \hat{x}^{\mathrm{GLSM}}(y) = x - K^{\mathrm{GLSM}} y = (E - K^{\mathrm{GLSM}} H)x - K^{\mathrm{GLSM}} \nu.$$

Since the following equations are valid for the LSM and GLSM,

$$E - K^{\mathrm{LSM}} H = E - \left(H^{\mathrm{T}} H\right)^{-1} H^{\mathrm{T}} H = 0; \tag{2.20}$$

$$E - K^{\mathrm{GLSM}} H = E - \left(H^{\mathrm{T}} Q H\right)^{-1} H^{\mathrm{T}} Q H = 0, \tag{2.21}$$

the errors corresponding to these two methods can be written as

$$\varepsilon^{\mathrm{LSM}}(y) = -K^{\mathrm{LSM}} \nu; \tag{2.22}$$

$$\varepsilon^{\mathrm{GLSM}}(y) = -K^{\mathrm{GLSM}} \nu. \tag{2.23}$$

For the MLSM, since $(E - K^{\mathrm{MLSM}} H) \neq 0$, the equation for the estimation errors will take the form:

$$\varepsilon^{\mathrm{MLSM}}(y) = x - \hat{x}^{\mathrm{MLSM}}(y) = (E - K^{\mathrm{MLSM}} H)(x - \overline{x}) - K^{\mathrm{MLSM}} \nu. \tag{2.24}$$

From the above formulas, we can make the following conclusions. The estimation errors corresponding to the LSM and GLSM in the linear case do not contain summands that depend on the vector of the parameters being estimated; they only depend on measurement errors. In these conditions, it is possible to speak about the **invariance (independence) of estimation errors** with respect to the vector being estimated. As for the MLSM, the estimate error depends, in addition, on the value of the vector being estimated; thus, it has no property of invariance.

The advantage of the methods under consideration is that at the stage of the algorithm design, we do not need any a priori information of a statistical nature. However, its absence makes the solution of the accuracy analysis problem difficult. This difficulty may be obviated by introducing an assumption of a random nature of both the measurement errors (in the case of LSM and GLSM) and the vector being estimated (in the case of MLSM).

Introduce such assumptions, assuming, in particular, that the measurement errors ν_i, $i = \overline{1.m}$ are zero-mean random variables with the known covariance matrix R. In this case, from (2.22), (2.23), it is inferred that the estimate errors for the LSM and GLSM will also be zero mean with the covariance matrices:

$$P^{\mathrm{LSM}} = M\left\{(K^{\mathrm{LSM}}\nu)(K^{\mathrm{LSM}}\nu)^{\mathrm{T}}\right\} = (H^{\mathrm{T}}H)^{-1} H^{\mathrm{T}} R H (H^{\mathrm{T}}H)^{-1}; \tag{2.25}$$

$$P^{\mathrm{GLSM}} = \left(H^{\mathrm{T}} Q H\right)^{-1} H^{\mathrm{T}} Q R Q H \left(H^{\mathrm{T}} Q H\right)^{-1}. \tag{2.26}$$

To calculate the covariance matrix of estimation errors for the MLSM, it is necessary not only to introduce an additional assumption about the random nature of the vector being estimated and specify statistical properties for vectors v and x separately, but also to define their mutual statistical properties. If, for example, we assume that x and v are uncorrelated random vectors with expectations $\bar{x}$ and $\bar{v} = 0$ and covariance matrices R, P^x, then, using (2.24), it is easy to verify that the estimation errors are also zero mean, so that for the corresponding covariance matrix, we can obtain the following formula:

$$P^{\text{MLSM}} = (E - K^{\text{MLSM}} H) P^x (E - K^{\text{MLSM}} H)^{\text{T}} + K^{\text{MLSM}} R (K^{\text{MLSM}})^{\text{T}}, \tag{2.27}$$

where matrix K^{MLSM} is given by Equation (2.18).

If in criterion (2.13) we assume $Q = R^{-1}$, the GLSM estimates and the covariance matrix of their errors can be written as

$$\hat{x}^{\text{GLSM}}(y) = (H^{\text{T}} R^{-1} H)^{-1} H^{\text{T}} R^{-1} y, \tag{2.28}$$

$$P^{\text{GLSM}} = \left(H^{\text{T}} R^{-1} H \right)^{-1}. \tag{2.29}$$

If, in addition, we assume that $D = \left(P^x \right)^{-1}$, then $K^{\text{MLSM}} = \left(\left(P^x \right)^{-1} + H^{\text{T}} R^{-1} H \right)^{-1} H^{\text{T}} R^{-1}$ and the equations for estimate (2.17) and the corresponding covariance matrix (see Exercise 2.3) will take the form:

$$\hat{x}^{\text{MLSM}}(y) = \bar{x} + \left((P^x)^{-1} + H^{\text{T}} R^{-1} H \right)^{-1} H^{\text{T}} R^{-1} (y - H\bar{x}) \ ; \tag{2.30}$$

$$P^{\text{MLSM}} = \left(\left(P^x \right)^{-1} + H^{\text{T}} R^{-1} H \right)^{-1}. \tag{2.31}$$

The possibility of calculating estimation error covariance matrices provides a significant advance in solving the problem of accuracy analysis since this allows it to be characterized quantitatively. In particular, the diagonal elements of the obtained estimation error covariance matrices (2.25)–(2.27) represent the variances of the estimation errors of the components of vector x being estimated.

The values of the diagonal elements of matrix P^x are usually called **a priori estimation error variances**. This name is quite justified if it is assumed that before taking measurements, it makes sense to use the value of mathematical expectation $\bar{x}$ as a priori estimate. The comparison of a priori variances with the corresponding **a posteriori variances** (diagonal elements of covariance matrices (2.25)–(2.27)), i.e., with those obtained after taking measurements with the use of different algorithms, makes it possible to estimate the efficiency of these algorithms. Accordingly, the covariance matrix P^x is called **a priori matrix** and matrices (2.25)–(2.27), **a posteriori covariance matrices of estimation errors.**

If, in addition, along with the assumption of the known values of the expectations and covariance matrices, it is assumed that the measurement errors and the vector being estimated are Gaussian vectors, then, due to the linear nature of transformations

(2.22)–(2.24), this provides the Gaussian nature of the estimation errors for these methods as well. In other words, the PDF for the vector of estimation errors becomes known. The PDF availability provides a complete description of statistical properties for estimation errors. In particular, we can calculate for them such characteristics as a probable error, three-sigma limit, error, quantile, etc.

♦ **Example 2.4.** Assume that in Example 2.2, x is a random variable with expectation $\bar{x}$ and variance σ_0^2, and the measurement errors are zero-mean, uncorrelated with each other and x, random variables with variances r_i^2, $i=\overline{1.m}$ (measurements with unequal accuracy) and $r_i^2=r^2$, $i=\overline{1.m}$ (measurements with equal accuracy) in a particular case. Let us derive the equations for estimation errors and the corresponding variances for the three variants of the LSM as applied to the problem of estimating scalar value x.

Under the assumptions made, matrix R is a diagonal one with elements r_i^2, $i=\overline{1.m}$; matrix $H^{\mathrm{T}}=[1,1..1]$; matrix Q is a diagonal one with elements $q_i=1/r_i^2$ in the first case, and $q_i=1/r^2$, $i=\overline{1.m}$ in the second case. Thus, it is not difficult to obtain the equations given in Table 2.4.

Table 2.4

Equations for estimation errors and their variances for three variants of the LSM for the simplest example

Method	**Estimation errors**	**Variances**	
		measurements with unequal accuracy	measurements with equal accuracy $r_i^2 = r^2$
LSM	$\varepsilon^{LSM} = \frac{1}{m}\sum_{i=1}^{m} v_i$	$P^{LSM} = \frac{\sum_{i=1}^{m} r_i^2}{m^2}$	$P^{LSM} = \frac{r^2}{m}$
GLSM	$\varepsilon^{GLSM} = \sum_{i=1}^{m} \tilde{q}_i^{GLSM} v_i$, $\tilde{q}_i^{GLSM} = \frac{q_i}{\sum_{i=1}^{m} q_i}$	$P^{GLSM} = \left(\sum_{i=1}^{m} \frac{1}{r_i^2} \right)^{-1}$	$P^{GLSM} = \frac{r^2}{m}$
MLSM	$\varepsilon^{MLSM}(y) = \beta(x - \bar{x}) + \sum_{i=1}^{m} \tilde{q}_i^{MLSM} v_i$, $\tilde{q}_i^{MLSM} = \frac{q_i}{d + \sum_{i=1}^{m} q_i}$, $\beta = \frac{d}{d + \sum_{i=1}^{m} q_i}$	$P^{MLSM} = \left[\frac{1}{\sigma_0^2} + \sum_{i=1}^{m} \frac{1}{r_i^2} \right]^{-1}$	$P^{MLSM} = \frac{\sigma_0^2 r^2}{r^2 + \sigma_0^2 m}$

Table 2.5 shows the root-mean-square (RMS) value of the estimation errors, depending on the number of measurements for the GLSM and MLSM for equal variances of measurement errors r^2=1, $m = \overline{1.10}$.

Table 2.5

RMS value of the estimation errors, depending on the number of measurements for the GLSM and MLSM for different values of σ_0

Method	Measurement number									
	1	2	3	4	5	6	7	8	9	10
GLSM	1	0.7	0.58	0.5	0.45	0.41	0.38	0.35	0.33	0.32
MLSM $\sigma_0 \geq 10$, $\overline{x} = 0$	1	0.7	0.58	0.5	0.45	0.41	0.38	0.35	0.33	0.32
MLSM $\sigma_0 = 1$, $\overline{x} = 0$	0.7	0.58	0.5	0.45	0.41	0.38	0.35	0.33	0.32	0.3

From the results obtained it follows that for the GLSM and MLSM for $\sigma_0 \geq 10$ there are no differences in accuracy. In this case, the influence of a priori information is insignificant because $\sigma_0 >> r$. For $\sigma_0 = r = 1$, this effect is significant only for a small number of measurements. In fact, the RMS error for the GLSM coincides with that for the MLSM at the previous step. This behavior is easy to explain if we take into account a possible interpretation of a priori information as an additional measurement (2.46), which was described below in section 2.4. ♦

Now, let us concretize formulas (2.25), (2.26), (2.31) for the problem of estimating a two-dimensional vector considered in Example 2.3.

♦ **Example 2.5.** Now, let us obtain equations for the covariance matrices of the errors in estimating the coefficients of the polynomial of degree 1 using measurements of type (1.3), i.e., $y_i = x_0 + Vt_i + v_i$, assuming that the measurement errors are zero-mean random variables, uncorrelated with each other, with identical variances r^2, whereas the parameters being estimated are zero-mean ($\overline{x} = 0$) random variables, uncorrelated with each other and the measurement errors, with covariance matrix $P^x = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix}$.

Using Equations (1.4), (2.25), (2.26) with diagonal $Q = \{q_i\}$, in the criterion for the GLSM, it is easy to derive the following relations:

$$P^{\mathrm{LSM}} = r^2 \begin{bmatrix} m & \sum_{i=1}^{m} t_i \\ \sum_{i=1}^{m} t_i & \sum_{i=1}^{m} t_i^2 \end{bmatrix}^{-1}; \qquad (2.32)$$

$$P^{\mathrm{LSM}} = \begin{bmatrix} \sum_{i=1}^{m} q_i & \sum_{i=1}^{m} q_i t_i \\ \sum_{i=1}^{m} q_i t_i & \sum_{i=1}^{m} q_i t_i^2 \end{bmatrix}^{-1}. \qquad (2.33)$$

Assuming $Q = R^{-1}$, and, in addition, for the MLSM, assuming that $\bar{x} = 0$, $D = \left(P^x\right)^{-1}$, and also, taking into consideration the fact that $R = r^2 E$, it is easy to see that the estimation error covariance matrices in the LSM and GLSM coincide, i.e., $P^{\mathrm{GLSM}} = P^{\mathrm{LSM}}$, and for the MLSM,

$$P^{\mathrm{MLSM}} = \begin{bmatrix} \frac{1}{\sigma_0^2} + \frac{m}{r^2} & \frac{1}{r^2}\sum_{i=1}^{m} t_i \\ \frac{1}{r^2}\sum_{i=1}^{m} t_i & \frac{1}{\sigma_1^2} + \frac{1}{r^2}\sum_{i=1}^{m} t_i^2 \end{bmatrix}^{-1}. \qquad (2.34)\ \blacklozenge$$

2.4. Solution of nonlinear estimation problems. Linearized and iterated algorithms

In the previous sections we obtained simple algorithms for calculation of errors corresponding to the LSM and its modifications. Introducing additional assumptions about the random nature of the vector being estimated and measurement errors, it is also easy to calculate the covariance matrices of estimation errors which are used to analyze the estimation accuracy. The simplicity of the procedures for calculation of errors and their accuracy characteristics is, in this case, the result of the linear nature of these problems. The estimation problem solution becomes much more complicated in the case of nonlinear measurement dependence on the parameters being estimated. At the same time, as it has been mentioned, a wide scope of nonlinear problems on navigation data processing can be effectively solved with the use of the algorithms obtained above. Their application in the nonlinear case is based on the linearization procedure described in 1.7. Let us consider this question in greater detail.

After linearization, the initial nonlinear problem can be easily reduced to the linear statement, in which the values of (1.25) represented in the form (1.26) are treated as measurements. Using these equations, it is easy to obtain linearized variants of the LSM and its modifications. In this case, the efficiency of the designed algorithms significantly depends on how the linearization point is chosen. The closer it is to the true value of the parameter being estimated, the more accurate the linearized

representation and the more accurate the estimate obtained on the basis of linearized algorithms are. Therefore, to improve the efficiency of the algorithms based on linearization, the linearization point should be chosen so that it will be as close to the true unknown value of the vector being estimated as possible. From the above it follows that the efficiency of the linearized algorithms can be improved if we use a rather evident procedure, namely, we will repeat the measurement processing many times in order to use the results for refinement of the linearization point location. Let us explain the essence of this procedure.

After selecting the starting linearization point x^l and using the approximation,

$$s(x) \approx s(x^l) + \left.\frac{ds}{dx^{\mathrm{T}}}\right|_{x=x^l} (x - x^l) = s(x^l) + H(x^l)(x - x^l),$$

we form the initial estimate of the vector being estimated with the use of equation

$$\hat{x}^{(1)} = \overline{x} + K\,(x^l)\left[\, y - s\,(x^l) - H^{\,(1)}(x^l)\left(\overline{x} - x^l\right)\right], \qquad (2.35)$$

in which $K\,(x^l)$ is calculated in accordance with the rules of the chosen LSM. Repeat this procedure until the value of the estimate stops varying significantly. In a general case, for the MLSM, this algorithm takes the form:

$$\hat{x}^{(\gamma+1)} = \overline{x} + K\,(\hat{x}^{\,(\gamma)})\left[\, y - s\,(\hat{x}^{\,(\gamma)}) - H^{\,(\gamma)}(\hat{x}^{\,(\gamma)})\left(\overline{x} - \hat{x}^{\,(\gamma)}\right)\right]; \qquad (2.36)$$

$$K(\hat{x}^{\,(\gamma)}) = P(\hat{x}^{\,(\gamma)})H^{\mathrm{T}}(\hat{x}^{\,(\gamma)})R^{-1}; \qquad (2.37)$$

$$P(\hat{x}^{\,(\gamma)}) = \left((P^x)^{-1} + H^{\mathrm{T}}(\hat{x}^{\,(\gamma)})R^{-1}H(\hat{x}^{\,(\gamma)})\right)^{-1}; \qquad (2.38)$$

$$\gamma = 0,1,2..,\quad \hat{x}^{(0)} = \overline{x}.$$

For the MLSM, in these equations, we assume that $(P^x)^{-1} = 0$ and $\overline{x} = 0$, and for the LSM, in addition, $R^{-1} = E$. In estimation theory, such algorithms are called **iterated algorithms or algorithms with local iterations** [9, 20, 22].

Note that in view of the equality of the $KH = E$ type, when using the LSM or GLSM, the equation for the estimate (2.36) can be represented in the following form, recursive with respect to the number of the iteration and convenient for practical implementation: $\hat{x}^{(\gamma+1)} = \hat{x}^{(\gamma)} + \delta\hat{x}^{(\gamma+1)}$, where $\delta\hat{x}^{(\gamma+1)} = K\,(\hat{x}^{\,(\gamma)})\left[\, y - s\,(\hat{x}^{\,(\gamma)})\right]$.

If the measurement errors are assumed to be a zero-mean random vector with covariance matrix R and the vector being estimated is a random vector with mathematical expectation $\overline{x}$ and covariance matrix P^x, then, Equation (2.38) will determine **the calculated covariance matrix of estimation errors** with an accuracy of the assumption that the linearized description is valid. This term is due to the fact that the covariance matrix, obtained under the assumption that the linearized representation of function $s(x)$ is valid, differs from **the real covariance matrix**, which, for an arbitrary estimate $\tilde{x}(y)$ and the estimate generated with the use of linearized or iterated algorithms, is given as

$$\tilde{P} = E\left\{(x - \tilde{x}(y))(x - \tilde{x}(y))^{\mathrm{T}}\right\}. \qquad (2.39)$$

In this connection, there arises **the problem of consistency**, i.e., agreement between the calculated covariance matrices and their real values.

It is also important to emphasize that the resulting algorithms are no longer linear with respect to the measurements since there is a nonlinear dependence on the measurements because matrices $K(\hat{x}^{(\gamma)}) = K(\hat{x}^{(\gamma)}(y))$ depend on measurements. The calculated covariance matrix $P(\hat{x}^{(\gamma)}) = P(\hat{x}^{(\gamma)}(y))$ also depends on measurements.

The block diagram of the iterated algorithm is shown in Fig. 2.5.

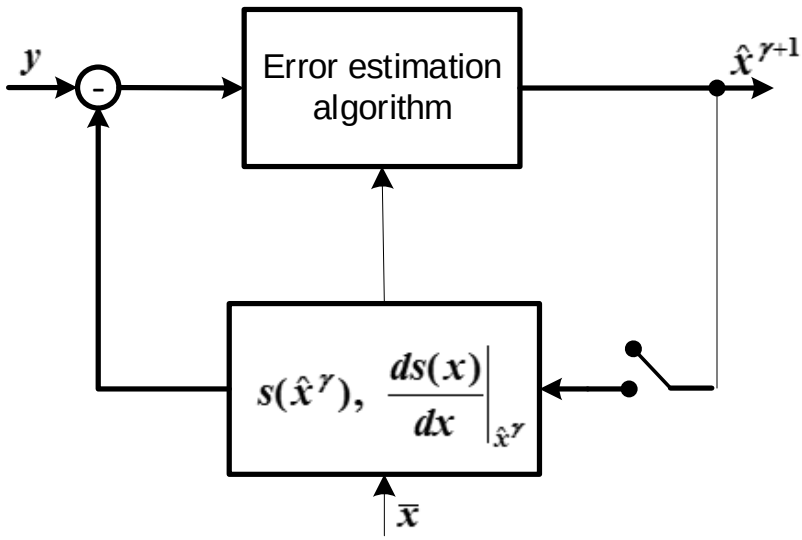


Fig. 2.5. Block diagram of the iterated estimation algorithm

♦ **Example 2.8.** Let us obtain two algorithms corresponding to the LSM in the problem of phase estimation. One of them is based on linearization, and the other one is an iterated algorithm. In doing so, we assume that the amplitude and the frequency are known. For simplicity, *A* is assumed to equal to unit.

In this case, criterion (2.1) for the LSM at $x = \phi_0$ can be written as

$$J^{\mathrm{LSM}}(x) = \sum_{i=1}^{m} (y_i - \sin(\omega t_i + x))^2. \qquad (2.40)$$

To find the estimate corresponding to criterion (2.40), it is necessary to find the point of its minimum value on the *x*-axis or, as a preliminary, try to solve the normal equation corresponding to the necessary condition for an extremum, which in this case takes the form:

$$\frac{dJ^{\mathrm{LSM}}(x)}{dx} = 2\sum_{i=1}^{m} (y_i - \sin(\omega t_i + x))\cos(\omega t_i + x) = 0.$$

The fact that the problem of criterion (2.40) minimization can be reduced to the problem of finding the roots of this nonlinear equation does not make the solution easy. We use a linearized description of function $\sin(\omega t_i + x\ )$:

$$\sin(\omega t_i + x\ ) \approx \sin(\omega t_i + x^l) + (x - x^l) A\cos(\omega t_i + x^l)\ ,\ i = \overline{1.m}\ ,$$

where x^l is the chosen linearization point. Then, the criterion in the neighborhood of the linearization point will be a parabola:

$$J^{\mathrm{LSM}}(x) \approx \sum_{i=1}^{m} (\tilde{y}_i(x^l) - (x - x^l)\cos(\omega t_i + x^l))^2\ , \qquad (2.41)$$

where $\tilde{y}_i(x^l) \overset{\Delta}{=} y_i - \sin(\omega t_i + x^l)$.

Figure 2.6 shows the curves for the criterion being minimized, corresponding to the original nonlinear function and its linearized description for the true value of the phase $\phi_0 = \dfrac{\pi}{2}$ and $x^l = \bar{\phi}_0 = \dfrac{3\pi}{4}$.

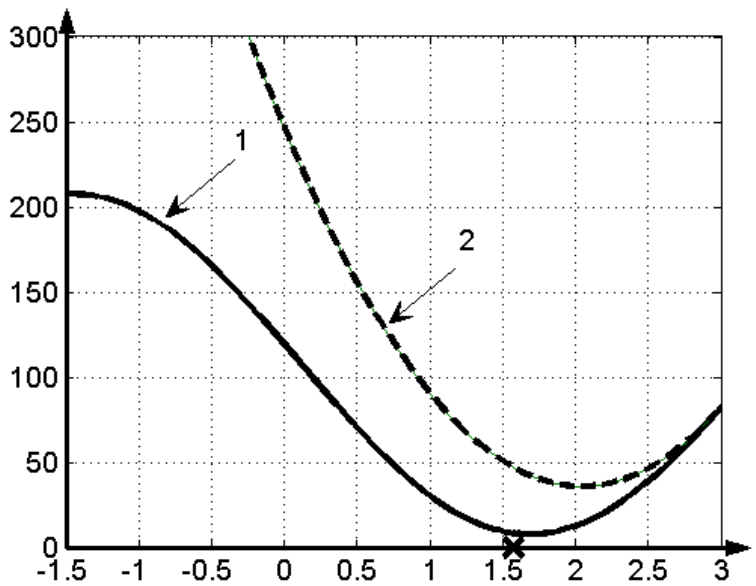


Fig. 2.6. A plot of criterion $J^{\mathrm{LSM}}(\phi_0)$ (1) and its approximate description (2) in the phase estimation problem

Taking $x^l = \hat{x}^{(0)} = \bar{\phi}_0$ as a starting point of linearization, where $\bar{\phi}_0$ is a prescribed a priori phase value, and using relation (2.35), taking into consideration $H^{(0)} = (\cos(\omega t_1 + \bar{\phi}_0), .. \cos(\omega t_m + \bar{\phi}_0))^{\mathrm{T}}$, we obtain the following formula for the estimate at the first iteration:

$$\hat{\phi}_0^{(1)} = \bar{\phi}_0 + \frac{\sum_{i=1}^{m} \cos(\omega t_i + \bar{\phi}_0)(y_i - \sin(\omega t_i + \bar{\phi}_0))}{\sum_{i=1}^{m} \cos^2(\omega t_i + \bar{\phi}_0)}.$$

This estimate corresponds to the minimum point of the parabola (2.41). Assuming further that the measurement errors are random variables, uncorrelated with each other, with identical variances r^2, it is easy to find the formula for the calculated variance corresponding to this estimate:

$$(\sigma^{\mathrm{LSM}}(\bar{\phi}_0))^2 = \frac{r^2}{\sum_{i=1}^{m} \cos^2(\omega t_i + \bar{\phi}_0)}.$$

Taking $x^l = \hat{\phi}_0^{(1)}$ as the next linearization point, after repeated processing of the measurement, we obtain a more accurate description of the behavior of the criterion being minimized in the neighborhood of the extremum point and, hence, a more accurate value of the estimate and the corresponding calculated variance. General equations for the estimate (2.36) and its calculated variance corresponding to the iterated algorithm for this problem take the form:

$$\hat{\phi}_0^{(\gamma+1)} = \bar{\phi}_0 + \frac{\sum_{i=1}^{m} \cos(\omega t_i + \hat{\phi}_0^{(\gamma)}) \left(y_i - \sin(\omega t_i + \hat{\phi}_0^{(\gamma)}) - \cos(\omega t_i + \hat{\phi}_0^{(\gamma)})(\bar{\phi}_0 - \hat{\phi}_0^{(\gamma)}) \right)}{\sum_{i=1}^{m} \cos^2(\omega t_i + \hat{\phi}_0^{(\gamma)})}; \quad (2.42)$$

$$(\sigma^{\mathrm{LSM}}(\hat{\phi}_0^{(\gamma)}))^2 = \frac{r^2}{\sum_{i=1}^{m} \cos^2(\omega t_i + \hat{\phi}_0^{(\gamma)})}.$$

Using such a procedure for calculation of the estimate until the latter does not change significantly any longer, we obtain the estimate with a much less error than that for a single iteration, i.e., with the use of a linearized algorithm.

This situation is illustrated below with the plots. The results of the estimate calculations, their errors and the calculated RMS errors for several iterations are given in Table 2.6. The true value of the phase was assumed to be $\phi_0 = \frac{\pi}{2}$ and the linearization point $\phi_0^{\,l} = \frac{3\pi}{4}$.

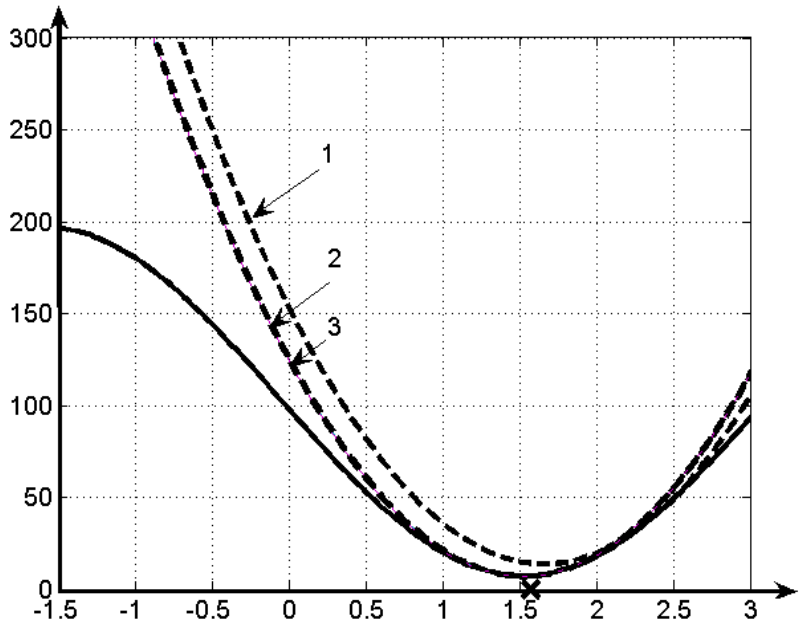


Fig. 2.7. Criterion $J^{:\mathrm{LSM}}(\phi_0)$ and its approximate description for different number of iterations

Note that the calculated variance does not practically change due to small changes of the derivative of function $\sin(\omega t_i + x)$ resulting from the change of the linearization point.

Table 2.6

The values of estimates $\hat{\phi}_0^{(\gamma)}$, their errors $\varepsilon^{(\gamma)}$ and calculated RMS errors $\sigma^{\mathrm{LSM}}(\hat{\phi}_0^{(\gamma)})$ for an iterated algorithm

$\phi_0{}^{l} = \frac{3\pi}{4}$, the true value $\phi_0 = \frac{\pi}{2}$			
Iteration number	$\hat{\phi}_0^{(\gamma)}$	$\varepsilon^{(\gamma)}$	$\sigma^{\mathrm{LSM}}(\hat{\phi}_0^{(\gamma)})$
1	1.6499	0.0791	0.0422
2	1.5233	-0.0475	0.0422
3	1.5164	-0.0544	0.0422
4	1.5161	-0.0547	0.0422
5	1.5161	-0.0547	0.0422

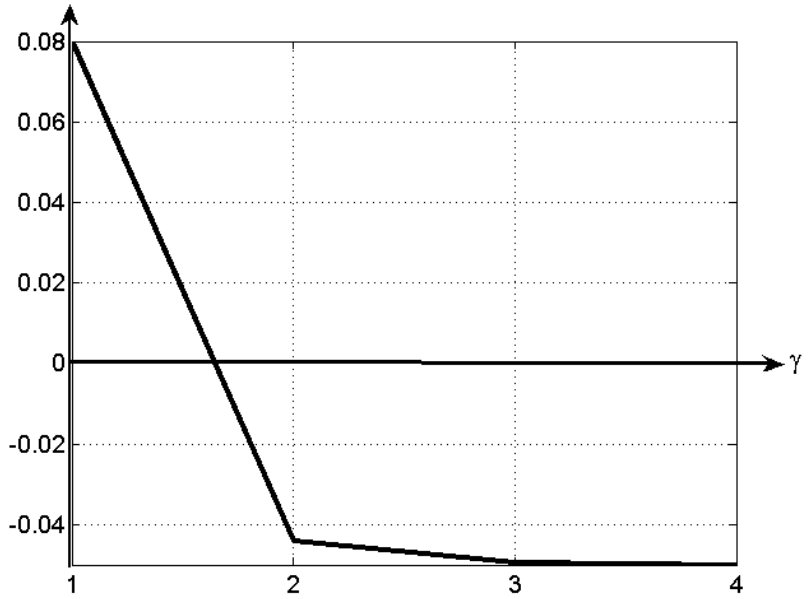


Fig. 2.8. Phase estimate errors vs. iteration number curve

Using the recursion procedure, with respect to the iteration number, we can write:

$$\hat{\phi}_0^{(\gamma+1)} = \hat{\phi}_0^{(\gamma)} + \delta\hat{\phi}_0^{(\gamma+1)},$$

where

$$\delta\hat{\phi}_0^{(\gamma+1)} = \frac{\sum_{i=1}^{m} \cos(\omega t_i + \hat{\phi}_0^{(\gamma)})\left(y_i - \sin(\omega t_i + \hat{\phi}_0^{(\gamma)})\right)}{\sum_{i=1}^{m} \cos^2(\omega t_i + \hat{\phi}_0^{(\gamma)})}. \quad \blacklozenge$$

Linearized and iterated algorithms are widely used to solve the problems of coordinate determination by ranging to beacons. In particular, we concretize these algorithms for solution of such a problem on a plane.

♦ **Example 2.9.** Assume that we have measurements (1.16) to two beacons. For simplicity, it is assumed that one of them is located on the ox_1-axis, and the other one, on the ox_2-axis. To begin with, we write the LSM algorithm based on linearization.

First, note that the criterion to be minimized has the form:

$$J^{\text{LSM}}(x_1, x_2) = \sum_{i=1}^{2}\left(y_i - \sqrt{\left(x_1^i - x_1\right)^2 + \left(x_2^i - x_2\right)^2}\right)^2, \qquad (2.43)$$

and its approximation, corresponding to the linearized description of function $s_i(x) = D_i(x) = \sqrt{\left(x_1^i - x_1\right)^2 + \left(x_2^i - x_2\right)^2}$ in the neighborhood of the linearization point, is a paraboloid:

$$J^{\mathrm{LSM}}(x_1, x_2) \approx \sum_{i=1}^{2}\left(\tilde{y}_i - H_{i1}(x)\left(x_1 - x_1^{\pi}\right) - H_{i2}(x)\left(x_2 - x_2^{\pi}\right)\right)^2, \qquad (2.44)$$

where $H_{i1}(x^l) = -\sin B_i(x^l)$, $H_{i2}(x^l) = -\cos B_i(x^l)$,

$$\tilde{y}_i(x^l) \overset{\Delta}{=} y_i - D_i(x_{\pi}) = y_i - \sqrt{\left(x_1^i - x_1^l\right)^2 + \left(x_2^i = x_2^l\right)^2}\ ,\ i = 1,2.$$

Introducing

$$H = -\begin{bmatrix} \sin B_1(x^l), \cos B_1(x^l) \\ \sin B_2(x^l), \cos B_2(x^l) \end{bmatrix} = \begin{bmatrix} H_1(x^l) \\ H_2(x^l) \end{bmatrix};\ \tilde{y} = \begin{bmatrix} \tilde{y}_1(x_l) \\ \tilde{y}_2(x_l) \end{bmatrix}; \delta x = x - x^l$$

and using (2.9), we can write:

$$\delta\hat{x}^{\mathrm{LSM}} = (H_1^{\mathrm{T}}(x_l)H_1(x_l) + H_2^{\mathrm{T}}(x_l)H_2(x_l))^{-1}(H_1^{\mathrm{T}}(x_l)\tilde{y}_1(x_l) + H_2^{\mathrm{T}}(x_l)\tilde{y}_2(x_l)),$$

where $H_i(x^l) = -\left(\sin B_i(x^l), \cos B_i(x^l)\right)$, $i = 1,2$.

Assuming further that the measurement errors are random variables, uncorrelated with each other, with identical variances r^2, and using Equation (2.25), we derive the following formula for the calculated estimation error covariance matrix:

$$P^{\mathrm{LSM}}(x_l) = r^2\left(H_1^{\mathrm{T}}(x_l)H_1(x_l) + H_2^{\mathrm{T}}(x_l)H_2(x_l)\right)^{-1}.$$

Assume that the linearization point is chosen at the origin of coordinates, i.e., $x^l = 0$. In this case, $\Pi_1(x^l) = 90°$, $\Pi_2(x^l) = 0$ and, therefore, $H = -E$. Taking this fact into consideration, we obtain:

$$\delta\hat{x}^{\mathrm{LSM}} = -\begin{bmatrix} \tilde{y}_1(x_l) \\ \tilde{y}_2(x_l) \end{bmatrix}; \qquad P^{\mathrm{LSM}} = \begin{bmatrix} r^2 & 0 \\ 0 & r^2 \end{bmatrix}.$$

Thus, the estimate of the vehicle's coordinates calculated with the use of the LSM takes the form:

$$\hat{x}_i^{\mathrm{LSM}} = x_i^{\pi} + \delta\hat{x}_i^{\mathrm{LSM}} = x_i^l + D_i(x_l) - y_i,$$

or, since the linearization point was chosen at the origin of coordinates,

$$\hat{x}_i^{\mathrm{LSM}} = \delta\hat{x}_i^{\mathrm{LSM}} = D_i(0) - y_i, \qquad i = 1,2.$$

Now, choosing $x^l = \delta\hat{x}_i^{\mathrm{LSM}}$, repeat the calculations; in so doing, note that at $\delta\hat{x}_i^{\mathrm{LSM}} \neq 0$, matrix H is no longer unitary because $\Pi_1(x^l)$ and $\Pi_2(x^l)$ will be different from 90 degrees and zero, respectively. To implement the iterated algorithm, the described procedures must be repeated until $\delta\hat{x}_i^{\mathrm{LSM}}$ becomes negligible at the next step.

It is clear that the obtained estimates coincide with the GLSM estimates if the weighting matrix Q is chosen diagonal with elements $1/r^2$, $\bar{x} = 0$, and matrix D is zero.

The algorithm is easy to generalize for the case of m measurements. In particular, assuming that $x \equiv \delta x = (x - x^l)$ is a zero-mean vector with covariance matrix P^x, and

the measurement errors v_i are zero-mean random variables, uncorrelated with each other and with vector x, with variances r_i^2, $i=\overline{1.m}$, for the calculated estimation error covariance matrices corresponding to different variants of the LSM, we can write:

$$P^{\text{LSM}}(x^{\prime\prime}) = \left(\sum_{i=1}^{m} M_i(x^l)\right)^{-1} \left(\sum_{i=1}^{m} r_i^2 M_i(x^l)\right) \left(\sum_{i=1}^{m} M_i(x^l)\right)^{-1}; \tag{2.45}$$

$$P^{\text{GLSM}}(x^{\prime\prime}) = \left(\sum_{i=1}^{m} \frac{1}{r_i^2} M_i(x^l)\right)^{-1}; \tag{2.46}$$

$$P^{\text{MLSM}}(x^{\prime\prime}) = \left((P^x)^{-1} + \sum_{i=1}^{m} \frac{1}{r_i^2} M_i(x^l)\right)^{-1}, \tag{2.47}$$

where

$$M_i(x^l) \stackrel{\Delta}{=} \begin{bmatrix} \sin^2 B_i(x^l) & 0.5\sin 2B_i(x^l) \\ 0.5\sin 2B_i(x^l) & \cos^2 B_i(x^l) \end{bmatrix}. \tag{2.48}$$

When $r_i^2 = r^2$, $i=\overline{1.m}$, the above equations are simplified to the form:

$$P^{\text{LSM}}(x^{\prime\prime}) = P^{\text{GLSM}}(x^l) = r^2 \left(\sum_{i=1}^{m} M_i(x^l)\right)^{-1}; \tag{2.49}$$

$$P^{\text{MLSM}}(x^l) = \left((P^x)^{-1} + \frac{1}{r^2} \sum_{i=1}^{m} M_i(x^l)\right)^{-1}. \tag{2.50}$$

When an iterated algorithm is used to obtain the calculated covariance matrix, x^l should be replaced for the value obtained for the last iteration.

Note that the number of measurements and the number of beacons may be different. ♦

From the equations obtained in Example 2.9 it follows that the calculated estimation accuracy in the problem of coordinate determination using beacons is largely determined by the relative position of the beacons. Indeed, introducing the radial RMS error in determining position coordinates σ_{p}, we can write:

$$\sigma_{\text{p}} \equiv DRMS = \sqrt{P^{\text{GLSM}}(1.1) + P^{\text{GLSM}}(2.2)} = r\sqrt{Sp(\sum_{i=1}^{m} M_i(x^l))^{-1}}. \tag{2.51}$$

Hence, it follows that for the identical values of r, σ_p is determined by the product of the RMS range measurement error by the coefficient

$$PDOP = \sqrt{Sp\left(\sum_{i=1}^{m} M_i(x^l)\right)^{-1}}, \tag{2.52}$$

depending on the relative position of point references. This coefficient is called

Position Dilution of Precision. For example, in a particular case of two point references with $\Pi_2=\Pi_1+90°$, it is easy to verify that $\sigma_p \equiv \sqrt{2}r$. It is obvious that the radial RMS error will decrease as the number of references m increases.

Exercises

Exercise 2.1. Assume that we have a criterion specified as

$$J^{\text{MLSM}}(x) = (y - Hx)^{\text{T}} R^{-1}(y - Hx) + (x - \bar{x})^{\text{T}}(P^x)^{-1}(x - \bar{x}). \quad (1)$$

Show that value x, at which this criterion reaches its maximum value, is determined as

$$\hat{x}(y) = \bar{x} + ((P^x)^{-1} + H^{\text{T}} R^{-1} H)^{-1} H^{\text{T}} R^{-1}(y - H\bar{x}).$$

Note. Solve this problem

a) using a set of normal equations;

b) by selecting a perfect square.

Exercise 2.2. Show that in the solution to the vector x estimation problem with measurements $y = Hx + v$, using the LSM and GLSM, the following relations [24] will hold:

$$(y - \hat{y}^{\text{LSM}})^{\text{T}} \hat{y}^{\text{LSM}} = 0,$$

$$(y - \hat{y}^{\text{GLSM}})^{\text{T}} Q\hat{y}^{\text{GLSM}} = 0,$$

where $\hat{y}^{\text{LSM}} = H\hat{x}^{\text{LSM}} = HK^{\text{LSM}} y$; $\hat{y}^{\text{GLSM}} = H\hat{x}^{\text{GLSM}} = HK^{\text{GLSM}} y$.

Exercise 2.3. Assume that in criterion (2.16), $Q = R^{-1}$, $D = (P^x)^{-1}$, and, therefore, in the equation for the estimate $\hat{x}^{\text{MLSM}}(y) = \bar{x} + K^{\text{MLSM}}(y - H\bar{x})$, corresponding to the MLSM, matrix K^{MLSM} is determined as $K^{\text{MLSM}} = ((P^x)^{-1} + H^{\text{T}} R^{-1} H)^{-1} H^{\text{T}} R^{-1}$. Assuming that x and v are uncorrelated random vectors with covariance matrices P^x and R, and the mathematical expectation of vector x is $\bar{x}$, show that the estimate error covariance matrix of the MLSM is determined as

$$P^{\text{MLSM}} = \left(\left(P^x \right)^{-1} + H^{\text{T}} R^{-1} H \right)^{-1}.$$

Exercise 2.4. Calculate the PDOP in the problem of estimating the position on a plane, using measurements (1.16), for the case of two beacons, assuming that the following assumptions are valid: the linearization point is at the origin of coordinates; one of the beacons is on the ox_1-axis, and the other one, on the ox_2-axis; measurement errors are zero-mean random variables, uncorrelated with each other, with similar variances r^2.

Test questions

1. Formulate estimation problem statements using the LSM, GLSM, and MLSM. Explain their main features.

2. What are measurement residuals and a system of normal equations? Give an example of a system of normal equations for a phase estimation problem.
3. Solve the problem of estimating the amplitude of the harmonic signal using the LSM.
4. Derive algorithms for calculation of estimates with the LSM and its modifications for a linear problem in general form. Illustrate the application of these algorithms by a solution to a simple problem of estimating a scalar constant value.
5. Derive equations for errors in estimations with the LSM and its modifications and the equations for the estimation error covariance matrix for the linear estimation problem. What additional assumptions about the properties of the measurement errors should be made to calculate the estimation error covariance matrix when solving the problem with the use of the LSM?
6. Explain how, using measurements $y = s(x) + v$, it is possible to obtain estimation algorithms based on linearization. Illustrate this with estimation of the harmonic signal phase and estimation of coordinates on a plane using range measurements to beacons.
7. What are the main features of the iterated algorithm? Under what conditions does this algorithm provide the estimate corresponding to the LSM?

3. Solution of estimation problems using the non-Bayesian approach

In this chapter we consider **a non-Bayesian approach** and **non-Bayesian algorithms.** Their common feature is the fact that measurement errors are assumed random, and their stochastic properties are assumed known. The latter means that the probability density function (PDF) $p_v(v)$ is known. The unknown estimated vector, as in the LSM and its modifications, is considered to be a nonrandom (determinate) vector [24].

3.1. Fundamentals and statement of the estimation problem in the non-Bayesian approach

If measurement errors are assumed random, with known PDF $p_v(v)$, the measurement can be treated as a random vector whose properties are determined by conditional PDF $p(y/x)$, when x is known. It fully refers to estimate $\tilde{x}(y)$ and its error $\varepsilon(y) = x - \tilde{x}(y)$, which are transformations of measurements y. Taking $y = s(x) + v$ into account and fixing x, the equation for $p(y/x)$ can be written as:

$$p(y/x) = p_v(y - s(x)), \tag{3.1}$$

where $p_v(\bullet)$ is the measurement error PDF.

Thus, assuming that in the equation $y = s(x) + v$, the measurement error is a Gaussian zero-mean vector with known covariance matrix R, $p(y/x)$ can be specified as follows:

$$p(y/x) = \frac{1}{(2\pi)^{m/2}\sqrt{\det R}} \exp\left(-\frac{1}{2}(y - s(x))^{\mathrm{T}} R^{-1} (y - s(x)) \right). \tag{3.2}$$

Moreover, if variables of v_i, $i = \overline{1.m}$ are considered to be random variables, independent of each other, with variances r_i^2, $i = \overline{1.m}$, then $p(y/x)$ is given as

$$p(y/x) = \frac{1}{(2\pi)^{\frac{m}{2}} \sqrt{\prod_{i=1}^{m} r_i^2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{m} \frac{(y_i - s_i(x))^2}{r_i^2} \right). \tag{3.3}$$

The quantitative characteristic of x estimation quality from y measurements can be introduced using scalar function $L(x - \tilde{x}(y))$, setting a certain penalty for the difference between the estimated values and the true ones, referred to as the **loss function**. In navigation data processing, the **quadratic loss function** is most frequently used to analyze the estimation quality:

$$L(x-\tilde{x}(y))=\sum_{i=1}^{n}\left(x_i-\tilde{x}_i(y)\right)^2=(x-\tilde{x}(y))^{\mathrm{T}}(x-\tilde{x}(y))=Sp\left\{(x-\tilde{x}(y))(x-\tilde{x}(y))^{\mathrm{T}}\right\}.$$

Introduce the criterion as a mathematical expectation of this function:

$$J(x)=E_{\mathbf{y}/x}\left\{L(x-\tilde{x}(y))\right\}=E_{\mathbf{y}/x}\left\{(x-\tilde{x}(y))^{\mathrm{T}}(x-\tilde{x}(y))\right\}. \quad (3.4)$$

As operations of the mathematical expectation and calculation of the matrix trace can be rearranged, the criterion can be written as follows:

$$J(x)=E_{\mathbf{y}/x}\left\{Sp(x-\tilde{x}(y))(x-\tilde{x}(y))^{\mathrm{T}}\right\}=Sp\tilde{P}(x), \quad (3.5)$$

where

$$\tilde{P}(x)=E_{y/x}\left\{\left(x-\tilde{x}(y)\right)\left(x-\tilde{x}(y)\right)^{\mathrm{T}}\right\}$$

is the estimate error covariance matrix.

It should be noted that the symbol of mathematical expectation, which depends on x. As a result, this determines the dependences of the matrix and the criterion on argument x.

Therefore, the problem of **algorithm design within the non-Bayesian approach** can be specified as follows. Based on minimization of criterion (3.4), design an algorithm to estimate unknown determinate (nonrandom) vector x from measurements $y=s(x)+\nu$, where ν is the *m*-dimensional random vector of the measurement error with known $p_{\nu}(\nu)$, i.e.,

$$\hat{x}(y)=\arg\min_{\tilde{x}(y)}E_{\mathbf{y}/x}\left\{\sum_{i=1}^{n}(x-\tilde{x}_i(y))^2\right\}.$$

The problem of **accuracy analysis within the non-Bayesian approach** is reduced to calculation of the estimation error covariance matrix (3.5).

Criterion (3.4) is referred to as a **root-mean-square criterion**, and the estimate minimizing it is called an **optimal root-mean-square non-Bayesian estimate.** Note that criterion (3.4) is fundamentally different from the observed criterions from the previous chapter as it deals with satisfying certain requirements for the estimate error rather than for calculated measured parameters.

Unfortunately, for this problem, we cannot specify a general rule for finding estimates that minimize criterion (3.4). Thus, the estimation algorithm is selected by comparing the values of criterions and properties of **the estimate** corresponding to different algorithms. The **estimate** properties, such as unbiasedness, consistency, and efficiency, are very important in comparison. Let us provide definitions of these notions and explain them.

In non-Bayesian approach, estimate $\tilde{x}(y)$ is called unbiased if its mathematical expectation coincides with the true value of parameter x, i.e.,

$$E_{y/x}\{\tilde{x}(y)\}=\int\tilde{x}(y)\,p(y/x)dy=x. \quad (3.6)$$

As a rule, when criterion (3.3) is minimized, an additional requirement on estimate unbiasedness is imposed. An estimate providing minimum of this criterion, when (3.6)

is true, is called **non-Bayesian unbiased estimate with minimum variance.**

To explain the notion of consistency, assume a sequence of scalar measurements $y_i = x + v_i, \quad i = \overline{1.m}$ used to calculate estimate $\tilde{x}_m$. This estimate is called **consistent** if it converges in probability to the true value of the parameter being estimated as the sample size m increases, i.e.,

$$\lim_{m\to\infty} \Pr(x - e < \tilde{x}_m < x + e) = 1, \tag{3.7}$$

where e is an arbitrarily small positive value.

The estimate consistency can be determined in vector case in a similar way.

The notion of estimate efficiency is associated with the so-called **Cramer-Rao inequality**. For unbiased estimates $\tilde{x}(y)$, **the Cramer-Rao inequality within the non-Bayesian approach** is formulated as follows [29, 30]:

$$\tilde{P}(x) \geq I^{-1}(x), \tag{3.8}$$

where

$$I(x) = E_{y/x}\left\{ \frac{\partial \ln p(y/x)}{\partial x} \left(\frac{\partial \ln p(y/x)}{\partial x} \right)^{\mathrm{T}} \right\}. \tag{3.9}$$

It follows from this inequality that matrix $I(x)$ is always equal to or less than the covariance matrix for any unbiased estimate. Matrix $I^{-1}(x)$ determines the **Cramer-Rao Lower Bound (CRLB)** of the estimate error. For inequality to hold true, $p(y/x)$ should meet the regularity requirements, i.e., absolute integrability and existence of the first and second derivatives with respect to x. The estimate for which $\tilde{P}(x) = I^{-1}(x)$ is called an **efficient non-Bayesian estimate**. The matrix in the right-hand part of (3.9) is called a **Fisher information matrix [**24, 29, 30**]**.

If an algorithm to calculate the efficient estimate $\hat{x}(y)$ is designed, the following inequality is true, as follows from (3.8):

$$E_{y/x}\left\{ (x - \tilde{x}(y))(x - \tilde{x}(y))^{\mathrm{T}} \right\} \geq E_{y/x}\left\{ (x - \hat{x}(y))(x - \hat{x}(y))^{\mathrm{T}} \right\}, \tag{3.10}$$

meaning that whatever another algorithm is selected for calculating the unbiased estimate $\tilde{x}(y)$, its estimation error covariance matrix will always be more than or equal to the matrix inverse of the Fisher information matrix.

It follows from the above that the problem of finding the non-Bayesian unbiased estimate with a minimum variance is equivalent to the problem of finding the unbiased efficient estimate if it exists.

Using Cramer-Rao inequality proves very helpful in accuracy analysis since it allows evaluation of **potentially achievable accuracy** without designing the estimation procedure.

The above definitions of the estimate properties make it possible to compare various estimates in more detail.

3.2. Maximum likelihood estimation

In the non-Bayesian approach, the most popular estimator is based on maximizing $p(y/x)$ as a function of x with the known measurement y. In estimation theory, $p(y/x)$ as a function of x with fixed measurements y is called a **likelihood function**, and the estimation method based on its maximization is referred to as the **maximum likelihood method** or **maximum likelihood estimation** [24, 29]. Note that $p(y*/x)$ multiplied by small measurement increment Δy in the scalar case approximately defines the probability $\Pr(y^* < y < y^* + \Delta y) \approx p(y^*/x)\Delta y$, i.e., $\Pr(y^* < y < y^* + \Delta y) \approx p(y^*/x)\Delta y$. Thus, the essence of the procedure for maximizing the likelihood function is to select the value of an unknown parameter at fixed values of the measurements at which this probability reaches its maximum value. Often, instead of the likelihood function, we use its logarithm or the **logarithmic likelihood function** $\ln p(y/x)$**.** These functions are usually defined with accuracy to the arbitrary constant coefficient.

It is known from the estimation theory that the maximum likelihood estimate has a number of important properties: it is consistent; unbiased and normal (Gaussian) in asymptotic approximation with unlimited increase of sample size $m \to \infty$. **Moreover, if an efficient non-Bayesian estimate exists, it is the estimate maximizing the likelihood function** [29]. These features of the maximum likelihood estimate explain its popularity in the **non-Bayesian** approach. However, it should be borne in mind that this estimate is not a general solution to minimizing the criterion for finding unbiased estimates with a minimum variance and, with a limited sample size, it is not always unbiased.

From the above it follows that the **maximum likelihood estimate** is calculated by selecting the value of x maximizing $p(y/x)$, i.e.,

$$\hat{x}^{\text{mlf}}(y) = \arg\max_x p(y/x), \tag{3.11}$$

or

$$\hat{x}^{\text{mlf}}(y) = \arg\max_x \ln p(y/x).$$

To provide the maximum likelihood function, the estimate should meet the necessary maximum condition:

$$\frac{d}{dx} p(y/x)\Big|_{\hat{x}^{\text{mlf}}(y)} = 0,$$

or

$$\frac{d}{dx} \ln p(y/x)\Big|_{\hat{x}^{\text{mlf}}(y)} = 0. \tag{3.12}$$

These equations are called **likelihood equations**.

As with the LSM, (3.12) is the necessary condition to provide the maximum

likelihood function, each solution should be checked for a sufficient condition of the type

$$\frac{d^2}{dxdx^{\mathrm{T}}}\ln p(y/x)\Big|_{\hat{x}^{\mathrm{mlf}}(y)} \le 0. \tag{3.13}$$

Consider an example.

♦ **Example 3.1**. Specify the algorithm to find the maximum likelihood estimate of scalar x by scalar measurements

$$y_i = x + v_i, i = \overline{1.m}, \tag{3.14}$$

where $v_i, i = \overline{1.m}$ are Gaussian random variables, independent of each other, and they have the same variances r^2, i.e., $R = r^2 E$.

For this example, the likelihood function is given by (3.3), and $s(x) = x$, so the estimation algorithm is reduced to minimization of the criterion

$$J^{\mathrm{mlf}}(x) = -\frac{1}{2r^2}\sum_{i=1}^{m}(y_i - x)^2,$$

wherefrom

$$\hat{x}^{\mathrm{mlf}}(y) = \frac{1}{m}\sum_{i=1}^{m} y_i .$$

Thus, in this example, the maximum likelihood estimate is the arithmetic average of all measurements.

Analyze the features of this estimate.

Since $E_{y/x}\left\{\hat{x}^{\mathrm{mlf}}(y)\right\} = \frac{1}{m}E_{y/x}\left\{\sum_{i=1}^{m}(x+v_i)\right\} = x$, the estimate is unbiased, and its error variance is calculated as

$$E_{y/x}\left\{\left(\hat{x}^{\mathrm{mlf}}(y) - x\right)^2\right\} = \frac{1}{m^2}E_{y/x}\left\{\left(\sum_{i=1}^{m} v_i\right)^2\right\} = \frac{r^2}{m}.$$

The error variance tends to zero as the sample size m increases, which means that estimate $\hat{x}^{\mathrm{mlf}}(y)$ is consistent.

Calculate the CRLB in this example. Here, we have

$$\frac{\partial \ln p(y/x)}{\partial x} = H^{\mathrm{T}}\frac{1}{r^2}(y - Hx),$$

where H^{T} is the row consisting of ones.

With account for

$$E_{y/x}\left\{(y - Hx)(y - Hx)^{\mathrm{T}}\right\} = r^2 E_m,$$

we obtain $\tilde{P}(x) \ge \frac{r^2}{m}$.

Thus, in estimating a scalar from measurements (3.14), with the measurement errors being Gaussian random values, independent of each other, with the same variances r^2, the error variance of an unbiased estimate cannot be less than $\frac{r^2}{m}$. It follows that in this example the maximum likelihood estimate is an unbiased efficient estimate, and, therefore, an unbiased estimate with a minimum variance. ♦

3.3. General solution of the linear Gaussian estimation problem using the maximum likelihood method

In Example 3.1, it was rather simple to find the unbiased efficient estimate. Moreover, it can be easily seen that the designed algorithm coincides with the LSM. It is due to the fact that in this example, we solved a linear Gaussian estimation problem. In this case, the solution of the estimation problem using the maximum likelihood method is implemented in the easiest form. Let us illustrate it in detail.

Suppose a linear problem of estimating the constant *n*-dimensional vector $x = (x_1,...x_n)^{\mathrm{T}}$, $\dot{x} = 0$ from the *m*-dimensional measurement vector $y = (y_1,...y_m)^{\mathrm{T}}$, $y = Hx + v$ is solved. In so doing, the measurement errors *v* is a zero-mean random Gaussian vector with covariance matrix R. Let us obtain the maximum likelihood estimate and analyze its features.

With these assumptions, the likelihood function coincides with (3.2), where $s(x) = Hx$, i.e.,

$$p(y / x) = \frac{1}{(2\pi)^{m/2}\sqrt{\det R}} \exp\left(-\frac{1}{2}(y - Hx)^{\mathrm{T}} R^{-1}(y - Hx) \right),$$

and the logarithmic likelihood function can be written as

$$J^{\mathrm{mlf}}(x) = (y - Hx)^{\mathrm{T}} R^{-1}(y - Hx). \qquad (3.15)$$

Therefore,

$$\hat{x}^{\mathrm{mlf}}(y) = \arg\max_x N(y; Hx, R) = \arg\min_x J^{\mathrm{mlf}}(x). \qquad (3.16)$$

Note that criterion (3.15) coincides with criterion (2.13) for the GLSM, with $Q = R^{-1}$, which yields the following:

$$\hat{x}^{\mathrm{mlf}}(y) = K^{\mathrm{mlf}} y; \qquad (3.17)$$

$$P^{\mathrm{mlf}} = K^{\mathrm{mlf}} R (K^{\mathrm{mlf}})^{\mathrm{T}} = \left(H^{\mathrm{T}} R^{-1} H \right)^{-1}, \qquad (3.18)$$

where

$$K^{\mathrm{mlf}} = \left(H^{\mathrm{T}} R^{-1} H \right)^{-1} H^{\mathrm{T}} R^{-1}. \qquad (3.19)$$

Estimate (3.17) is unbiased.

Indeed, while $E_{y/x}\{y\} = Hx$ and $E - K^{\mathrm{mlf}} H = 0$, then $E_{y/x}\left\{(x - K^{\mathrm{mlf}} y)\right\} = 0$.

Estimate error $\varepsilon^{\text{mlf}}(y) = x - \hat{x}^{\text{mlf}}(y)$ can be represented in the form similar to (2.23):

$$\varepsilon^{\text{mlf}}(y) = -K^{\text{mlf}}\nu, \tag{3.20}$$

therefore, it depends only on the measurement errors rather than on the estimated vector x. Thus, the error of the maximum likelihood estimate in the considered problem is invariant to the estimated vector. It should be noted that the covariance matrix does not depend on the estimated vector either.

In this problem, the matrix characterizing the CRLB can be found rather easily, and it can be proven that estimate (3.17) is efficient. Since $s(x) = Hx$, with account for

$$\frac{\partial \mathrm{l}\, p(y/x)}{\partial x} = H^{\mathrm{T}} R^{-1}(y - Hx)$$

and the fact that the regularity conditions are met, the mathematical expectation in (3.9) can be easily calculated, and

$$I(x) = H^{\mathrm{T}} R^{-1} H \tag{3.21}$$

can be obtained for the Fisher information matrix.

Comparison of $I^{-1}(x)$ and covariance matrix $P^{\text{mlf}}(x)$ set by (3.18) shows that they coincide.

Thus, in the linear Gaussian estimation problem, the maximum likelihood estimate (3.17) is an unbiased efficient non-Bayesian estimate with covariance matrix (3.18) or, which is the same, unbiased non-Bayesian estimate with a minimum variance.

Now, discuss the relation to the LSM. Note that equations (3.17), (3.18) are identical to (2.28), (2.29), corresponding to the GLSM with $Q = R^{-1}$. It is quite logical since the problem of minimizing the GLSM criterion coincides with the problem of maximizing the likelihood function as criterion (3.15) agrees with the GLSM criterion.

It follows from the above that the maximum likelihood estimate in the x estimation problem from measurements (1.11) with Gaussian measurement errors coincides with the GLSM estimate if the weight matrix in its criterion is $Q = R^{-1}$.

Since the maximum likelihood estimate coincides with the GLSM estimate, the following statement is true for the GLSM estimates with properly selected matrices.

In the linear Gaussian estimation problem, the GLSM estimate, with the weight matrix in criterion $Q = R^{-1}$, is an unbiased efficient non-Bayesian estimate with covariance matrix (3.18).

In conclusion, we note that in the linear problem (1.10) (1.11), the unbiased estimate with the minimum variance can be actually derived without the PDF of measurement errors. It can be done by minimizing criterion (3.4) in the class of linear estimates and assuming that the two first moments are set for the measurement error ν.

3.4. Solution of the nonlinear Gaussian estimation problem using the maximum likelihood method

Now, discuss the algorithm to find the maximum likelihood estimate in the nonlinear problem (1.20), (1.21) with unknown *n*-dimensional vector $x = (x_1,...x_n)^{\mathrm{T}}$, $\dot{x} = 0$ and *m*-dimensional measurement vector $y = (y_1,...y_m)^{\mathrm{T}}$, $y = s(x) + v$. Here, $s(\bullet) = (s_1(x),..s_m(x))^{\mathrm{T}}$ is a generally nonlinear *m*-dimensional function, $v = (v_1,...v_m)^{\mathrm{T}}$ is an *m*-dimensional vector of the measurement error, that is, a zero-mean random Gaussian vector with covariance matrix R, as in the previous section. Here, the logarithmic likelihood function can be written as

$$J^{\mathrm{mlf}}(x) = \ln p(y / x) = -\frac{1}{2}(y - s(x))^{\mathrm{T}} R^{-1}(y - s(x)). \qquad (3.22)$$

Hence, to find an estimate, we need to search for the maximum of this criterion or to solve the system of nonlinear equations

$$\frac{\partial \ln p(y / x)}{\partial x} = \frac{ds^{\mathrm{T}}(x)}{dx} R^{-1}(y - s(x)) = 0 \qquad (3.23)$$

with a further check of condition (3.13).

It can be easily noted that criterion (3.22) for the considered Gaussian case coincides with criterion (2.5) of the GLSM accurate to the constant factor if weight matrix Q is selected as $Q = R^{-1}$ in the GLSM. This yields the conclusion similar to that for the linear problem. **The maximum likelihood estimate in the x estimation problem from measurements (1.21) with Gaussian measurement errors coincides with the GLSM estimate if the weight matrix in its criterion is $Q = R^{-1}$.**

It follows from the above that the algorithm design methods and accuracy analysis detailed in 2.4 can be interpreted as maximum likelihood estimation methods. Particularly, linearized and iterative algorithms can be used to calculate the estimates. In turn, matrix (2.39) should be used in accuracy analysis. In this case, conditional PDF $p(y / x)$ is used in calculating the mathematical expectation.

Clearly, obtaining this matrix with the use of the Monte Carlo method, for example, requires a large size of computations [6]. Here, the application of Cramer-Rao inequality proves very helpful since in a nonlinear Gaussian problem, similar to the linear case, the matrix characterizing the CRLB can be found rather easily. Assuming that function $s(x)$ meets the regularity conditions and taking

$$\frac{\partial \ln p(y / x)}{\partial x} = \frac{ds^{\mathrm{T}}(x)}{dx} R^{-1}(y - s(x)), \qquad (3.24)$$

we obtain

$$I(x) = \frac{ds^{\mathrm{T}}(x)}{dx} R^{-1} \frac{ds(x)}{dx^{\mathrm{T}}}. \qquad (3.25)$$

It can be seen that if the linearized description of $y = s(x) + v$ is acceptable and the problem is approximately solved by the maximum likelihood method using (3.17), the covariance matrix calculated according to (3.18) (calculated covariance matrix) coincides with $I^{-1}(x)$ while

$$I^{-1}(x) \approx I^{-1}(x_l) = \left(H^{\mathrm{T}}(x_l) R^{-1} H(x_l)\right)^{-1} = P^{\mathrm{mlf}},$$

where $H(x_l)$ is calculated according (1.23).

Let us conctretize the derived equations for two examples.

♦ **Example 3.2.** Obtain the formula for the CRLB in estimating scalar x from measurements

$$y_i = s_i(x) + v_i, i = \overline{1.m}, \tag{3.26}$$

where $v_i, i = \overline{1.m}$ are Gaussian random variables, independent of each other, with the same variances r^2, i.e., $R = r^2 E$.

Using (3.25), write

$$I(x) = \frac{1}{r^2} \sum_{i=1}^{m} \left(\frac{ds_i(x)}{dx} \right)^2 .$$

We introduce the value characterizing the derivative values

$$\bar{g}(x) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left(\frac{ds_i(x)}{dx} \right)^2},$$

then, the equation for the CRLB can be written as

$$P(x) \geq I^{-1}(x) = \frac{r^2}{\bar{g}^2(x) m}. \tag{3.27}$$

Note that $I^{-1}(x) = (\sigma^{\mathrm{LSM}}(x))^2$, where $(\sigma^{\mathrm{LSM}}(x))^2$ is the calculated error variance of the linearized LSM estimate with linearization point $x = x^l$.

The obtained relations can be easily specified for Example 2.8. The value characterizing the derivative is given by

$$\bar{g}^2(x) = \frac{\sum_{i=1}^{m} \cos^2(\omega t_i + x)}{m}. \quad ♦$$

♦ **Example 3.3.** In a nonlinear problem of positioning by ranges (1.16) to *m* beacons, obtain an equation for the CRLB matrix assuming that measurement errors are Gaussian zero-mean random values, independent of each other, with variances r_i^2, $i = \overline{1.m}$.

With account for (3.25), equation $I^{-1}(x)=\left(\sum_{i=1}^{m}\frac{1}{r_i^2}M_i(x)\right)^{-1}$ can be written for the CRLB matrix, where $M_i(x)$ is determined by (2.48), i.e.,

$$M_i(x^l)\overset{\Delta}{=}\begin{bmatrix}\sin^2 B_i(x^l) & 0.5\sin 2B_i(x^l)\\ 0.5\sin 2B_i(x^l) & \cos^2 B_i(x^l)\end{bmatrix}.$$

If linearized description in a priori uncertainty area is true, we will see that the GLSM covariance matrix with $x=x^l$ substituted with the true x, as follows from Example 2.9, coincides with the estimated CRLB matrix. ♦

Generally, when using the Cramer-Rao inequality for accuracy analysis in nonlinear problems, it should be remembered that the real error covariance matrix of the algorithms being studied is set by

$$\tilde{P}=E\left\{(x-\tilde{x}(y))(x-\tilde{x}(y))^{\mathrm{T}}\right\}.$$

This matrix can significantly differ from $I^{-1}(x)$, even if $I^{-1}(x)\approx I^{-1}(x_l)$. Particularly, the difference can be observed in the linearized algorithm while description (1.22) is approximate. At the same time, inequality (3.8), realized as

$$\tilde{P}(x)\geq\left(\frac{ds^{\mathrm{T}}(x)}{dx}R^{-1}\frac{ds(x)}{dx^{\mathrm{T}}}\right)^{-1},$$

is always true if $s(x)$ provides regularity of likelihood function (3.2). It means that whatever algorithm is selected to solve the problem, its error covariance matrix will always exceed the matrix in the right-hand part of inequality calculated with the prescribed value of the unknown vector x to be estimated.

Exercises

Exercise 3.1. Suppose that a scalar x is estimated from measurements of type (3.26) written as

$$y_i=s_i(x)+\varepsilon_i, \tag{1}$$

where the measurement errors

$$\varepsilon_i=d+v_i \tag{2}$$

are a sum of a zero-mean Gaussian random variable with variance σ_d^2 describing the systematic error component and zero-mean Gaussian random variables with the same variances $r_i^2=r^2$, $i=\overline{1.m}$, independent of each other and of d. Obtain the equation for the likelihood function $J^{\mathrm{mlf}}(x)$ and the CRLB.

Exercise 3.2. Show that the equation for $J^{\mathrm{mlf}}(x)$ from Exercise 3.1 can be represented in the form:

$$J^{\text{mlf}} = \sum_{i=1}^{m} \frac{(y_i - s_i(x) - \hat{d}_{i-1}(x))^2}{r^2 + \tilde{\sigma}_{i-1}^2}, \tag{1}$$

where

$$\hat{d}_i(x) = \hat{d}_{i-1}(x) + \frac{\tilde{\sigma}_{i-1}^2}{\tilde{\sigma}_{i-1}^2 + r^2}(y_i - s_i(x) - \hat{d}_{i-1}(x)), \ ; \tag{2}$$

$$\tilde{\sigma}_i^2 = \frac{\tilde{\sigma}_{i-1}^2 r^2}{\tilde{\sigma}_{i-1}^2 + r^2} \quad i = \overline{1.m}; \ \tilde{\sigma}_0^2 = \sigma_d^2, \ \hat{d}_0(x) = 0. \tag{3}$$

Exercise 3.3. Suppose that we have two navigation systems, which generate measurements

$$y_1 = x + v_1,$$

$$y_2 = x + v_2,$$

where $x = (x_1, x_2)^{\text{T}}$ is a two-dimensional vector generating the vehicle plane coordinates. Assume that this vector is determinate (nonrandom) and two-dimensional measurement error vectors v_1, v_2 are zero-mean Gaussian vectors with covariance matrices R_1, R_2.

Obtain the algorithm for an efficient estimate and the relevant covariance matrix. Compare the results with the GLSM solution assuming that Q is a block diagonal matrix with blocks R_1^{-1} and R_2^{-1}.

Exercise 3.4. Consider the problem of estimating the phase of the harmonic oscillation using measurements (1.14) assuming that the amplitude and the frequency are known, and the measurement errors are noncorrelated Gaussian random values with the same variance r^2.

Obtain the linearization and iterative estimation algorithms by the maximum likelihood method. Compare the results with the LSM solution.

Exercise 3.5. Show that maximum likelihood estimates do not depend on linear nonsingular transformations applied to the measurements used in solving the nonlinear Gaussian problem of estimating vector x from measurements (1.21).

Exercise 3.6. Consider a linear problem of estimating vector x from measurements $y = Hx + v$, where the estimates are determined in the form $\hat{x}(y) = Ky$. Obtain the equation for matrix K satisfying the condition $E - KH = 0$ and minimizing criterion (3.4), i.e., $J = E_{y/x}\left\{(x - \hat{x}(y))^{\text{T}}(x - \hat{x}(y))\right\}$.

Beforehand, make sure that in satisfying the above condition to calculate the selected criterion J, it will suffice to assume only the random nature of the measurement error vector v and specify only its two first moments.

Assuming further that v is a zero-mean random vector with known covariance matrix R, obtain the equation for the error covariance matrix.

Test questions

1. Formulate the estimation problem within the non-Bayesian approach. Name the distinctive features of this formulation as compared with the formulation using the LSM and its modifications.
2. Provide definitions of unbiased and efficient estimates, and unbiased non-Bayesian estimate with a minimum variance.
3. Formulate the Cramer-Rao inequality and explain it. What is an efficient estimate?
4. Explain the idea of the maximum likelihood method. What do the likelihood function and likelihood equation mean?
5. Obtain an equation for the Fisher information matrix for linear and nonlinear Gaussian estimation problems.
6. Solve the linear Gaussian estimation problem by the maximum likelihood method. Is the derived estimate efficient? What is the interrelation between the maximum likelihood method and the LSM method?
7. What is the relation of the CRLB matrix with the real error covariance matrix in the nonlinear Gaussian problem and with the calculated covariance matrix derived using a linearized algorithm?
8. Formulate the problem of calculating unbiased non-Bayesian estimate with a minimum variance in the class of linear unbiased estimates. Name the conditions under which this solution coincides with the solution to the problem of calculating an unbiased non-Bayesian estimate without restrictions on the estimate class.

4. Solution of the estimation problem using the Bayesian approach. Optimal estimates

Now consider the estimation problem (1.21) in the context of the Bayesian approach. The main feature of this approach consists in the assumption about the random nature of vector x to be estimated and the measurement error v. This allows for the introduction of the joint PDF $p(x,v)$, which, in turn, makes it possible to proceed to PDF $p(x,y)$, and then, to introduce the conditional (posterior) PDF $p(x/y)$. Taking into consideration the importance of the posterior PDF concept for the estimation problem solved in the context of the Bayesian approach, it will be discussed more comprehensively.

4.1. Bayes formula and posterior probability density function

So, suppose that the joint PDF $p(x,y)$ is known. For $p(x,y)$, the following **formula for** PDF **multiplication** is valid:

$$p(x,y) = p(x/y)\,f(y) = p(y/x)\,p(x). \tag{4.1}$$

In these formulas, $p(x/y)$ and $p(y/x)$ are conditional PDFs that determine the statistical properties of vectors x and y, provided that the vector on the right side of the slash is fixed. In the solution of estimation problems, conditional PDFs $p(x/y)$ and $p(y/x)$ are also called **posterior PDFs or posterior densities**, which emphasizes the fact that these densities correspond to a posterior situation in which one of the vectors associated with the vector being estimated is fixed. Accordingly, functions $p(x)$, $p(y)$ are usually called **prior PDFs or densities**. From (4.1) it follows that

$$p(x/y) = \frac{p(x,y)}{p(y)} = \frac{p(x,y)}{\int p(x,y)dx}. \tag{4.2}$$

Function $p(y) = \int p(x,y)dx$, depending on measurements y, provides normalization, i.e., equality to unity for $\int p(x/y)dx = 1$.

Formula (4.2) is known as the **Bayes formula or Bayes rule**. Equations (4.2) provide a possibility for finding conditional PDF using the known joint PDF $p(x,y)$. From these equations follows that if vectors x and y are independent, conditional and prior PDFs coincide.

Write the equations that allow us to find the parameters of the **conditional Gaussian PDF** [9, 19]. Assume that the joint PDF of the two Gaussian vectors x and

y of n and m dimensions is determined as

$$p(x,y) = N\left(\left(x^{\mathrm{T}}, y^{\mathrm{T}}\right)^{\mathrm{T}}; \left(\overline{x}^{\mathrm{T}}, \overline{y}^{\mathrm{T}}\right)^{\mathrm{T}}, P\right),$$

where

$$P = \begin{bmatrix} P^x & P^{xy} \\ (P^{xy})^{\mathrm{T}} & P^y \end{bmatrix}.$$

It is clear that the PDF for each vector can represented in the following form:

$$p(x) = N\left(x; \overline{x}, P^x\right);$$

$$p(y) = N\left(y; \overline{y}, P^y\right).$$

Using relation (4.2) and the rules for inversion of a block matrix, we can show that the conditional PDF $p(x/y)$ is also Gaussian, i.e., $p(x/y) = N\left(x; \hat{x}(y), P^{x/y}\right)$, and its parameters are determined as

$$\hat{x}(y) = \overline{x} + P^{xy}\left(P^y\right)^{-1}(y - \overline{y}); \tag{4.3}$$

$$P^{x/y} = P^x - P^{xy}(P^y)^{-1}(P^{xy})^{\mathrm{T}}. \tag{4.4}$$

These formulas define the rule for finding the parameters of the conditional Gaussian PDF for the two Gaussian vectors.

♦ **Example 4.1.** Suppose that we have a scalar Gaussian random value x with PDF $p(x) = N(x; \overline{x}, \sigma_0^2)$ and a set of values

$$y_i = s(x) + v_i,\ i = \overline{1.m},$$

in which $s(x)$ is a known, in the general case, nonlinear function. It is assumed that v_i, $i = \overline{1.m}$ are Gaussian random values, independent of each other and x, with PDF $p(v_i) = N\left(v_i; 0, r^2\right)$. Write the equation for posterior PDF. Using (4.1), (4.2), under the assumptions made, we derive:

$$p(x/y) = \frac{1}{c(y)} \exp - \frac{1}{2}\left(\frac{x^2}{P^x} + \frac{1}{r^2}\sum_{i=1}^{m}(y_i - s(x))^2\right),$$

where $c(y)$ is **the normalizing factor** defined as

$$c(y) = \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\frac{x^2}{P^x} + \frac{1}{r^2}\sum_{i=1}^{m}(y_i - s(x))^2\right)\right\} dx.$$

Function $c(y)$ coincides with $p(y)$ to the accuracy of the coefficient. For the special case of $s(x) = x$, in these equations, $s(x)$ should be replaced for x. For this particular case, it is not difficult to show that the posterior PDF is Gaussian $p(x/y) = N\left(x; \hat{x}(y), P^{x/y}\right)$ and its parameters are defined as

$$\hat{x}(y) = \overline{x} + \frac{\sigma_0^2}{r^2 + \sigma_0^2 m}\left(\sum_{i=1}^{m}(y_i - \overline{x})\right);$$

$$P = \left[\frac{1}{\sigma_0^2} + \frac{m}{r^2}\right]^{-1} = \frac{\sigma_0^2 r^2}{r^2 + \sigma_0^2 m}.$$

To be sure that this is true, first, write the parameters of the Gaussian PDF of the composite vector $z = (x, v_1, v_2, ...v_m)^T$, then find the parameters of the Gaussian PDF for the composite vector $\tilde{z} = (x, y_1, y_2, ...y_m)^T = (x, y^T)^T$. It is not difficult to do, taking into consideration that $\tilde{z} = Tz$, where T is the matrix which is easy to determine from the equation $y_i = x + v_i$, $i = \overline{1.m}$. Further, we should use equations (4.3), (4.4). It is easy to make sure that in this example, with $s(x) = x$, function $p(y)$ will be determined as

$$p(y) = N(y; \overline{x}I_{m\times 1}; \sigma_0^2 I_{m\times m} + r^2 E_{m\times m}),$$

where $I_{m\times m}$, $E_{m\times m}$ are the $m \times m$ matrix consisting of unities and a unit matrix. ♦

4.2. Fundamentals, statement, and a general solution of the Bayesian estimation problem

Assume that we need to estimate the *n*-dimensional constant vector x using the *m*-dimensional vector of measurements $y = s(x) + v$, where the unknown vector *x* to be estimated and the measurement error v are random vectors. This allows (see Example 4.1) us to introduce the joint PDF $p(x, v)$, which, in turn, makes it possible to proceed to PDF $p(x, y)$, and then, to introduce the conditional (posterior) PDF $p(x / y)$.

Now, let us formulate the estimation problem solved in the context of the Bayesian approach and give a general solution to this problem.

As in the previous section, we introduce the quadratic loss function:

$$L(x - \tilde{x}(y)) = \sum_{i=1}^{n}\left(x_i - \tilde{x}_i(y)\right)^2 = (x - \tilde{x}(y))^{\mathrm{T}}(x - \tilde{x}(y))$$

and the associated criterion as the expectation of the quadratic loss function:

$$J^B = E_{x,y}\left\{L(x - \tilde{x}(y))\right\} = E_{x,y}\left\langle Sp\left\{(x - \tilde{x}(y))(x - \tilde{x}(y))^{\mathrm{T}}\right\}\right\rangle = Sp\left\{\tilde{P}\right\}.$$

In the Bayesian approach, this criterion is called **the Bayesian risk**. Note that it differs from the similar criterion introduced in the previous section because in this case, the expectation is calculated based on the random characters of both x and y.

In the context of the Bayesian approach, the problem of vector x estimation using measurements y is formulated as follows: find an estimate that will minimize the expected value of the loss function.

The estimate, minimizing the mathematical expectation of the quadratic loss function, is referred to as an **optimal root-mean-square Bayesian estimate**. This

estimate is further called the **optimal Bayesian estimate** or simply the **optimal estimate**.

The following statement is very important in estimation theory: **the optimal Bayesian estimate is a mathematical expectation corresponding to posterior PDF** $p(x/y)$**,** i.e.,

$$\hat{x}(y) = \int xp(x/y)dx. \tag{4.5}$$

Indeed, writing the criterion as

$$J^B = \int \int (x - \tilde{x}(y))^{\mathrm{T}} (x - \tilde{x}(y))\, p(x/y)dx p(y)dy$$

and differentiating it with respect to the estimate, we can write

$$\frac{d}{d\tilde{x}(y)} \int (x - \tilde{x}(y))^{\mathrm{T}} (x - \tilde{x}(y))\, p(x/y)dx = -2\int (x - \tilde{x}(y))^{\mathrm{T}}\, p(x/y)dx = 0.$$

This implies that

$$\int x^{\mathrm{T}} p(x/y)dx = \tilde{x}^{\mathrm{T}}(y) \int p(x/y)dx.$$

Taking into account the normalization condition, we obtain (4.5). Thus, to find the optimal estimate in the Bayesian approach requires the calculation of multiple integral (4.5).

For the accuracy analysis, in the Bayesian approach we use **conditional and unconditional posterior covariance matrices of estimation errors:**

$$P(y) = E_{x|y}\{[x - \hat{x}(y)][x - \hat{x}(y)]^{\mathrm{T}}\} = \int [x - \hat{x}(y)][x - \hat{x}(y)]^{\mathrm{T}}\, p(x/y)dx, \tag{4.6}$$

$$P = E_{x,y}\{[x - \hat{x}(y)][x - \hat{x}(y)]^{\mathrm{T}}\} = \int\int [x - \hat{x}(y)][x - \hat{x}(y)]^{\mathrm{T}}\, p(x,y)dxdy. \tag{4.7}$$

Matrix $P(y)$ determines the estimation accuracy for the certain (prescribed) measurement sample, and matrix P determines the estimation accuracy of the average for the whole measurement ensemble.

It is important to note the following. Unlike the algorithms obtained with the use of the LSM or the non-Bayesian approach, in the Bayesian approach, we can provide not only the algorithm for calculation of the error, but also the procedure used for the calculation of the corresponding covariance matrix $P(y)$, which characterizes the calculated estimation accuracy for a certain measurement sample used to find the estimates. This is of fundamental importance for solving estimation problems in the navigation data processing.

Bearing in mind the above, **in the Bayesian approach,** the term "**the problem of the optimal algorithm design**" is used to mean finding a procedure that provides not only the calculation of estimate (4.5), but also the corresponding accuracy characteristic in the form of conditional posterior covariance matrix (4.6).

The term "**the problem of accuracy analysis**" is used **in the Bayesian approach** to mean obtaining unconditional posterior covariance matrix P (4.7).

4.3. Properties of optimal estimates

Definition 1.

Bayesian estimate $\tilde{x}(y)$ is called **unbiased** if the following equality holds:

$$E_y\{\tilde{x}(y)\} = \overline{x}.$$

Note that this definition differs from the one introduced in considering the non-Bayesian approach. At the same time, it seems quite logical since in this case, the vector to be estimated is assumed random.

Property 1. Estimate (4.5) is unbiased. This property is easy to verify because

$$E_y\{\hat{x}(y)\} = \int \hat{x}(y)\, p_y(y)dy = \iint xp(x \mid y)\, p_y(y)dxdy =$$

$$= \iint xp(y \mid x)\, p_x(x)dydx = \int x \int p(y \mid x)dy)\, p_x(x)dx = E_x\{x\}.$$

Property 2 (property of orthogonality). The error of the optimal estimate is not correlated with (orthogonal to) the measurements, i.e.,

$$E_{x,y}\{(x - \hat{x}(y))y^{\mathrm{T}}\} = 0. \tag{4.8}$$

This equation is easy to prove if we calculate the mathematical expectation sequentially, first, using $p(x/y)$, and then, using $p(y)$. Similarly, we can show that the estimate error is not correlated with (orthogonal to) the optimal estimate, i.e.,

$$E_{x,y}\{(x - \hat{x}(y))\hat{x}^{\mathrm{T}}(y)\} = 0.$$

Property 3. The covariance matrices of estimate error (4.6) (4.7) satisfy the following inequalities:

$$\tilde{P} - P \geq 0\ ; \qquad \tilde{P}(y) - P(y) \geq 0, \tag{4.9}$$

where $\tilde{P}(y), \tilde{P}$ are conditional and unconditional matrices given by the formulas similar to (4.6), (4.7), characterizing the accuracy of $\tilde{x}(y)$.

Recall that for the matrix, $P \geq 0$ means nonnegative definiteness of the corresponding quadratic form. In connection with the above inequality, we can say that the optimal Bayesian estimate minimizes the estimate error covariance matrix, whereas the posterior covariance matrix P itself characterizes **the potential accuracy of optimal estimation in the Bayesian approach.** Accordingly, the diagonal elements of these matrices define the potential accuracy of the estimation components x_j, $j = \overline{1.n}$.

Property 4. Optimal estimate (4.5) minimizes determinants of matrices P and $P(y)$ [24, 29].

Property 5. Assume that the m-dimensional vector $\tilde{x}$ is the linear transformation of the n- dimensional vector x, that is, $\tilde{x} = Tx$, where T is the known $m \times n$ matrix. The optimal estimate of vector $\tilde{x}$ is defined as $\hat{\tilde{x}}(y) = T\hat{x}(y)$, where $\hat{x}(y)$ is the optimal estimate of vector x.

This is easy to verify using (4.5).

Definition 2.

In the context of the Bayesian approach, we also introduce the notion of estimate efficiency, which, as in the non-Bayesian approach, follows from the **Cramer-Rao inequality**. Within Bayesian approach, this inequality is formulated as follows [29, 30]:

$$P = E_{x,y}\{(x - \hat{x}(y))(x - \hat{x}(y))^{\mathrm{T}}\} \geq (I^B)^{-1}, \tag{4.10}$$

where I^B is defined as

$$I^B = E_{x,y}\left\{\frac{\partial \ln p_{x,y}(x,y)}{\partial x}\left(\frac{\partial \ln p_{x,y}(x,y)}{\partial x}\right)^{\mathrm{T}}\right\}. \tag{4.11}$$

The following sequence of inequalities is valid:

$$\tilde{P} \geq P \geq (I^B)^{-1}. \tag{4.12}$$

Estimate $\hat{x}(y)$ is called **an effective Bayesian estimate** if in (4.12), the inequality sign becomes equality. For inequality to hold true, $p(x,y)$ should meet the regularity requirements, i.e., absolute integrability and existence of the first and second derivatives with respect to x and y.

The sign of more or equal in the inequality $P \geq (I^B)^{-1}$ means that the optimal Bayesian estimate does not always have to be efficient. This matrix $(I^B)^{-1}$ determines the CRLB, and thus, it characterizes **the potentially achievable accuracy of optimal estimation in the Bayesian approach.** Matrix I^B is called **the Fisher information matrix**. Note that in the Bayesian approach, the CRLB is prescribed for the unconditional covariance matrix. The ratio between $(I^B)^{-1}$ and matrix $P(y)$, which for a nonlinear problem depends on measurement, is not prescribed.

4.4. Solution of the linear Gaussian estimation problem. Interrelation with the least squares method

It is rather simple to solve the estimation problem in the linear Gaussian case within the Bayesian approach.

Suppose it is required to find the optimal Bayesian estimate of vector x using measurements (1.2), i.e., $y = Hx + v$, where x and v are assumed Gaussian vectors. In view of the Gaussian nature of x, v and linearity of measurements, the joint PDF $p(x,y)$, and consequently, the posterior density will also be Gaussian.

To find this PDF, we should act in the same manner as in the example considered in 4.1: write the joint PDF for vector $z = (x^T, v^T)^T$, then find the parameters of the Gaussian PDF of vector $\tilde{z} = (x^T, y^T)^T$, taking into consideration that $\tilde{z} = Tz$, and then, find the parameters of the conditional density PDF. Calculating the values of $\bar{y} = H\bar{x}$,

P^{xy}, P^{y}, it is easy to specify the formulas for estimation of the covariance matrix. In the special case when vectors x and v are independent of each other, i.e., $p(z) = p(x)p(v)$, taking into consideration the notation used $P^{v} = R$, $P^{x/y} = P$, we have:

$$\overline{y} = H\overline{x}; \quad P^{xy} = P^{x}H^{\mathrm{T}}; \qquad P^{y} = HP^{x}H^{\mathrm{T}} + R.$$

Hence, it follows:

$$\hat{x}(y) = \overline{x} + K(y - H\overline{x}); \tag{4.13}$$

$$K = P^{x}H^{\mathrm{T}}\left(HP^{x}H^{\mathrm{T}} + R\right)^{-1}; \tag{4.14}$$

$$P = P^{x} - P^{x}H^{\mathrm{T}}\left(HP^{x}H^{\mathrm{T}} + R\right)^{-1}HP^{x} = (E - KH)P^{x}. \tag{4.15}$$

Using the matrix inversion lemma [5]

$$\left[P^{-1} + H^{\mathrm{T}}R^{-1}H\right]^{-1} = P - PH^{\mathrm{T}}\left(HPH^{\mathrm{T}} + R\right)^{-1}HP,$$

matrix P and K can be represented in the form:

$$P = \left(\left(P^{x}\right)^{-1} + H^{\mathrm{T}}R^{-1}H\right)^{-1}; \tag{4.16}$$

$$K = PH^{\mathrm{T}}R^{-1}. \tag{4.17}$$

In the linear Gaussian problem, the optimal, in the RMS sense, the estimate is efficient because

$$\frac{\partial \ln p_{x,y}(x,y)}{\partial x} = -\left(\left(P^{x}\right)^{-1}(x - \overline{x}) - H^{\mathrm{T}}R^{-1}(y - s(x))\right).$$

Substituting this equation into (4.11), we can write

$$P = P(y) = (I^{B})^{-1} = \left(\left(P^{x}\right)^{-1} + H^{\mathrm{T}}R^{-1}H\right)^{-1}. \tag{4.18}$$

It is important to note that the posterior PDF in this problem is Gaussian, and the algorithm for calculation of the optimal Bayesian estimate is linear relative to measurements y.

It is also essential that the conditional covariance matrix $P(y)$ does not dependent on measurements and, therefore, it will coincide with the unconditional covariance matrix P, which does not hold in the general nonlinear case.

Find the interrelation between the optimal estimates and the estimates sought by the LSM. Verify the validity of the following statement.

Proposition. The estimates obtained by the modified least squares method (MLSM) coincide with optimal Bayesian estimates in the linear Gaussian problem, when, first, there is no correlation between the vector to be estimated and the vector of measurement errors and second, the criterion in the MLSM is chosen such as $\overline{x}$ is the mathematical expectation of vector x, and matrices $D = (P^{x})^{-1}$, $Q = R^{-1}$.

As it was noted above, the posterior PDF in the linear Gaussian problem is also Gaussian, and therefore, it is a symmetrical function and the corresponding expectation $\hat{x}(y)$, i.e., the optimal estimate (4.5) coincides with the value of x at which this PDF reaches its maximum value.

Write the following formula for the joint PDF:

$$p(x,y) = p(y/x)\,p(x) = p_v(y - Hx)\,p(x).$$

Thus, for the posterior PDF we will have:

$$p(x/y) = c\exp\left\{-\frac{1}{2}\left((y-Hx)^{\mathrm{T}} R^{-1}(y-Hx) + (x-\bar{x})^{\mathrm{T}}\left(P^{x}\right)^{-1}(x-\bar{x})\right)\right\}, \quad (4.19)$$

where c is the normalization factor which is independent of x.

Taking into consideration this representation for $p(x/y)$, it is not difficult to understand that maximization of the posterior density is equivalent to minimization of the criterion

$$J(x) = (y-Hx)^{\mathrm{T}} R^{-1}(y-Hx) + (x-\bar{x})^{\mathrm{T}}(P^{x})^{-1}(x-\bar{x}), \quad (4.20)$$

which coincides with the criterion used in the MLSM, provided that matrices Q и D are chosen properly. It is this fact that explains the coincidence of the MLSM estimate at independent x and v with the optimal Bayesian estimate, which, in turn, coincides in the linear Gaussian problem with the linear optimal estimate. Accordingly, the covariance matrices of errors also coincide.

4.5. Linear optimal estimates. Statement of the problem and its general solution

Often, for simplicity of the resultant estimation algorithm, researchers introduce restrictions on the class of estimates. For example, it is often thought that the Bayesian estimate being sought is linearly dependent on measurements, i.e., it is defined as

$$\tilde{x}(y) = \bar{x} + K(y - \bar{y}). \quad (4.21)$$

Bearing this in mind, the problem of vector x estimation using measurements y can be formulated as follows: find an estimate that minimizes the mathematic expectation of the quadratic loss function in the class of linear estimates of the form (4.21). It is easy to verify that the estimate defined in the form (4.21) is unbiased.

Thus, the problem is to find the linear unbiased estimates with a **minimum variance**, or in other words, the problem of finding optimal, in the RMS sense, linear estimates. Further, for brevity, we will simply speak about **optimal linear estimates**. The algorithm that provides for finding of such estimates is called an **optimal linear algorithm.**

Substituting (4.21) in the formula for criterion J^{B}, we can write

$$J^B = E_{x,y}\left\{(x-\bar{x}-K(y-\bar{y}))^{\mathrm{T}}(x-\bar{x}-K(y-\bar{y}))\right\} =$$

$$= E_{x,y}\left\{Sp[\left(x-\bar{x}+K(y-\bar{y})\right)\left(x-\bar{x}+K(y-\bar{y})\right)^{\mathrm{T}}]\right\} =$$

$$= SpE_{x,y}\{\ (x-\bar{x})(x-\bar{x})^{\mathrm{T}} + K(y-\bar{y})(x-\bar{x})^{\mathrm{T}} +$$

$$+(x-\bar{x})(y-\bar{y})^{\mathrm{T}}K^{\mathrm{T}} + K(y-\bar{y})(y-\bar{y})^{\mathrm{T}}K^{\mathrm{T}}\ \} =$$

$$= Sp\left\{P^x + KP^{yx} + P^{xy}K^{\mathrm{T}} + KP^yK^{\mathrm{T}}\right\}. \qquad (4.22)$$

Thus, the algorithm for the estimate calculation is reduced to the problem of parametric optimization of criterion J^B with respect to matrix K. In other words, finding of the posterior PDF itself is not necessary here.

The following statement is valid.

For the linear estimate (4.21) in the problem solution of vector x estimation with the use of the measurement vector y to provide a minimum of criterion (4.22), it is necessary and sufficient that matrix K^{lin}, used in the calculation of this estimate, should satisfy this equation.

$$K^{lin}P^y = P^{xy}. \qquad (4.23)$$

This equation can be treated as a simplest variant of the **Wiener-Hopf equation [**19, 29**]**.

If matrix P^y is nonsingular, then from (4.23) it follows that

$$K^{lin} = P^{xy}(P^y)^{-1} \qquad (4.24)$$

and, therefore,

$$\hat{x}(y) = \bar{x} + K^{lin}(y-\bar{y}) = \bar{x} + P^{xy}(P^y)^{-1}(y-\bar{y}). \qquad (4.25)$$

Using (4.25), it is not difficult to write the equation for the covariance matrix:

$$P^{lin} = E_{x,y}\{(x-\hat{x}(y))(x-\hat{x}(y))^T\} =$$

$$= P^x + K^{lin}P^y(K^{lin})^T - K^{lin}P^{yx} - P^{xy}(K^{lin})^T.$$

Transforming this equation with the use of (4.24), we obtain

$$P^{lin} = P^x - P^{xy}(P^y)^{-1}P^{yx} = P^x - K^{lin}P^{yx}. \qquad (4.26)$$

Thus, the procedure for the calculation of optimal, in the RMS sense, linear estimates is given by (4.24)–(4.26). In this case, to design the algorithm, it is necessary to have a priori expectations $\bar{x}$ and $\bar{y}$ and matrices P^x, P^y, P^{xy}. Knowledge of these matri**ces also provides a solution to the problem of accuracy analysis** since they are used to find a posteriori error covariance matrix P^{lin} corresponding to the optimal linear algorithm.

In the particular case when in the linear problem in the measurements $y = Hx + v$, vectors x and v are not correlated, and v is zero mean, it is easy to verify that (see Exercise 4.2) formulas (4.25), (4.26) for the optimal linear estimation and the covariance matrix of its errors are reduced to formulas (4.13)–(4.17). Recall that these

formulas define the algorithm for finding the optimal estimate sought without any restriction on the class of estimates in the form of conditional expectation (4.5).

Thus, we have come to the following important and valid conclusion.

In the linear Gaussian problem, the optimal linear estimate and the optimal estimate coincide.

Consider an example.

♦ **Example 4.2.** Find the formula for the optimal linear estimate and the corresponding variance of the scalar random variable x uniformly distributed on the interval $[0,b]$, using the measurements $y_i = x + v_i$, $i = \overline{1.m}$, in which v_i are random values, independent of each other and x, uniformly distributed on the interval $[0,a]$.

As applied to the problem under consideration, for the optimal linear estimate and the corresponding variance we obtain:

$$\hat{x}^{lin}(y) = \bar{x} + \frac{\sigma_0^2}{r^2 + \sigma_0^2 m}\sum_{i=1}^{m}(y_i - \bar{x} - \bar{v}),$$

$$P^{lin} = \frac{\sigma_0^2 r^2}{r^2 + \sigma_0^2 m},$$

which, after substituting the values of expectations $\bar{x} = b/2$, $\bar{v} = a/2$ and variances $\sigma_0^2 = b^2/12$, $r^2 = a^2/12$, can be written as follows:

$$\hat{x}^{lin}(y) = \frac{b}{2} + \frac{b^2}{(a^2 + b^2 m)}\sum_{i=1}^{m}(y_i - b/2 - a/2)\ ; \qquad P^{lin} = \frac{a^2 b^2}{12(a^2 + b^2 m)}.$$

In particular, at $a << b$, we have

$$\hat{x}^{lin}(y) \approx \frac{1}{m}\sum_{i=1}^{m}(y_i - a/2);\ \ P^{lin} \approx \frac{a^2}{12m}.$$

The above algorithm is an optimal linear algorithm not only for random variables with uniform distribution, but also for any zero-mean random variables with an arbitrary distribution and prescribed variances $\sigma_0^2 = b^2/12$, $r^2 = a^2/12$. In particular, this algorithm will be optimal for zero-mean Gaussian random variables with the same variances. ♦

In view of the above said, we can formulate the following conclusion.

The algorithm for calculation of optimal linear estimates in a linear problem is fully determined by the first two moments for the composite vector $z = \left(x^{\mathrm{T}} \quad , v^{\mathrm{T}}\right)^{\mathrm{T}}$ and it does not depend on the PDF $p(x,v)$. In other words, the linear optimal algorithm in a linear problem does not depend on the PDF of the vector being estimated and measurement errors. Note that in a nonlinear problem, this statement is not valid.

We should emphasize that the estimation problem of vector *x* using measurements *y*,

considered in the framework of the Bayesian approach, is closely related to the so-called **regression problem**. The essence of this problem is to describe the properties of one random vector x at a fixed value of another vector y, statistically dependent on x. From the mathematical point of view, this problem reduces to obtaining the best approximation, in some sense, of vector x using function $\tilde{x}(y)$ in the presence of a joint PDF $p(x,y)$.

If the regression problem is formulated as the problem of finding function $\tilde{x}(y)$ that minimizes the Bayesian criterion J^B, then it is obvious that its solution will coincide with the estimation problem solution and will be defined by formula (4.5). The most popular is the **linear regression problem**, i.e., one in which the finding of $\tilde{x}(y)$ minimizing J^B is carried out in the class of functions linearly dependent on the vector y. To solve this problem, it will suffice to have information only about the first two moments of the composite vector, including x and y. It is clear that in this case the solution of the linear regression problem will be determined by formula (4.25). The difference between the estimation problem and the regression one lies only in the fact that the regression problem does not use the functional dependence (1.2). In this sense, the estimation problem under consideration can be interpreted as a special case of the regression problem in the presence of the functional dependence (1.2).

4.6. Improvement of estimation accuracy by using a nonlinear algorithm

Generally, the estimation accuracy provided by nonlinear algorithms can be improved as compared with the accuracy of optimal linear algorithms. Let us illustrate this with Example 4.2 considered in the previous section. With this aim, we obtain an optimal estimate for this example. First, let us write the formulas for $p(x,y) = p(y/x)\,p(x)$ and $p_y(y)$. In view of the specific character of the problem, these functions can be represented in the form:

$$p(x,y) = \begin{cases} \dfrac{1}{b}\,p(y/x), x \in [0,b], \\ 0 \quad , \qquad x \notin [0,b], \end{cases}$$

$$p(y) = \int p(y/x)\,p(x)dx = \frac{1}{b}\int_0^b p(y/x)dx\,.$$

To obtain values of $p(y)$ at fixed values of measurements, we need to know $p(y/x)$. We can write:

$$p(y/x) = \prod_{i=1}^{m} p_v(y_i - x)\,,$$

where $$p(y_i / x) = p_v(y_i - x) = \begin{cases} \frac{1}{a}, x \in [y_i - a, y_i] \\ 0, x \notin [y - a, y] \end{cases}, \; i = \overline{1.m},$$

and hence,

$$p(y / x) = \begin{cases} c^*, x \in \Omega, \\ 0, x \notin \Omega. \end{cases} \tag{4.27}$$

In this formula, c^* is a certain constant, and domain Ω is a segment formed by the intersection of all intervals $[y_i - a, y_i]$, $i = \overline{1.m}$, i.e.,

$$\Omega \equiv \bigcap_{l=1}^{m} [y_i - a, y_i] = [d_1, d_2] = [y_{\max} - a, y_{\min}]. \tag{4.28}$$

The bounds of this interval $d_1 = y_{\max} - a$ and $d_2 = y_{\min}$ are determined by the maximum $y_{\max}$ and minimum $y_{\min}$ measured values. Taking into account (4.27), (4.28) and the fact that

$$\hat{x}(y) = \frac{\int_0^b x p(y / x) dx}{\int_0^b p(y / x) dx},$$

for the optimal Bayesian estimate we can write:

$$\hat{x}(y) = \frac{1}{c_2 - c_1} \int_{c_1}^{c_2} x dx = \frac{1}{c_2 - c_1} \left(\frac{x^2}{2} \Big|_{c_1}^{c_2} \right) = \frac{(c_2 + c_1)}{2}, \tag{4.29}$$

where $[c_1, c_2]$ is a segment representing an intersection of the prior domain $[0, b]$ and domain Ω, so that

$$c_1 = \max\{0, d_1\}; \qquad c_2 = \min\{b, d_2\}.$$

It is important to emphasize that estimate (4.29) is nonlinearly dependent on measurements. Formula (4.29) is a consequence of the fact that the posterior PDF in this problem corresponds to the uniform distribution on the interval $[c_1, c_2]$.

In view of the above said, it is easy to show that

$$P(y) = \int_0^b (x - \hat{x}(y))^2 p(x / y) dx = \frac{(c_2 - c_1)^2}{12}. \tag{4.30}$$

In particular, if $d_1 = y_{\max} - a$ and $d_2 = y_{\min}$ do not go beyond the prior domain $[0, b]$, then, for the optimal estimate and the conditional variance, the following formula is valid:

$$\hat{x}(y) = \frac{y_{\min} + y_{\max}}{2} - \frac{a}{2}; \qquad (4.31)$$

$$P(y) = \int_0^b (x - \hat{x}(y))^2 \, p(x / y) dx = \frac{(y_{\min} + a - y_{\max})^2}{12}.$$

From (4.31) it actually follows that the optimal estimate is the arithmetic mean of the maximum and minimum values of measurements calculated with the account of the known mathematic expectation for measurement errors.

In order to compare the accuracies of the linear and nonlinear estimates, the method of statistical tests was used to calculate the values of the RMS errors $\sigma^{opt}(m)$ and $\sigma^{lin}(m)$ characterizing the accuracy of the nonlinear and linear estimates.

Table 4.1 and Fig. 4.1 show the results obtained at $b = 1$, $a = 0.1$ and different number of measurements $m = 1, 2 ..... 100$, with the number of samples being $L = 1000$.

Table 4.1

The values of the RMS errors of the optimal linear and nonlinear estimates

Number of measurements	10	20	100
Optimal linear estimate	0.0091	0.0065	0.0029
Optimal nonlinear estimate	0.0064	0.0034	0.00075

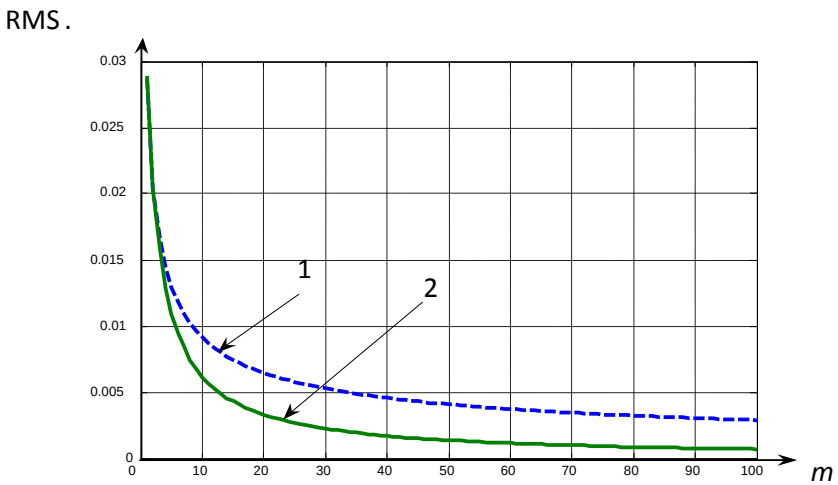


Fig. 4.1. RMS values for the optimal linear (*1*) and nonlinear (*2*) algorithms at different numbers of measurements

From the results obtained it follows that the accuracy of the optimal nonlinear algorithm is significantly higher than that of the algorithm which is optimal in the class of linear algorithms, which is consistent with the statement given at the beginning of the section. Formally, this difference is explained by the fact that the form of the posterior PDF (Fig. 4.2) is substantially different from the Gaussian one.

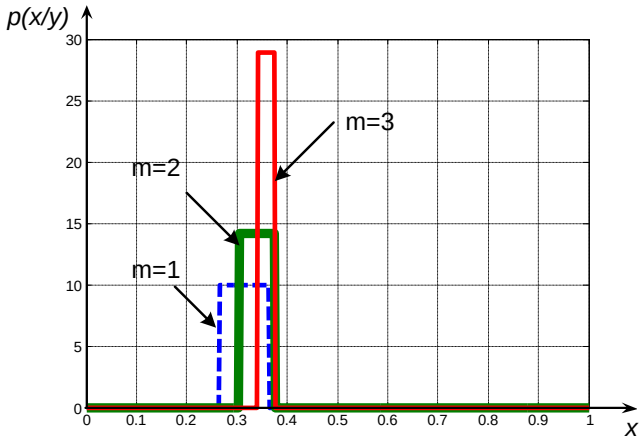


Fig. 4.2. Posterior PDF at m=1, 2, 3

We can give the following explanation for this difference. Note that if the condition $a^2 << b^2 m$ is satisfied, the algorithm for calculation of the optimal linear estimate will virtually be reduced to finding the arithmetic mean for all values of $y_i - a/2$, $i = \overline{1.m}$ that belong to the segment $[y_{\min} - a/2, y_{\max} - a/2]$. At the same time, in the optimal nonlinear algorithm, the posterior domain of the values $[c_1, c_2]$ is determined by the intersection of the prior domain $[0, b]$ and the segment $[y_{\min}, y_{\max} - a]$.

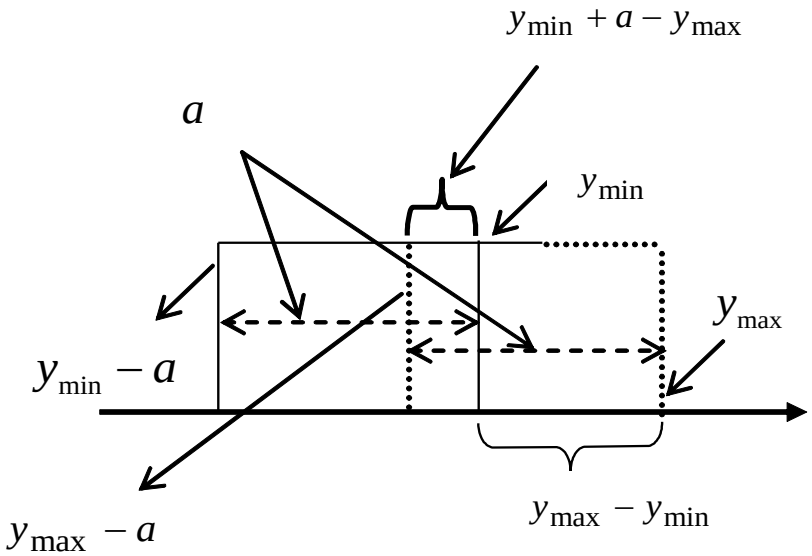


Fig. 4.3. Formation of the posterior domain for the optimal nonlinear estimate

It is obvious that as the number of measurements increases, the probability of a measurement implementation with a minimum (close to zero) and a maximum (close to a) error increases, i.e., $y_{\min} \to x$, a $y_{\max} \to x + a$. Hence, it follows that the length of

the segment $y_{\max} - y_{\min}$ that determines the domain of the values of all measurements used in calculating the arithmetic mean, tends to the value a. At the same time, the length of the segment $y_{\min} + a - y_{\max}$ which determines the size of the posterior domain for the optimal nonlinear estimate, will tend to zero (Fig. 4.3).

Exercises

Exercise 4.1. Write the formulas for the optimal estimate and its variance for the problem of estimating a scalar random variable x using measurements $y_i = x + v_i$, when x and v_i, $i = \overline{1.m}$ are independent of each other, and $p(x) = N(x; \bar{x}, \sigma_0^2)$, $p(v) = N(v; 0, R)$, where R is a diagonal matrix with elements r_i^2, $i = \overline{1.m}$. Simplify the obtained formulas for the case when $r_i^2 = r^2$, $i = \overline{1.m}$ and compare them with the similar formulas in the MLSM.

Exercise 4.2. Find such PDFs for measurement errors and scalar x being estimated in the conditions of Example 4.1 under which the obtained algorithm will be an optimal algorithm.

Exercise 4.3. Show that for the optimal Bayesian estimate (4.5) of zero-mean vector x, the following equation is valid:

$$E_{x,y}\left\{\hat{x}(y)\hat{x}^T(y)\right\} = P^x - P,$$

where P^x and P are prior and posterior unconditional covariance matrices.

Exercise 4.4. Solving the problem of finding the optimal linear unbiased estimate, as it is formulated in 4.4, and assuming that vectors x and y are zero mean, i.e., $\bar{x} = 0$ and $\bar{y} = 0$, prove the following statement.

For the linear estimate $\tilde{x}(y) = Ky$ of vector x with the use of measurements (1.21) to provide the minimum criterion (4.24), it is necessary and sufficient that matrix K^{lin}, used in the calculation of this estimate, should satisfy the equation

$$K^{lin} P^y = P^{xy}. \qquad (1)$$

Exercise 4.5. Write the formulas for the optimal, in the RMS sense, linear estimate and the corresponding posterior error covariance matrix, assuming that we have measurement $y = Hx + v$ and the linear estimation problem is to be solved under the conditions where the vector of parameters x to be estimated and the measurement error vector v are assumed random correlated vectors with zero expectations for them, and the covariance matrix is given in the form:

$$P^{x,v} = \begin{bmatrix} P^x & B \\ B^{\mathrm{T}} & R \end{bmatrix}.$$

Of essential importance here is the fact that the type of the PDF is not assumed to be Gaussian; moreover, it is considered that the type of this density is not specified.

Exercise 4.6. Obtain the formula for the optimal, in the RMS sense, linear estimate

and the corresponding covariance matrix, assuming that the previous exercise 4.5 is to be solved in the conditions where there are no measurement errors, and $B = 0$.

Exercise 4.7. Obtain the formula for the optimal, in the RMS sense, linear estimate and the corresponding covariance matrix, assuming that exercise 4.5 is to be solved at $B = 0$, and measurements are given in the form: $y = Hx + v + u$, where u is the known m-dimensional vector.

Test questions

1. What is posterior PDF and how can we find it in the estimation problem?
2. Formulate the problem statement for obtaining the optimal, in the RMS sense, estimate of vector x using measurements $y = s(x) + v$ without imposing restrictions on the class of the estimates used; give its general solution.
3. Call the properties of optimal estimates.
4. How is the problem statement for obtaining the optimal in the RMS sense, linear estimate of vector x, using measurements y, modified as compared with the problem statement that does not impose any restrictions on the class of the estimates used?
5. Give the necessary and sufficient conditions of optimality for the linear estimate of vector x using measurements y. Describe the algorithm for calculating the optimal linear estimate.
6. Under what assumptions will the optimal, in the RMS sense, estimate of vector x using measurements $y = Hx + v$ coincide with the estimates obtained with the use of the modified LSM?
7. Give an example of a problem in which the estimation error variance for the optimal algorithm is significantly lower than that for the linear optimal algorithm?

5. Algorithms of integrated measurement processing

This section discusses the problem of integrated processing of measurements and design of various algorithms used to solve this problem.

The problem of integrated (joint) processing of measurements consists in obtaining the estimate of the sought vector using measurements from all available sensors and systems.

On the one hand, the purpose of this section is to show that the problem of integrated processing of measurements is essentially an ordinary estimation problem. On the other hand, we are going to discuss the special features of the algorithms used to solve the problem of integrated measurement processing depending on the level of the available a priori information about the stochastic properties of unknown parameters and measurement errors.

5.1. Statement of the integrated measurement processing problem

The simplest case where we deal with the problem of integrated measurement processing is finding an n-dimensional vector of unknown parameters x by data from two sensors (systems) whose outputs can be presented as

$$y_1 = x + v_1;$$

$$y_2 = x + v_2,$$

where v_1, v_2 are the errors of the sensors. Both sensors or systems are supposed to measure the same parameters. For example, various navigation systems, such as inertial and satellite systems, generate position and velocity of a vehicle, etc.

It can be easily noted that this problem is a particular case of problem (1.10), (1.11), i.e., estimation of x from measurements $y = Hx + v$. It can be proved by setting $y = (y_1^{\mathrm{T}}, y_2^{\mathrm{T}})^{\mathrm{T}}, v = (v_1^{\mathrm{T}}, v_2^{\mathrm{T}})^{\mathrm{T}}$, $H = \begin{bmatrix} E_{n\times n} \\ E_{n\times n} \end{bmatrix}$, where $E_{n\times n}$ is an $n \times n$ unity matrix.

Thus, the problem of integrated processing of measurements from two sensors (systems) is reduced to the standard statement of the linear estimation problem.

Consider another example of the integrated processing problem. Suppose that x should be estimated from two sets of measurements:

$$y_1 = x + v_1,;$$

$$y_2 = \tilde{s}(x) + v_2,$$

where y_2 is an l-dimensional vector; $\tilde{s}(x) = (\tilde{s}_1(x), ... \tilde{s}_l(x))^{\mathrm{T}}$ is an l-dimensional nonlinear vector-function depending on n-dimensional unknown vector x; v_1 and v_2

are the vectors of the corresponding dimensions.

Measurement y_1 provides information directly on all components of the sought vector, and measurement y_2 provides some function of this vector x. It should be noted that the dimensionality of measurement y_2 may not coincide with the dimensionality of the estimated vector and can be arbitrary.

Introducing $l+n$-dimensional vector-function $s(x)$ and vector v

$$s(x)=\begin{bmatrix} x \\ \tilde{s}(x) \end{bmatrix},\ v=\begin{bmatrix} v_1 \\ v_2 \end{bmatrix},$$

we come to the nonlinear estimation problem (1.20), (1.21) formulated in 1.7, i.e., to estimation of x from measurements $y=s(x)+v$.

Thus, the problem of integrated measurement processing in this case can be also treated as a standard problem but already nonlinear estimation problem. As before, the attention will be focused on linear problems or problems that can be reduced to linear ones. After linearization of function $s(x)$, for example, at the point $x^l=0$, the problem under consideration can be reduced to a linear estimation problem, namely, estimation of x from $y=Hx+v$. Indeed, it can be easily done if we write:

$$y_1=x+v_1;$$

$$\tilde{y}_2\approx y_2-s(0)=\tilde{H}x+v_2,$$

where $\tilde{H}=\left.\dfrac{d\tilde{s}(x)}{dx^{\mathrm{T}}}\right|_{x=0}$ is an $l\times n$ matrix, and H is taken to be a $(l+n)\times n$ matrix, i.e.,

$$H=\left.\begin{bmatrix} E_{n\times n} \\ \dfrac{d\tilde{s}(x)}{dx^{\mathrm{T}}} \end{bmatrix}_{x=0}\right|.$$

This type of problem is encountered, for example, in aiding (correction) of a navigation system using some aiding measurements. In particular, distances to beacons (landmarks) can be used as aiding data. For example, assume that we have measurements $y_1=x+v_1$ from a navigation system and aiding distance measurements to the beacon ($l=1$), where $x=(x_1,x_2)^{\mathrm{T}}$ is a two-dimensional vehicle position vector, $y_1=(y_1^{(1)},y_1^{(2)})^{\mathrm{T}}$, $v_1=(v_1^{(1)},v_1^{(2)})^{\mathrm{T}}$. Then it is possible to write

$$y_1^{(1)}=x_1+v_1^{(1)};$$

$$y_1^{(2)}=x_2+v_1^{(2)};$$

$$y_2=\sqrt{\left(x_1^o-x_1\right)^2+\left(x_2^o-x_2\right)^2}+v_2,$$

where $x^o=(x_1^o,x_2^o)^{\mathrm{T}}$ are the known beacon coordinates.

After linearizing function $s(x)$ at the point $x^l=0$, the aiding measurement y_2 can be represented as

$$\tilde{y}_2(x^l) = y_2 - s(0) = \tilde{H}x + v_2 = -x_1 \sin B - x_2 \cos B + v_2 .$$

Thus, introducing $v = (v_1^{(1)}, v_1^{(2)}, v_2)^{\mathrm{T}}$, $H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -\sin B & -\cos B \end{bmatrix}$, all measurements $y = (y_1^{(1)}, y_1^{(2)}, \tilde{y}_2)^{\mathrm{T}}$ can be written as $y = Hx + v$. Here, B is the angle determining the orientation of the unit vector $(\sin B, \cos B)^T$ with respect to OX_2.

In the most general case, the problem of integrated processing in the linear statement can be formulated as follows.

Estimate the *n*-dimensional vector x using the set of m vector measurements

$$y_j = H_j x + v_j, \quad j = \overline{1.m},$$

where H_j are $m_j \times n$ matrices; v_j are the m_j-dimensional vectors of measurement errors; j is the measurement number.

Introducing $m_\Sigma \times n$ matrix H, where $m_\Sigma = \sum_{j=1}^{m} m^j$, composite vectors of measurements y and their errors v in the form

$$H = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_m \end{bmatrix}, \; y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \; v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix},$$

we note that in this general case the problem is reduced to a usual statement: estimate x from $y = Hx + v$.

The next sections discuss different variants of designing integrated processing algorithms.

5.2. Complementary filter. Invariant processing scheme

If data from two or more sensors or measuring devices are available, the **complementary filter** is often used. Such a filter is used in the conditions where a priori information on vector x is absent, but we have only a priori statistical information on the sensor errors. [4]. **The essence of the complementary filter** is to use differential measurements Δy that do not contain the sought vector x to obtain an estimate of the errors of one sensor v_1 against the background of the errors of the other sensor v_2 (Fig. 5.1).

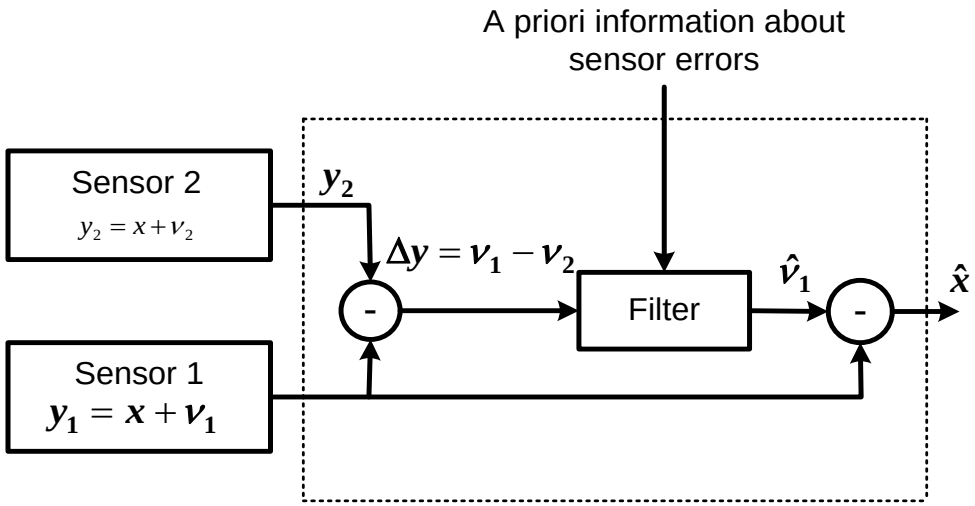


Fig. 5.1. Complementary filter or the invariant scheme of an integrated processing algorithm

Let us illustrate the use of this filter and analyze the properties of the obtained estimation errors by the example of the simplest integrated processing problem considered above. Assume that we have two measurements

$$y_1 = x + v_1;$$

$$y_2 = x + v_2,$$

and the sensor errors v_1, v_2 are zero-mean vectors noncorrelated with each other with the known covariance matrices $R_j > 0,\ j = 1,2$.

In order to design the complementary filter, let us generate the difference measurements

$$\Delta y = y_1 - y_2 = v_1 - v_2, \tag{5.1}$$

which exclude the sought vector x and use Δy to estimate the errors of one sensor against the background of the errors of the other sensor. To be definite, suppose that v_1 is the estimated vector, and v_2 is the measurement error vector.

The available data on mathematical expectations and covariance matrices of v_1, v_2 make it possible, using the results of section 4.5, to obtain for v_1, linear RMS optimal estimate and the corresponding covariance matrix, which can be easily shown to be defined as

$$\hat{v}_1 = \left(R_1^{-1} + R_2^{-1}\right)^{-1} R_2^{-1}(y_1 - y_2); \tag{5.2}$$

$$P^{v_1} = \left(R_1^{-1} + R_2^{-1}\right)^{-1}. \tag{5.3}$$

The estimate of the sought vector x can be calculated in the form

$$\hat{x} = y_1 - \hat{v}_1 = x + v_1 - \hat{v}_1 = \left(R_1^{-1} + R_2^{-1}\right)^{-1}\left[\left(R_1^{-1} + R_2^{-1}\right)y_1 - R_2^{-1}(y_1 - y_2)\right].$$

The formula for estimate $\hat{x}$ is easy to transform as:

$$\hat{x} = \left(R_1^{-1} + R_2^{-1} \right)^{-1} \left[R_1^{-1} y_1 + R_2^{-1} y_2) \right]. \quad (5.4)$$

Obviously, the estimate error $\varepsilon = x - \hat{x} = v_1 - \hat{v}_1$ does not depend on the vector x being estimated, and thus, it is invariant to x, which is why the processing scheme corresponding to the complementary filter is called **invariant**. The corresponding algorithms are also referred to as **invariant algorithms (filters)**. Sometimes, estimates are said to be invariant. The error covariance matrix for the obtained estimate (5.4) coincides with matrix (5.3) for estimate v_1.

Write the measurements y_1, y_2 as $y = Hx + v$, where $y = (y_1^{\mathrm{T}}, y_2^{\mathrm{T}})^{\mathrm{T}}, v = (v_1^{\mathrm{T}}, v_2^{\mathrm{T}})^{\mathrm{T}}, H = \begin{bmatrix} E_n \\ E_n \end{bmatrix}$ and $R^v = \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix}$. The GLSM described in chapter 2 can be used to solve this estimation problem. Assume that in criterion (2.13) $Q = R^{-1}$, it is easy to verify that estimate (5.4) is exactly the same as the estimate derived using the GLSM algorithm (Exercise 5.1).

It follows from the above that **in this case the complementary filter (invariant scheme) can be treated as a specially organized GLSM estimation procedure.** Its distinctive feature is that any assumptions about stochastic properties on the sought vector x need not be made in designing the estimation algorithm.

It should be noted that if we additionally assume that v_1 and v_2 are Gaussian, estimate (5.2) for error v_1 of the first sensor becomes the optimal Bayesian estimate, and estimate (5.4) coincides with the maximum likelihood estimate, i.e., the estimate derived using the invariant algorithm coincides with the maximum likelihood estimate. In these conditions, we can say that the complementary filter (invariant scheme) provides maximum likelihood estimates whose error also does not depend on the vector of the parameters being estimated.

The technique described above can be also applied to process measurements given in the form

$$y_1 = x + v_1; \quad (5.5)$$

$$y_2 = Hx + v_2. \quad (5.6)$$

To get the complementary filter (invariant scheme), difference measurement should be generated:

$$\Delta y = y_2 - Hy_1 = -Hv_1 + v_2.$$

Then, assuming errors v_1, v_2 to be noncorrelated zero-mean vectors with the known covariance matrices $R_j > 0$, $j = 1, 2$, within the Bayesian approach, the linear optimal estimate of vector v_1 and the corresponding covariance matrix of its estimation error can be easily calculated using measurements $\Delta y = y_2 - Hy_1$:

$$\hat{v}_1 = \left(R_1^{-1} + H^{\mathrm{T}} R_2^{-1} H \right)^{-1} H^{\mathrm{T}} R_2^{-1} (y_2 - Hy_1),$$

$$P^{v_1} = \left(R_1^{-1} + H^{\mathrm{T}} R_2^{-1} H\right)^{-1}.$$

Subtracting the derived error estimate $\hat{v}_1$ of the first sensor from its outputs y_1, we get the following for the estimate of x:

$$\hat{x} = y_1 - \left(R_1^{-1} + H^{\mathrm{T}} R_2^{-1} H\right)^{-1} H^{\mathrm{T}} R_2^{-1} (y_2 - H y_1). \qquad (5.7)$$

It can be easily proven (problem 5.2) that in this case as well, the estimate and its error covariance matrix coincide with the GLSM or maximum likelihood estimates if measurement errors are assumed Gaussian.

It is important to emphasize that in order to design an invariant scheme, it would be sufficient to have at least one measurement of type $y = x + v$ among the measurements used, which provides direct measurement of the sought vector x.

5.3. Non-invariant processing scheme

It should be noted that the invariant scheme described in 5.2 provides the Bayesian estimates, optimal in RMS sense, using difference measurements Δy only for the sensor errors and not for the sought parameters. To derive the optimal Bayesian estimate of the sought vector x, we should make an assumption of its random nature. Besides, it is advisable to consider the integrated measurement processing problem as the traditional estimation problem for vector *x* by measurements $y=Hx+v$ and use the Bayesian algorithms described in sections 4.4. or 4.5 for its solution. Since the error of optimal Bayesian estimate $x - \hat{x}(y) = (E - KH)(x - \bar{x}) + Kv$ depends not only on measurement errors but also on the sought parameter, the algorithm used to find this estimate is sometimes referred to as a **non-invariant algorithm**. The algorithm scheme is shown in Fig. 5.2. Here, unlike the previous case, both measurements are equally processed in the algorithm using a priori data both on the measurement errors and the vector being estimated.

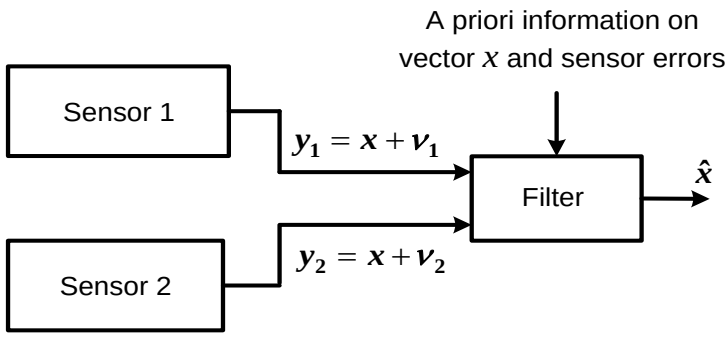


Fig. 5.2. Non-invariant (Bayesian) scheme of the integrated processing algorithm

The principal difference between the invariant and non-invariant algorithms is that in the first case, a priori information about stochastic properties only on measurement errors is used, whereas in the second case, such information on the sought parameters

vector is additionally applied.

The advantage of the invariant scheme is that no assumptions about stochastic properties should be made on the vector to be estimated. In some cases, it is justified since providing adequate description for x is often complicated. However, it should be remembered that if information on the sought parameters of the vector is available, its non-use can significantly degrade the accuracy.

Consider an example illustrating a possible loss in accuracy in the case that a priori information about the properties of the estimated vector is not taken into account.

♦ **Example 5.1.** Suppose that an aircraft height is measured using a satellite navigation system (SNS) receiver and barometric altimeter data at times t_i ($i = \overline{1.m}$). Write these measurements as

$$y_i^{SNS} = h_i + v_i^{SNS},\ i = \overline{1.m}; \tag{5.8}$$

$$y_i^{BA} = h_i + v_i^{BA},\ i = \overline{1.m}. \tag{5.9}$$

Introduce vectors $x = (h_{1,}....h_m)^{\mathrm{T}}$, $v^{SNS} = (v_1^{SNS} v_2^{SNS},...v_m^{SNS})^{\mathrm{T}}$, $v^{BA} = (v_1^{BA} v_2^{BA},...v_m^{BA})^{\mathrm{T}}$ and matrix $H = E_m$, then measurements (5.8), (5.9) can be represented as $y = Hx + v$.

Assume for the beginning that a priori information on vector x is absent, and we have only a priori statistical information on the sensor errors; they are zero-mean Gaussian noncorrelated errors, and for each of them, the covariance matrices are set as $R_j = r_j^2 E_m$, $j = SNS, BA$. Let us solve the problem under these assumptions using a complementary algorithm and difference measurements

$$\Delta y_i = y_i^{SNS} - y_i^{BA} = v_i^{SNS} - v_i^{BA}.$$

Using (4.13) it is easily to derive the formula for the estimate (exercise 5.3) and show that the covariance matrix for vector x is diagonal, and with $r_j^2 = r$, $j = SNS, BA$, the equality $P = \dfrac{r^2}{2} E_m$ is true for it. Hence, it follows that the height error variance at each discrete moment is only decreased twofold.

Now, assume that we have also a priori stochastic information on vector x and take it into account in estimating x. In particular, suppose that the height is constant over the observation interval and it is a zero-mean Gaussian random variable with variance σ_0^2. In this case, the a priori covariance for vector x is $P^x = \sigma_0^2 I_{m\times m}$, where $I_{m\times m}$ is the matrix of ones. Using the materials of section 4.4, we can derive the Bayesian estimate and the corresponding covariance matrix (exercise 5.4). It should be noted that since $P^x = \sigma_0^2 I_{m\times m}$ is singular, (4.16) cannot be used to calculate the covariance matrix, so we use (4.15). With account for the specific matrix $H^{\mathrm{T}} = [E_m, E_m]$, we obtain

$$P = \sigma_0^2 I_{m\times m} - \sigma_0^4 I_{m\times 2m}\left(\sigma_0^2 I_{2m\times 2m} + r^2 E_{2m}\right)^{-1} I_{2m\times m}.$$

Using the relationship following from (A1.59), it is possible to write:

$$\left(\sigma_0^2 I_{2m\times 2m} + r^2 E_{2m}\right)^{-1} = \frac{1}{r^2}\left(E_{2m} - \frac{\sigma_0^2}{2m\sigma_0^2 + r^2} I_{2m\times 2m}\right).$$

It can be shown that the diagonal elements of the covariance matrix for the vector being estimated, which determine the height estimation accuracy, are given by

$$\sigma_h^2 = \frac{\sigma_0^2 r^2}{2m\sigma_0^2 + r^2}. \qquad (5.10)$$

The meaning of this formula is quite clear as with the assumptions made, we actually solve the problem of estimating the scalar value from $2m$ measurements with independent errors with variances r^2.

Neglecting the contribution of a priori information about random vector *x* and taking $r^2 << \sigma_0^2$, we derived that height error variance decreases by $2m$ times, i.e., $\sigma_h^2 \approx \frac{r^2}{2m}$, which is significantly greater than in the previous case. ♦

5.4. Centralized and decentralized data processing schemes

Consider the problem of integrated processing of measurements formulated above:

$$y_j = H_j x + v_j, \quad j = \overline{1.i}, \qquad (5.11)$$

where H_j are $m_j \times n$ matrices; v_j are m_j-dimensional error vectors; j is the sensor number.

Introduce composite vectors of measurements Y_i and their errors V_i with dimensions $m_{\Sigma} = m_{\Sigma} = \sum_{j=1}^{i} m_j$, and $m_{\Sigma}\times n$ matrix $\mathbf{H}_i$:

$$\mathbf{H}_i = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_i \end{bmatrix}, \; Y_i = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix} V_i = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_i \end{bmatrix}.$$

Then measurements (5.11) can be written as

$$Y_i = \mathbf{H}_i x + V_i.$$

Assume x and v_j to be random zero-mean vectors with preset covariance matrices P^x, R_j, $j = \overline{1.i}$. For simplicity, assume that all these vectors are noncorrelated with each other. Then, covariance matrix $\mathbf{R}_i$ for the composite vector V_i is given by

$$\mathbf{R}_i = \begin{bmatrix} R_1 & 0 & 0 & 0 \\ 0 & R_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & R_i \end{bmatrix}.$$

With the assumptions made, we can use the optimal Bayesian algorithms for the solution of this problem and easily derive the following formulas for the optimal, in the RMS sense, estimate and its covariance matrix:

$$\hat{x}_i(Y_i) = P_i \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} Y_i \tag{5.12}$$

$$P_i = \left((P^x)^{-1} + \sum_{j=1}^{i} H_j^{\mathrm{T}} R_j^{-1} H_j \right)^{-1}. \tag{5.13}$$

Hence,

$$\hat{x}_i(Y_i) = \left((P^x)^{-1} + \sum_{j=1}^{i} H_j^{\mathrm{T}} R_j^{-1} H_j \right)^{-1} \left(\sum_{j=1}^{i} H_j^{\mathrm{T}} R_j^{-1} y_j \right). \tag{5.14}$$

Note that (5.12) is a linear optimal estimate in the RMS sense. This means that the estimate minimizes the RMS criterion in a class of algorithms linearly dependent on measurements. This allows us, in designing algorithms, to use information only about the mathematical expectations and the covariance matrices of the corresponding random vectors.

To get the optimal estimate (5.12), two processing schemes can be applied to the whole set of measurements. The first one is called the **centralized processing scheme** (Fig. 5.3). The name is explained by the fact that all computations are performed centrally within one algorithm.

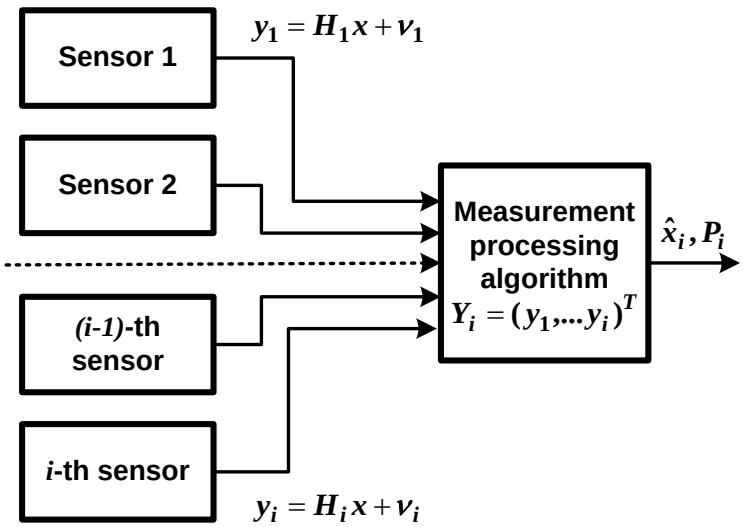


Fig. 5.3. Centralized optimal estimation scheme

Another scheme can be used to calculate the estimate.

Suppose that inequality

$$(P^x)^{-1} << H_j^{\mathrm{T}} R_j^{-1} H_j, \ j = \overline{1.i} \tag{5.15}$$

is true, meaning that a priori information on vector x can be neglected. Introduce partial optimal estimates calculated using only partial j-th measurements,

$$\hat{x}^{(j)} \cong \left(P^{(j)}\right) H_j^{\mathrm{T}} R_j^{-1} y_j, \tag{5.16}$$

where

$$P^{(j)} = (H_j^{\mathrm{T}} R_j^{-1} H_j)^{-1}, \tag{5.17}$$

Taking into account (5.14) (5.16), it can be easily found that the formula for the linear optimal estimate calculated by using the whole measurement set can be reduced to

$$\hat{x}_i \cong \left(\sum_{j=1}^{i} \left(P^{(j)}\right)^{-1} \right)^{-1} \left(\sum_{j=1}^{i} \left(P^{(j)}\right)^{-1} \hat{x}^{(j)} \right). \tag{5.18}$$

In other words, the optimal error in using the whole measurement set is a weighted sum of partial optimal estimates.

Assumption (5.15) need not be made if the partial estimates are calculated using

$$P^{(j)} = \left(\frac{1}{i} (P^x)^{-1} + H_j^{\mathrm{T}} R_j^{-1} H_j \right)^{-1} \tag{5.19}$$

instead of (5.17).

The described scheme is presented in Fig. 5.4. It is called the **decentralized scheme** because partial estimates and their covariance matrices can be calculated in separate algorithms, and the sought estimate is found by weighting these partial estimates.

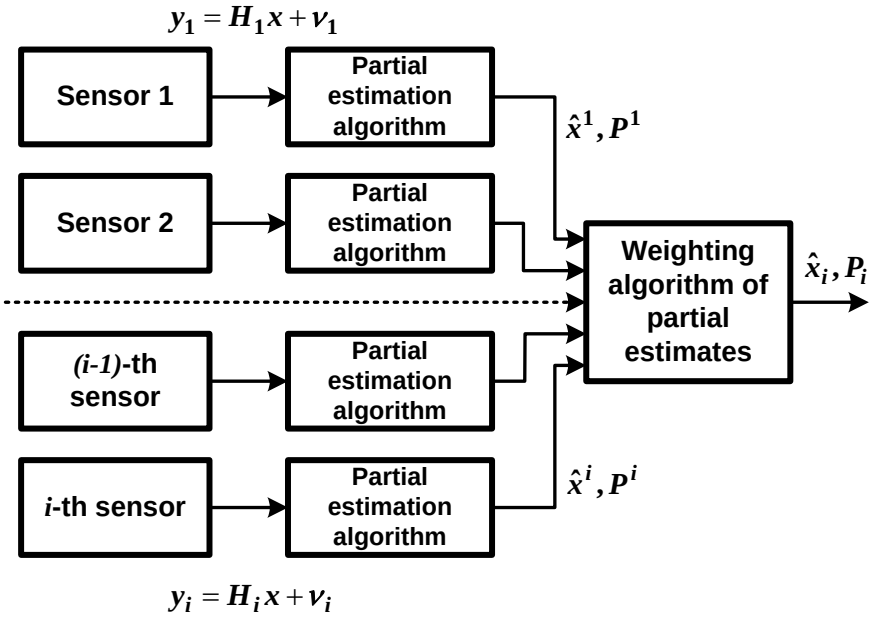


Fig. 5.4. Decentralized processing scheme using the results of partial algorithms

Note that if we additionally assume that x and v_j are Gaussian vectors, then the above estimates are optimal Bayesian estimates minimizing the RMS criterion without any restrictions on the class of the algorithms used.

5.5. Recursion processing scheme

The recursion scheme is the most popular one in navigation information processing. The idea is that the sought estimate is not obtained by processing the whole measurement set; it is generated by successive processing of each available measurement and the results derived at the previous processing step. These algorithms are referred to as **recursive algorithms**.

The idea of the algorithm design can be explained by a simple example of estimating an unknown scalar from scalar measurements

$$y_i = x + v_i .$$

If the measurement errors are assumed to be noncorrelated random variables with the same variances, the GLSM estimate can be determined as an arithmetic mean of the accumulated measurements, i.e., $\hat{x}_i = \frac{1}{i}\sum_{j=1}^{i} y_j$. Writing the chain of equalities

$$\hat{x}_i = \frac{1}{i}\sum_{j=1}^{i} y_j = \frac{y_i + \sum_{j=1}^{i-1} y_j}{i} = \frac{y_i}{i} + \frac{i-1}{i}\hat{x}_{i-1} = \hat{x}_{i-1} + \frac{1}{i}(y_i - \hat{x}_{i-1}),$$

we obtain

$$\hat{x}_i = \hat{x}_{i-1} + \frac{1}{i}(y_i - \hat{x}_{i-1}). \tag{5.20}$$

Similarly, obtain the recursive algorithm for the problem (5.11) considered in the previous section, taking that the measurements are successively input to the processing scheme. Introduce the estimate and its error covariance matrix for the processing of the $(i-1)$-th measurement:

$$\hat{x}_{i-1}(Y_{i-1}) = \left((P^x)^{-1} + \sum_{j=1}^{i-1} H_j^{\mathrm{T}} R_j^{-1} H_j \right)^{-1} \left(\sum_{j=1}^{i-1} H_j^{\mathrm{T}} R_j^{-1} y_j \right), \tag{5.21}$$

$$P_{i-1} = \left((P^x)^{-1} + \sum_{j=1}^{i-1} H_j^{\mathrm{T}} R_j^{-1} H_j \right)^{-1}. \tag{5.22}$$

Then, the following chain of equalities can be written:

$$\hat{x}_i(Y_i) = P_i \left(\sum_{j=1}^{i} H_j^{\mathrm{T}} R_j^{-1} y_j \right) = P_i \left(H_i^{\mathrm{T}} R_i^{-1} y_i + \sum_{j=1}^{i-1} H_j^{\mathrm{T}} R_j^{-1} y_j \right) =$$

$$= P_i\left(H_i^{\mathrm{T}} R_i^{-1} y_i + P_{i-1}^{-1} P_{i-1}\left(\sum_{j=1}^{i-1} H_j^{\mathrm{T}} R_j^{-1} y_j \right)\right) = P_i\left(H_i^{\mathrm{T}} R_i^{-1} y_i + P_{i-1}^{-1}\hat{x}_{i-1}(Y_{i-1})\right) =$$

$$= P_i\left(H_i^{\mathrm{T}} R_i^{-1} y_i + P_{i-1}^{-1}\hat{x}_{i-1}(Y_{i-1})\right) + P_i H_i^{\mathrm{T}} R_i^{-1}\left(H_i\hat{x}_{i-1}(Y_{i-1}) - H_i\hat{x}_{i-1}(Y_{i-1})\right) =$$

$$= P_i H_i^{\mathrm{T}} R_i^{-1}(y_i - H_i\hat{x}_{i-1}(Y_{i-1})) + P_i\left(P_{i-1}^{-1} + H_i^{\mathrm{T}} R_i^{-1} H_i\right)\hat{x}_{i-1}(Y_{i-1}) =$$

$$= \hat{x}_{i-1}(Y_{i-1}) + P_i H_i^{\mathrm{T}} R_i^{-1}(y_i - H_i\hat{x}_{i-1}(Y_{i-1}));$$

$$P_i = \left((P^x)^{-1} + \sum_{j=1}^{i} H_j^{\mathrm{T}} R_j^{-1} H_j \right)^{-1} =$$

$$= \left((P^x)^{-1} + \sum_{j=1}^{i-1} H_j^{\mathrm{T}} R_j^{-1} H_j + H_i^T R_i^{-1} H_i \right)^{-1} = \left(P_{i-1}^{-1} + H_i^{\mathrm{T}} R_i^{-1} H_i\right)^{-1},$$

from which, obviously, the sought recursive relationships follow:

$$\hat{x}_i(Y_i) = \hat{x}_{i-1}(Y_{i-1}) + P_i H_i^{\mathrm{T}} R_i^{-1}(y_i - H_i\hat{x}_{i-1}(Y_{i-1})); \tag{5.23}$$

$$P_i = \left(P_{i-1}^{-1} + H_i^{\mathrm{T}} R_i^{-1} H_i\right)^{-1}. \tag{5.24}$$

It should be emphasized that these relationships have recursive nature both for the estimate and for the covariance matrix. With a matrix

$$K_i = P_i H_i^{\mathrm{T}} R_i^{-1} \tag{5.25}$$

introduced, the estimate can be given by

$$\hat{x}_i = \hat{x}_{i-1}(y_{i-1}) + K_i(y_i - H_i\hat{x}_{i-1}(y_{i-1})). \tag{5.26}$$

With account for (4.24), (4.26), error covariance matrix P_i and matrix K_i can be calculated as

$$P_i = P_{i-1} - P_{i-1}H_i^{\mathrm{T}}(H_i P_{i-1} H_i^{\mathrm{T}} + R_i)^{-1} H_i P_{i-1} = (E - K_i H_i)P_{i-1}, \tag{5.27}$$

$$K_i = P_{i-1}H_i^{\mathrm{T}}(H_i P_{i-1} H_i^{\mathrm{T}} + R_i)^{-1}. \tag{5.28}$$

As will be shown in further sections, actually, these are Kalman filter formulas for the problem of constant vector estimation studied here.

Exercises

Exercise 5.1. Solve the vector estimation problem having measurements

$$y_1 = x + v_1;$$

$$y_2 = x + v_2,$$

using the GLSM, assuming that in criterion (2.13), $Q = \begin{bmatrix} R_1^{-1} & 0 \\ 0 & R_2^{-2} \end{bmatrix}$.

Assuming that the measurement errors are noncorrelated with each other centered

zero-mean vectors with covariance matrices R_1 and R_2, using (2.29), write the covariance matrix for the errors of the obtained estimate.

Compare the derived equations with the results corresponding to the complementary filter for the problem under consideration.

Exercise 5.2. Let there be an *n*-dimensional unknown vector *x* to be estimated from measurements (5.5), (5.6) written as

$$y_1 = x + v_1;$$
$$y_2 = Hx + v_2,$$

where y_2 is an *l*-dimensional vector; H is an $l \times n$ matrix. Use the GLSM to derive the *x* estimate with $Q = R^{-1}$.

Assuming additionally that v_1 and v_2 are noncorrelated zero-mean vectors with known covariance matrices $R_j > 0,\ j = 1,2$, and R is the covariance matrix of vector $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, derive the formula for the estimation error covariance matrix using (2.29). Make sure that the obtained estimates coincide with the estimates derived using a complementary filter (invariant scheme). Under which conditions will this estimate agree with the maximum likelihood estimate?

Exercise 5.3. For Example 5.1 obtain the formula for the estimate of vector *x* and the corresponding covariance matrix when the measurements (5.8.), (5.9) and a complementary filter are used.

Exercise 5.4. For Example 5.1 obtain the formula for the estimate of vector *x* and the corresponding covariance matrix when measurements (5.8.), (5.9) and the optimal Bayesian algorithm are used.

Exercise 5.5. The aircraft height measurements are taken at times $t_i = \Delta t(i-1)$, $i = \overline{1.m}$ with equal intervals Δt using a satellite receiver and a barometric altimeter. The measurement errors are independent random values with variances r_{SNS}^2 и r_{BA}^2. The vehicle height is described as a first-order polynomial $h_i = x_0 + Vt_i$.

Formulate the problem of integrated processing of these measurements to derive optimal estimates of vector $x = (x_1, x_2)^{\mathrm{T}} = (x_0, V)^{\mathrm{T}}$ assuming that its components are Gaussian random values with mathematical expectation $(\overline{x}_0, 0)^{\mathrm{T}}$ and variances σ_0^2, σ_V^2, independent of each other and of measurement errors. Write the formula for the error covariance matrix for optimal estimates.

Exercise 5.6. Consider the problem of aiding the navigation system using distance measurements to the beacon, discussed in this section. Suppose that it can be solved in a linear statement; thus, the measurements can be written as

$$y_1^{(1)} = x_1 + v_1^{(1)};$$
$$y_1^{(2)} = x_2 + v_1^{(2)};$$

$$y_2 = \tilde{H}x + v_2 = -x_1 \sin B - x_2 \cos B + v_2,$$

where $x = (x_1, x_2)^{\mathrm{T}}$ is a two-dimensional vector setting the vehicle 2D position, and angle B sets the orientation of the unit vector $(\sin B, \cos B)^{\mathrm{T}}$ with respect to axis Ox_2.

Let $v_1 = (v_1^{(1)}, v_1^{(2)})^{\mathrm{T}}$ be a zero-mean measurement error vector with covariance matrix R that has the parameters of error ellipse a, b, τ, and error v_2 is a zero-mean random value noncorrelated with this vector with a variance coinciding with b^2.

Specify the complementary filter for the position estimation algorithm and make sure that this algorithm agrees with the GLSM with $Q = R^{-1}$.

Find angle B such that the distance RMS (DRMS) error is minimum.

Exercise 5.7. Assume that we have two sensors with outputs

$$y_1 = x + v_1;$$

$$y_2 = x + v_2,$$

where $x = (x_1, x_2)^{\mathrm{T}}$ is a two-dimensional vector setting the vehicle 2D position. Two-dimensional vectors are assumed to be noncorrelated zero-mean random vectors with covariance matrices $R_2 > 0$.

Obtain the formula for the estimate error covariance matrix corresponding to the GLSM assuming that Q is a block-diagonal matrix with blocks R_1^{-1} and R_2^{-1}.

Assuming that dimensions of minor and major semiaxes of these ellipse matrices are the same, i.e., $a_1 = a_2 = a$, and $b_1 = b_2 = b$, determine which mutual orientation of these ellipses provides minimum (maximum) DRMS error.

Test questions

1. Formulate the problem of integrated measurement processing and illustrate it by examples. What is a complementary filter?
2. Explain the idea of the invariant algorithm by the example of processing data from two sensors. Explain the name of the algorithm. Provide the measurement processing scheme.
3. How is the algorithm of the complementary filter related to the GLSM and the maximum likelihood algorithms?
4. Outline the specific feature of the non-invariant algorithm. Provide the measurement processing scheme. Discuss its advantages and disadvantages as compared with the complementary filter.
5. Explain the ideas of the centralized and decentralized schemes for processing data from several sensors.
6. Explain how recursive estimation algorithms can be derived in estimating the vector of constant parameters.

6. Random sequences

This chapter provides basic definitions and properties of discrete random processes, as well as methods used to describe them. These processes are often referred to as sequences or discrete time series. Hereinafter, we will use the term “sequence”.

6.1. Definition of a random sequence and its description

Consider a sequence of scalar values $x_1, x_2, x_3, ...x_i, ...$ A **sequence is called random** if each x_i is a random variable. Figure 6.1 shows a sample of a random sequence. If the subscript corresponds to time t_i, we deal with a temporal random sequence. Further, we consider a temporal random sequence and the subscript is treated as discrete time.

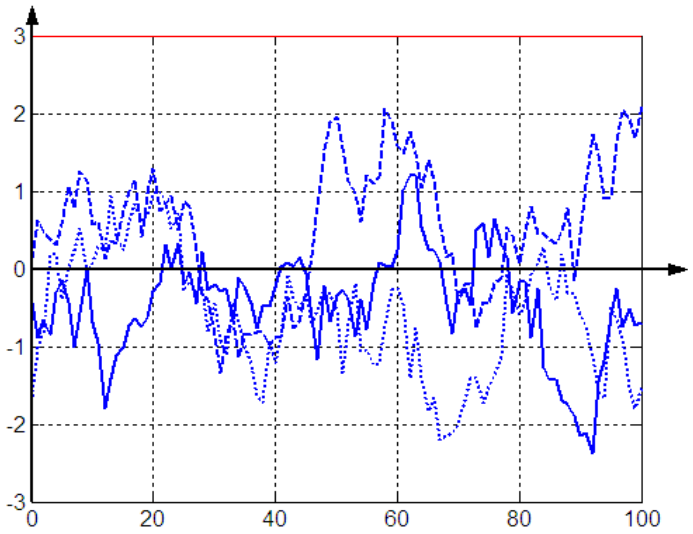


Fig. 6.1. A set of samples of random sequences

Assume that x_i, $i = 1, 2..$ is a random sequence. If we fix two arbitrary time points, for example, t_i and t_j, we can form a two-dimensional random vector whose components are two random variables x_i and x_j. Joint statistical properties of this vector can be determined with the use of either a cumulative probability density (CDF) or a probability density function (PDF). It is also possible to introduce similar functions for a vector comprising a greater number of variables of the sequence. It is clear that the description of a random sequence will be fully specified if we can determine PDF $p(x_1, x_2.....x_i)$ for any finite set of variables in a sequence at arbitrary time points $t_1, t_2.....t_i$.

We can also introduce some characteristics of a random sequence, the most

important of which are the **mathematical expectation and the variance of the random sequence.** Knowing PDF $p(x_i)$ for an arbitrary i-th time point, these characteristics can be defined as:

$$\overline{x}_i = E\{x_i\} = \int x_i p(x_i) dx_i ; \quad (6.1)$$

$$\sigma_i^2 = E\left\{\left(x_i - \overline{x}_i\right)^2\right\} = \int \left(x_i - \overline{x}_i\right)^2 p(x_i) dx_i . \quad (6.2)$$

If joint PDF $p(x_i, x_j)$ is assumed known for the sequence values at arbitrary time points, it is possible to introduce another very important characteristic:

$$k(i,j) = E\left\{(x_i - \overline{x}_i)(x_j - \overline{x}_j)\right\} = \int\int (x_i - \overline{x}_i)(x_j - \overline{x}_j) p(x_i, x_j) dx_i dx_j , \quad (6.3)$$

which is called the correlation function of a random sequence.

A similar characteristic of two different sequences is called a **cross-correlation function**. Function (6.3) sets the value of the correlation coefficient between random variables, corresponding to the *i*-th and j-th time points. By virtue of (4.1), the following equation

$$p(x_i, x_j) = p(x_i / x_j) p(x_j)$$

is true. Here, $p(x_i / x_j)$ is a conditional PDF for the variable of the sequence at the i-th time under the condition that the value is fixed at the j-th time.

Since at $i = j$, $p(x_i / x_j) = \delta(x_i - x_j)$, i.e., the transition density is a delta-function, it is clear that

$$k(i,i) = \sigma_i^2 , \quad (6.4)$$

i.e., at a fixed time, the sequence variance coincides with the value of the correlation function at coincident values of the argument $i = j$.

Random sequences may be not only scalar, but vector as well. The term **a random *n*-dimensional vector sequence** is used in reference to a sequence whose variables are *n*-dimensional random vectors.

Assume that x_i, $i = 1, 2..$ is an *n*-dimensional random sequence. Note that hereinafter, we use the subscript to indicate the discrete time point which corresponds to the vector and its components.

If we need to simultaneously specify the time point and the number of the component, we use double subscripting, i.e., $x_i = (x_{i1}, ..., x_{in})^{\mathrm{T}}$. The first of the subscripts indicates the time, while the second, the number of the component.

All of the definitions given above can be generalized for the case of the vector sequence. In particular, the correlation function will be **a correlation matrix** defined as

$$k(i,j) = \int\int (x_i - \overline{x}_i)(x_j - \overline{x}_j)^{\mathrm{T}} p(x_i, x_j) dx_i dx_j . \quad (6.5)$$

For the same values of the argument $i = j$, the correlation matrix coincides with the covariance matrix, i.e.,

$$k(i,i) = P_i = \int (x_i - \overline{x}_i)(x_i - \overline{x}_i)^{\mathrm{T}} p(x_i) dx_i . \quad (6.6)$$

6.2. Stationary (time-invariant) random sequences. Discrete white noise

Consider some of the most common types of random sequences.

An important class of random sequences is **stationary random sequences**.

A stationary (time-invariant) sequence, **in a broad sense,** is a sequence whose expectation does not depend on time, whereas the correlation function depends on the difference $(i-j)$:

$$\overline{x}_i = \overline{x};$$

$$k(i-j) = E\{(x_i - \overline{x}_i)(x_j - \overline{x}_j)\}.$$

From this formula it follows that the variance of the stationary sequence does not depend on time since

$$\sigma^2 = k(0).$$

If the above-mentioned properties are not met, the sequence is called a **nonstationary** one (time-varying).

A strictly stationary (time-invariant) sequence is a sequence for which the PDF $p(x_1,,.....x_k)$, introduced for any finite set of sequence values, remains unchanged when all time points change simultaneously by a value μ, i.e., $p(x_1,,.....x_k) = p(x_{1+\mu},,.....x_{k+\mu})$.

The sequence that has a zero mathematical expectation $(\overline{x}=0)$ is called **a zero-mean sequence**. The samples shown in Fig. 6.1 correspond to a zero-mean sequence. A sample of a nonzero-mean random sequence is shown in Fig. 6.2.

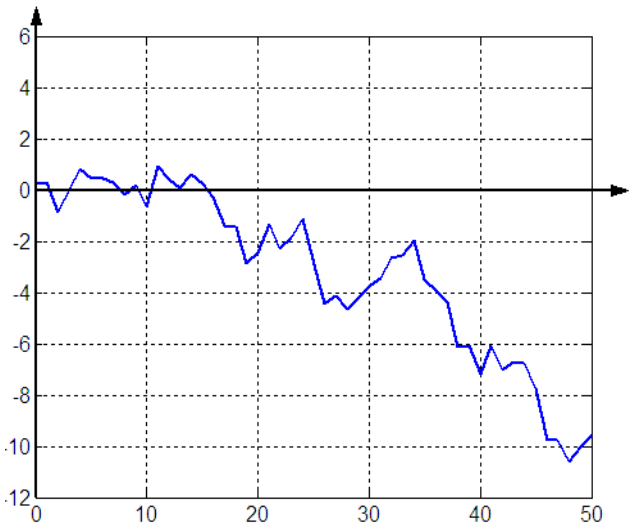


Fig. 6.2. Realization of a nonzero-mean random sequence

It is clear that a random sequence with a nonconstant expectation is nonstationary. An example of a **nonzero-mean** sequence is shown in
Fig. 6.2.

Consider a sequence the variables of which are independent of each other at different time points:

$$p(x_1, x_2, ...x_k) = \prod_{j=1}^{k} p(x_j). \tag{6.7}$$

It is easy to understand that the correlation function of this sequence has the following form:

$$k(i, j) = \sigma_i^2 \delta_{ij}, \tag{6.8}$$

where δ_{ij} is the Kronecker symbol.

A sequence with a correlation function of the form (6.8) is called **discrete white noise**. In other words, discrete white noise is a sequence the variables of which are uncorrelated with each other at different time points. If the white noise is zero-mean and the variances for all times are identical, the noise is stationary. An example of a sample of such noise is shown in Fig. 6.3.

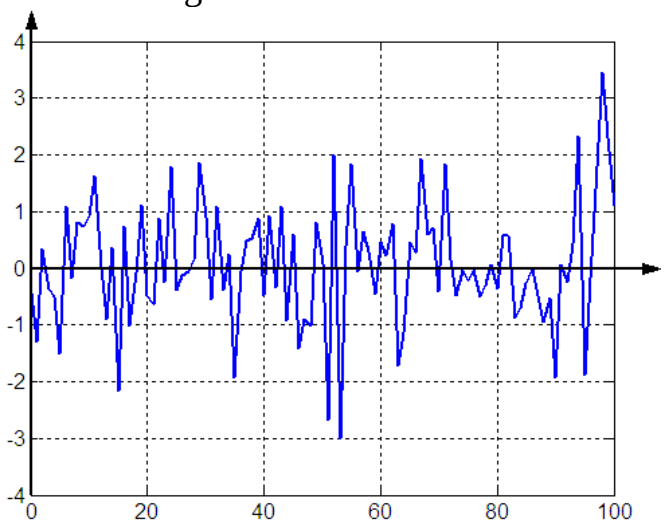


Fig. 6.3. Realization of stationary white noise

Note that the correlation function characterizes the temporal properties of the sequence, defining the level of statistical dependence of its values at different time points quantitatively. At the same time, the forms of PDF $p(x_1)$, $p(x_1, x_2)$, $p(x_1, x_2, x_3)$, etc. are also of vital importance. For example, we can introduce Gaussian sequences.

A random sequence is called a **Gaussian sequence** if PDF $p(x_1, ...x_k)$ for any set of its values is Gaussian at arbitrary time points. In particular, if this requirement is fulfilled for white noise, such noise is called **discrete Gaussian white noise**. It is important to emphasize that for Gaussian white noise, formula (6.7) is valid for any set of values. Thus, besides (6.8), for any $i \neq j$, we can be write

$$p(x_i, x_j) = p(x_i) p(x_j).$$

Gaussian sequences have a very important feature consisting in the fact that if we

know expectation (6.1) and correlation function (6.3), it is possible to determine PDF $p(x_1,.....x_k)$ for any set of variables of the sequence at arbitrary time points. In other words, the knowledge of (6.1) and (6.3) provides a complete description of its statistical properties. This is a consequence of the fact that the Gaussian PDF is completely determined by the first two time points.

♦ **Example 6.1.** Assume that expectation $\overline{x}_i$ and correlation function $k(i)$ are known for a scalar stationary Gaussian sequence. We need to write the PDF $p(x_1, x_2, x_3)$ for the vector composed of the values of sequence $x = (x_1, x_2, x_3)^{\mathrm{T}}$.

Since vector $x = (x_1, x_2, x_3)^{\mathrm{T}}$ is Gaussian, then, to find the corresponding PDF, it will suffice to determine the mathematical expectation $\overline{x} = (\overline{x}_1, \overline{x}_2, \overline{x}_3)^{\mathrm{T}}$ and the covariance matrix

$$P^x = \begin{bmatrix} k(0) & k(t_1 - t_2) & k(t_1 - t_3) \\ k(t_2 - t_1) & k(0) & k(t_2 - t_3) \\ k(t_3 - t_1) & k(t_2 - t_3) & k(0) \end{bmatrix}.$$

Hence, it follows that

$$p(x_1, x_2, x_3) = N(x_1, x_2, x_3; \overline{x}_1, \overline{x}_2, \overline{x}_3, P^x). \quad ♦$$

6.3. Markov sequences

Consider scalar sequence x_i. Fix time points in ascending order and form vector $x_1, x_2.....x_k$.. For this vector, we can write a joint PDF $p(x_1,.....x_k)$, for which (see Exercise 6.1), as follows from the formula for multiplication of PDF (4.1), the following formula holds true:

$$p(x_1,.....x_k) = p(x_k / x_{k-1}, x_{k-2},..x_1)\, p(x_{k-1} / x_{k-2}, x_{k-3},..x_1)...p(x_1). \quad (6.9)$$

The question of special interest in studying the temporal properties of sequences is how statistical properties of their values at current time x_k depend on the values at previous time points $x_1, x_2.....x_{k-1}$. This dependence is determined by the PDF of the form $p(x_k / x_1,..., x_{k-1})$. Markov sequences are an important class of random sequences. Their distinctive feature is that the statistical properties of their values at a current time point only depend on the values at the nearest previous time point. In view of the above-said, a Markov sequence can be defined as follows.

A Markov sequence is a random sequence for which the properties of its variables x_k at time point t_k with the known value at the nearest moment t_{k-1} depend only on the sequence values at this moment and do not depend on the sequence values at time points $t < t_{k-1}$, i.e.,

$$p(x_k / x_{k-1}, x_{k-2}....x_1) = p(x_k / x_{k-1}). \quad (6.10)$$

The conditional PDF of the Markov sequence $p(x_k / x_{k-1})$ is called **transition PDF** or **transition density**.

For the Markov sequence, Equation (6.9) can be transformed as follows:

$$p(x_k,...x_1) = p(x_1)\prod_{j=2}^{k} p(x_j / x_{j-1}).$$

Thus, specification of the transition PDF and PDF for the initial time $p(x_1)$ is sufficient to find the joint PDF for a set of Markov sequence values at arbitrary time points. The above types of random sequences can be easily generalized to the vector case.

6.4. Shaping filter

In applied problems, an important role is played by sequences specified with recursive difference equations of the form

$$x_i = \Phi_i x_{i-1} + \Gamma_i w_i, i = 1,2..., \tag{6.11}$$

where x_i is an n-dimensional vector; w_i is uncorrelated with x_0, zero-mean p-dimensional discrete white noise with the correlation function of the form

$$M\{w_i w_j^{\tau}\} = \delta_{ij} Q_i, \tag{6.12}$$

where Q_i is a $p \times p$ covariance matrix; Φ_i, Γ_i are the known $n \times n$ и $n \times p$- matrices.

Vector x_0, which specifies the sequence value at the initial time, is also assumed to be random with the known expectation $\bar{x}_0$ and covariance matrix P_0.

It is easy to see that for this sequence, formula (6.10) holds good, i.e., the sequence defined by (6.11) is a Markov sequence. If the sequence value at the initial time x_0 and white noise w_i are Gaussian

$$p(x_0) = N(x_0; \bar{x}_0, P_0); \tag{6.13}$$

$$p(w_i) = N(w_i; 0, Q_i), \tag{6.14}$$

then sequence x_i is **a Gaussian Markov sequence.**

This is a consequence of the fact that linear transformations of Gaussian vectors generate the Gaussian vector, and vectors x_0 and $w_1,....w_i$ are jointly Gaussian (see Exercise 6.2).

Equation (6.11) is called **a shaping filter of a random sequence**. This name is motivated by the fact that a random sequence is generated as a result of transformation of discrete white noise with the use of difference equation (6.11). The input white noise w_i is called **generating (forcing) noise or system noise**. Matrix Φ_i is called a **matrix of dynamics**, matrix Γ_i is a **matrix of generating noise**, and vector x_i is the **state vector**.

Taking into account the results of Exercise 6.3, we can see that the time evolution

of the expectation and the covariance matrix for sequence (6.11) is determined by the recursive relations:

$$\bar{x}_i = \Phi_i \bar{x}_{i-1}; \tag{6.15}$$

$$P_i = M\left\{(x_i - \bar{x}_i)(x_i - \bar{x}_i)^{\mathrm{T}}\right\} = \Phi_i P_{i-1} \Phi_i^{\mathrm{T}} + \Gamma_i Q_i \Gamma_i^{\mathrm{T}}, \tag{6.16}$$

allowing the values at the current step to be calculated using the corresponding values of the previous step.

From (6.15) it is obvious that the sequence formed with the use of (6.11) will be zero-mean at zero-mean generating noise, when the vector of initial conditions is zero-mean.

If necessary, it is easy, using the transformation rules of random vectors and (6.11), (6.12), to obtain covariance matrix $P^X{}_i$ for the composite vector $X_i = \left(x_0^{\mathrm{T}}, x_1^{\mathrm{T}}, ..., x_i^{\mathrm{T}}\right)^{\mathrm{T}}$ with the sequence values at all previous time points, including the initial variable (see Exercise 6.6).

The advantage of describing the sequence using the shaping filter (6.11) and formulas (6.15), (6.16) is the fact that this makes it possible to recursively solve the problem of the sequence formation itself, as well as the problem of calculating the corresponding statistical characteristics in the form of mathematical expectations and covariance matrices. In particular, the diagonal elements of the covariance matrix determine variances $\sigma_i^2(j) = P_i[j, j]$, $j = \overline{1.n}$ for each component of the state vector. In view of the above-said, shaping filters (6.11) are widely used in solving applied problems for simulation of random sequences. For this purpose, it is necessary to form random vector x_0, vectors w_i, using a random number generator, and then, using recursive formula (6.11), obtain the required values of samples formed in the general case with the use of the equation

$$z_i = H_i x_i. \tag{6.17}$$

Further, when considering the examples of how to obtain samples with the use of shaping filters, it is assumed that the sensors of Gaussian random variables are used.

♦ **Example 6.2.** A particular case of sequence (6.11) is a sequence in which there is no generating noise, i.e., the sequence of the form

$$x_i = \Phi_i x_{i-1}. \tag{6.18}$$

To determine its expectation, it is necessary to use (6.15). The covariance matrix can be calculated if we specify the initial covariance matrix P_0 and equation $P_i = \Phi_i P_{i-1} \Phi_i^{\mathrm{T}}$.

For unity matrix Φ_i, i.e., $\Phi_i = E$, we obtain a sequence

$$x_i = x_{i-1}, \tag{6.19}$$

which is a constant vector with expectation $\bar{x}_0$ and covariance matrix P_0. ♦

♦ **Example 6.3.** If $\Phi_i = E$ and $\Gamma_i = E$ in the equation for the shaping filter (6.11),

and there is generating noise, then

$$x_i = x_{i-1} + w_i . \tag{6.20}$$

It is clear that such a sequence can be represented as a sum of two summands, uncorrelated with each other, one of which is a constant random vector, and the other one is a sum of uncorrelated vectors of discrete white noise from the first time point to the current time, i.e.,

$$x_i = x_0 + \sum_{j=1}^{i} w_j . \tag{6.21}$$

Since generating noise is considered zero-mean, then $\overline{x}_i = \overline{x}_0$ for any i. In this case, Equation (6.16) is defined concretely as

$$P_i = P_{i-1} + Q_i .$$

If the covariance matrices for the generating noise are constant, i.e., $Q_i = Q$, we can write:

$$P_i = P_0 + iQ . \tag{6.22}$$

From the last two equations follows that, despite the fact that the matrices defining the shaping filter are time-invariants (constant), the resulting sequence is not stationary (time-invariant) because the corresponding covariance matrix increases with time increase.

Sequence (6.20) is called **a sequence with uncorrelated increments.** This definition is due to the fact that for it, increments $(x_i - x_j)$ on disjoint time intervals are uncorrelated with each other. This is easy to see since on these intervals increments are determined by the values of white noise corresponding to different time points. If all vectors are Gaussian, such increments are independent of each other, and sequence (6.20) is called **a Gaussian sequence with independent increments.**

In the scalar case, sequence (6.20) is called **a Wiener sequence**, also known as a **random walk**. ♦

♦ **Example 6.4.** In section 1.1, we gave an example of how to describe height variation during the measurement time in the form:

$$h_i = x_0 + Vt_i , \; i = \overline{1.m} , \tag{6.23}$$

where x_0, V is the initial height and the vertical velocity, which is assumed constant; $t_i = (i-1)\Delta t$ are the time points from the beginning of the observation.

In section 1.1 this model for height variation was introduced using two unknown variables. Now we can show that this model can be obtained with the use of the shaping filter. Introduce the state vector

$$x_i = (x_{i1}, x_{i2})^{\mathrm{T}} = (h_i, V)^{\mathrm{T}} . \tag{6.24}$$

Since

$$\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{i-1,1} \\ x_{i-1,2} \end{bmatrix},$$

where Δt is the measurement interval, then at $\Gamma_i = 0$, $H_i = [1,0]$,

$$\Phi_i = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \tag{6.25}$$

we obtain

$$z_i = H_i x_i = [1,0] x_i = h_i . \tag{6.26}$$

If the initial covariance matrix is given as

$$P_0 = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_V^2 \end{bmatrix},$$

then using (6.16), it is easy to verify that

$$\sigma_{h_i}^2 = P_i[1,1] = \sigma_0^2 + \sigma_V^2 \left[(i-1)\Delta t\right]^2 .$$

If we add the generating noise with variance q_1^2 to the first equation

$$\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{i-1,1} \\ x_{i-1,2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w_i ,$$

then, instead of (6.23), we obtain the following more general model to describe the height variation:

$$h_i = x_0 + Vt_i + \sum_{j=1}^{i} w_j .$$

Since the matrix of dynamics is given by (6.25), and the matrix at generating noise $\Gamma_i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, in this case, using (6.16), we obtain

$$\sigma_{h_i}^2 = P_i[1,1] = \sigma_0^2 + \sigma_V^2 \left[(i-1)\Delta t\right]^2 + q^2 (i-1).$$

Figure 6.4 gives some examples of height variation realizations obtained by the simulation with the use of the shaping filter and a sensor of random numbers for the two cases mentioned above with the following data: time – 100 s, Δt =1 s, $\sigma_0 = 10\ m$, $\sigma_V = 0.1\text{m/s}$ with $q_1 = 0$ (a) and $q_1 = 1\text{m}$ (b). The triple RMS values define $\pm 3\sigma_{h_i}$ and, in the Gaussian case, they determine the domain of the most probable values of this sequence at each time point.

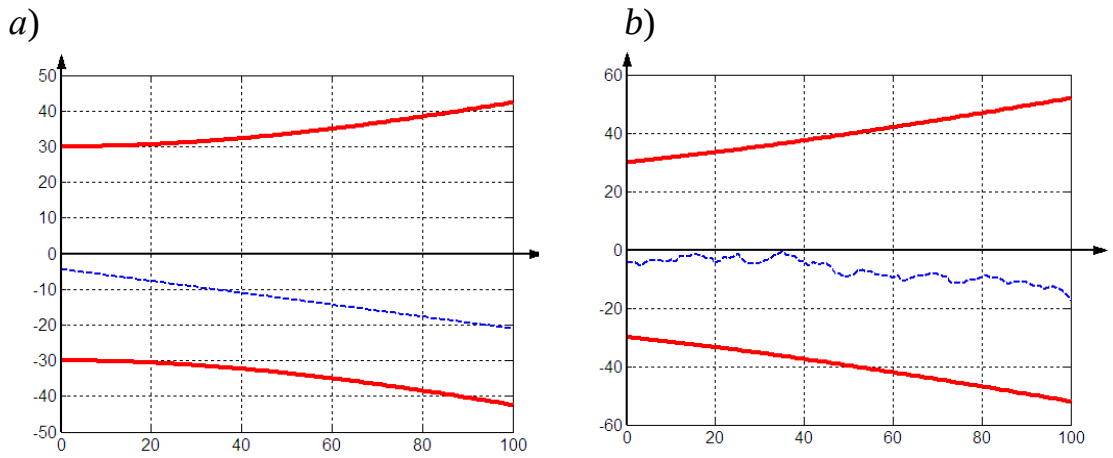


Fig. 6.4. Samples of the height variation and the corresponding triple RMS values in the absence and in the presence of generating noise ♦

6.5. Dynamics of the covariance matrix for the Markov sequence

From (6.16) it follows that in the general case, the covariance matrix of the Markov sequence described with the use of the shaping filter varies with time, and in so doing, P_i can either increase, i.e., $P_i \geq P_{i-1}$, or decrease: $P_i \leq P_{i-1}$.

Let us explain it for a scalar example

$$x_i = \Phi x_{i-1} + w_i, \tag{6.27}$$

where $\Gamma = 1$ and $Q_i = q^2$ for all i. In this case, after introducing notation $P_i = \sigma_i^2$, (6.16) can be written as

$$\sigma_i^2 = \sigma_{i-1}^2 \Phi^2 + q^2.$$

If for Φ^2, q^2 and σ_0^2 the following inequality $\sigma_0^2 \Phi^2 + q^2 > \sigma_0^2$, i.e., $\sigma_0^2 > \dfrac{q^2}{1-\Phi^2}$, is true, then the variance σ_i^2 increases with the increase of i, i.e., $\sigma_i^2 > \sigma_{i-1}^2$. However, if $\sigma_0^2 \Phi^2 + q^2 < \sigma_0^2$, i.e., $\left(\sigma_0^2 < \dfrac{q^2}{1-\Phi^2} \right)$, then variance σ_i^2 decreases. When there exists a solution to the equation $\sigma_\infty^2 \Phi^2 + q^2 = \sigma_\infty^2$, the variance of sequence (6.27) no longer changes with the increase of i and takes the value $\sigma_\infty^2 = \dfrac{q^2}{1-\Phi^2}$. It is possible when $\Phi^2 < 1$. If we assume that $\sigma_0^2 = \dfrac{q^2}{1-\Phi^2}$, the variance will be constant for all i. It is also clear that when $\Phi > 1$, the variance can only increase, which is why there is no positive solution to the equation $\sigma_0^2 \Phi^2 + q^2 = \sigma_0^2$; at $\Phi = 1$, the variance will also increase.

In the vector case, when $\Phi_i = \Phi$, $\Gamma_i = \Gamma$, $Q_i = Q$ are time invariant (constant)

matrices and there exists solution P_∞ for the equation:

$$P_\infty = \Phi P_\infty \Phi^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}}, \qquad (6.28)$$

at zero-mean generating noise, sequence (6.11) is a **stationary Markov sequence.**

♦ **Example 6.5.** Assume that we are given a Wiener sequence (6.20) and $P_0 = \sigma_0^2$, $Q = q^2$.

Consider an example of this sequence and the calculated RMS values corresponding to it at each time point: $\sigma_i = \sqrt{P_i} = \sqrt{\sigma_0^2 + q^2 i}$ at different interrelations between the initial variance and the variances of generating noise. For definiteness, it is assumed that, as in Example 6.4, x_i describes the height variation.

As it was mentioned before, the values of this sequence at each time point represent a sum of two summands: the random value and the sum of the values of the white-noise sequence accumulated by the current time.

At $\sigma_0^2 = 0$, there is no first summand. A sample and the corresponding RMS value is shown as $\pm\sigma_i$ in Fig. 6.5, *a*, at $q = 1$ m, and $\sigma_0 = 0$. Here and in the following example, it is assumed for definiteness that the sequence is formed on the interval of 100 s and Δt =1 s. If the condition $q^2 i = \sigma_0^2$ is satisfied for the last time point *i*=100 (at $\sigma_0 = 10$ m and $q = 1$m), the contribution of two summands in the resulting value of the variance is the same (Fig. 6.5, *b*).

a) *b*)

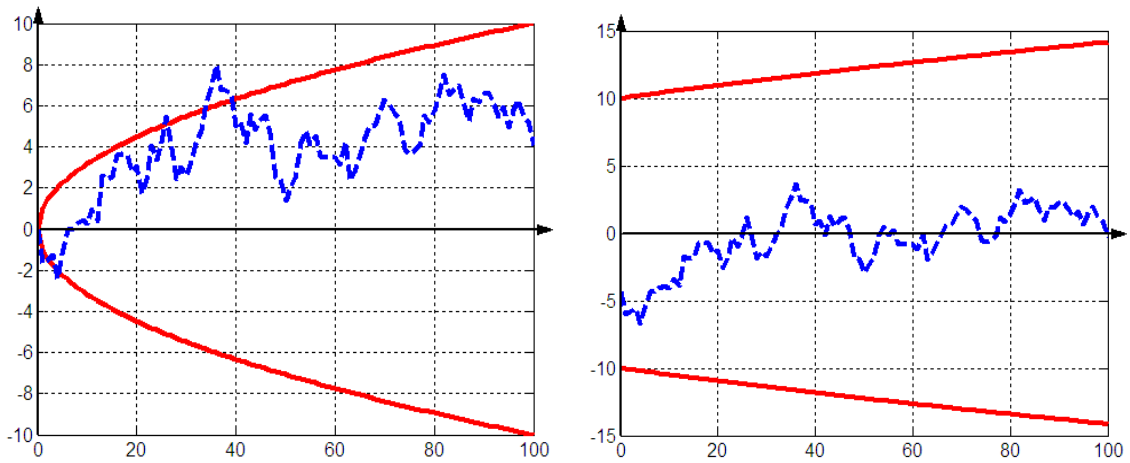


Fig. 6.5. Samples of the Wiener sequence and the corresponding RMS values at zero and nonzero initial values

If the condition $q^2 i << \sigma_0^2$ is satisfied (at $i = 100$, q=1 m, and $\sigma_0 = 100$ m), the samples will not differ significantly from the samples for the constant vector (Fig. 6.6).

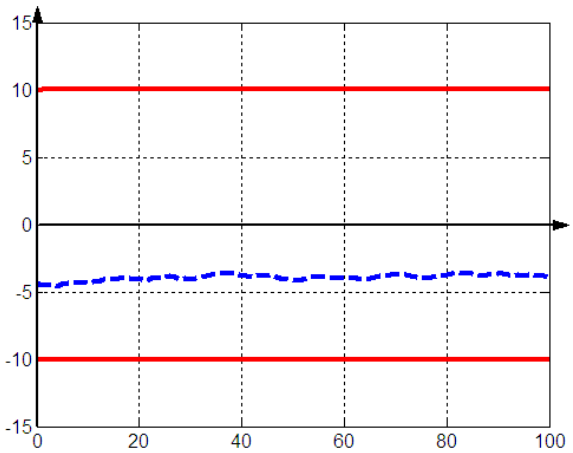


Fig. 6.6. Sample of the Wiener sequence and the corresponding RMS values in the case that condition $q_i^2 \ll \sigma_0^2$ is satisfied

For all three variants in this example, the sequences are nonstationary (time variant), although in the latter case, this nonstationarity does not manifest itself significantly, i.e., the RMS of the variable part is much less than the RMS of the constant component. ♦

♦ **Example 6.6.** Assume that we need to generate a sample of a zero-mean random sequence with the correlation function $k(i-j)=\sigma^2 e^{-\alpha|i-j|}$. Such sequence is usually called **an exponentially-correlated sequence.** It is evident that this can be done using the shaping filter of the form (6.11), where $\Phi = e^{-\alpha}$, $\Gamma = \sigma\sqrt{1-e^{-2\alpha}}$; x_0 is a zero-mean random variable with initial variance $P_0 = \sigma_0^2$, where $\sigma_0 = \sigma$, and w_i is zero-mean discrete white noise with a unit variance, independent of x_0 [25].

This is so because with such parameters, Equation (6.28) becomes an identity: $\sigma^2 = \sigma^2 e^{-2\alpha} + \sigma^2(1-e^{-2\alpha})$, which means that the variance of the sequence is constant. In addition, using the definition of the correlation function, we can write the following equation:

$$k(i,i-1) = k(i-(i-1)) = k(1) = E\{x_i x_{i-1}\} = E\{\Phi x_{i-1} x_{i-1} + \Gamma w_i x_{i-1}\} = \sigma^2 e^{-\alpha},$$

which is not difficult to generalize for the case $i-j>1$.

Below are the plots of the samples of such a sequence with $\alpha = 0.1$, $\sigma = 1m$ for three cases: the stationary case, provided that the conditions $\sigma_0^2\Phi^2 + \Gamma^2 = \sigma_0^2$, $\sigma = \sigma_0 = 1$ m are satisfied (Fig. 6.7); the nonstationary case with increasing variance, provided that the conditions $\sigma_0^2\Phi^2 + \Gamma^2 > \sigma_0^2$, $\sigma > \sigma_0 = 0$ are satisfied (Fig. 6.8, *a*); and the nonstationary case with a decreasing variance, provided that the conditions $\sigma_0^2\Phi^2 + \Gamma^2 < \sigma_0^2$, $\sigma < \sigma_0 = 3m$ are satisfied (Fig. 6.8, *b*). The plots show both the samples and the corresponding triple RMS values in the form of $\pm 3(\sigma_i = \sqrt{P_i})$.

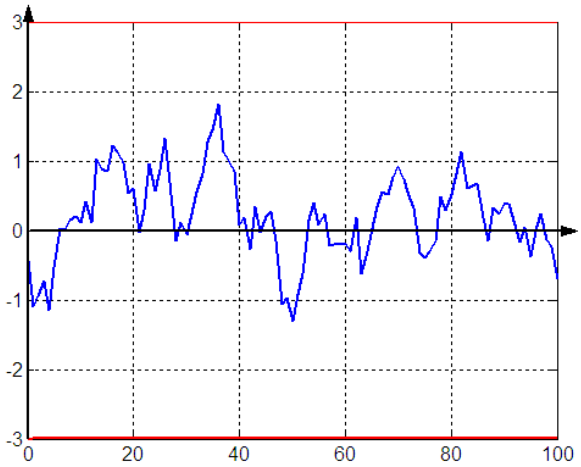


Fig. 6.7. Sample of an exponentially-correlated sequence (stationary case)

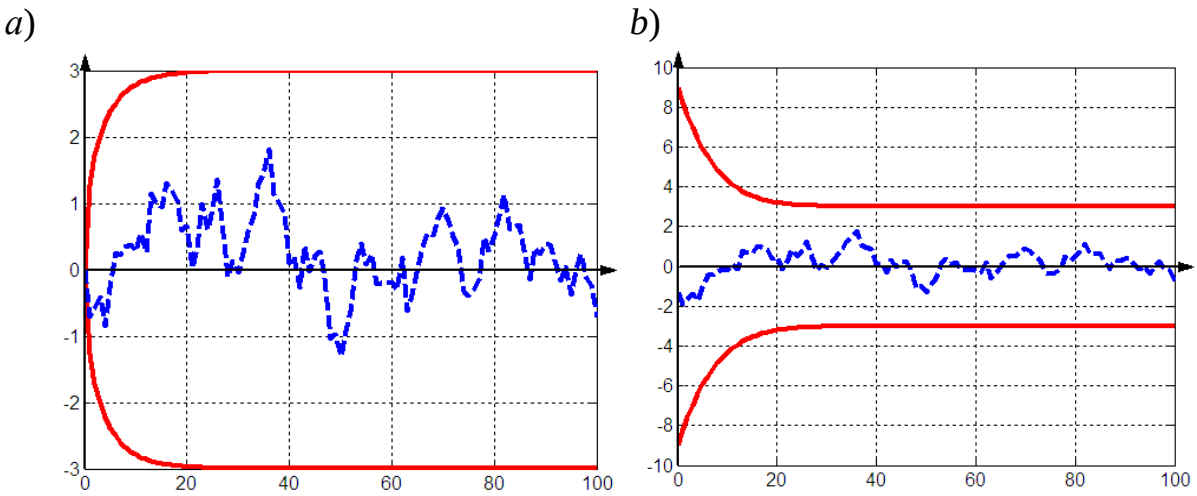


Fig. 6.8. Sample of an exponentially-correlated sequence and the corresponding triple RMS values (nonstationary case)

As it was mentioned above, in all cases, since the condition $\Phi^2 = e^{-2\alpha} < 1$ is satisfied, the variance reaches the steady- state value $\sigma_\infty^2 = \dfrac{\sigma^2(1-e^{-2\alpha})}{1-e^{-2\alpha}} = \sigma^2 = 1.$

From the above formulas it follows that if there is a need to generate a sequence with prescribed σ^2 and α at unit variance of discrete noise, this sequence can be obtained, provided that the condition $\Gamma^2 = \sigma^2(1-\Phi^2) = \sigma^2(1-e^{-2\alpha})$ is satisfied. At $\alpha << 1$, this condition can be written as $\Gamma^2 \approx 2\sigma^2\alpha$. ♦

Exercises

Exercise 6.1. Assume that we have a PDF $p(x_1,.....x_k)$. Show that the representation

$$p(x_1,.....x_k) = p(x_k / x_{k-1}, x_{k-2},..x_1)\, p(x_{k-1} / x_{k-2}, x_{k-3},..x_1)...p(x_1)$$

is valid.

Exercise 6.2. Explain why vectors x_0 and $w_1,....w_i$, having Gaussian PDF (6.13), (6.14), with x_0 uncorrelated with w_l, and w_l, w_j uncorrelated with each other at $l \neq j$, $l, j = \overline{1.i}$, are joint Gaussian.

Exercise 6.3. Show that the changes of the expectation and the covariance matrix in time for sequence (6.11) are defined by (6.15), (6.16):

$$\overline{x}_i = \Phi_i \overline{x}_{i-1};$$

$$P_i = M\left\{(x_i - \overline{x}_i)(x_i - \overline{x}_i)^{\mathrm{T}}\right\} = \Phi_i P_{i-1} \Phi_i^{\mathrm{T}} + \Gamma_i Q_i \Gamma_i^{\mathrm{T}}.$$

Exercise 6.4. Calculate the covariance matrix for the random vector generated according to the equation $x_i = \sum_{j=1}^{i} w_j$, in which w_j, $j = \overline{1.i}$ are uncorrelated with each other zero-mean random vectors with the same covariance matrix Q for each time point.

Exercise 6.5. Calculate the covariance matrix for the random vector generated according to the equation $x_i = \sum_{j=1}^{i} w_j$, in which for all $j = \overline{1.i}$, $w_j \equiv w$, we have the same zero-mean random vector with the covariance matrix Q.

Exercise 6.6. Derive the rule for finding the covariance matrix P^{X_i} of the composite vector $X_i = \left(x_0^{\mathrm{T}}, x_1^{\mathrm{T}},..., x_i^{\mathrm{T}}\right)^{\mathrm{T}}$, the components of which are defined by (6.11), (6.12). find this matrix for the case of $i = 1$.

Exercise 6.7. Derive the formula for a PDF for the composite vector $X_i = \left(x_0^{\mathrm{T}}, x_1^{\mathrm{T}},..., x_i^{\mathrm{T}}\right)^{\mathrm{T}}$ for the Gaussian Markov sequence specified by (6.11)–(6.14).

Exercise 6.8. Assume that in the previous problem, $i = 1$, and $\Gamma_1 = E$, i.e., $x_1 = \Phi_1 x_0 + w_1$. Taking into consideration the fact that the PDF for the composite vector $X_1 = (x_0^{\mathrm{T}}, x_1^{\mathrm{T}})^{\mathrm{T}}$ can be written as (see Problem 6.6)

$$p(X_1) = N(X_1; 0, P_1), \tag{1}$$

where

$$P_1 = \begin{bmatrix} P_0 & P_0 \Phi_1^{\mathrm{T}} \\ \Phi_1 P_0 & \Phi_1 P_0 \Phi_1^{\mathrm{T}} + Q_1 \end{bmatrix}, \tag{2}$$

show that representation (1) for the PDF is identical to that derived in the previous problem.

Exercise 6.9. Derive the formula for the correlation function of a zero-mean Wiener sequence for which the initial condition is zero and the generating noise has identical variance q^2.

Test questions

1. Give a definition of a random sequence. Explain what the mathematical expectation, variance and correlation function for a random sequence mean. Give examples of scalar random sequences.
2. What are zero-mean, stationary and Gaussian random sequences?
3. Give a definition of discrete white noise. Can discrete white noise be nonstationary or non-Gaussian?
4. Give a definition of a Markov sequence. Explain what the shaping filter is. Give some examples.
5. Derive recursive formulas for the mathematical expectation and the covariance matrix of a Markov sequence specified with the use of the shaping filter.
6. What is a Wiener sequence? What properties does it possess? Is the Wiener sequence a Markov sequence?
7. Explain why at constant matrices Φ , Γ , Q in (6.11), (6.12) the sequence is not stationary if condition (6.28) is not satisfied. Illustrate it with an example.

7. Filtering of random sequences. Discrete Kalman filter

In chapters 2–4, we discussed in detail various algorithm design methods depending on the level of a priori information as applied to the problem of estimating the constant vector. Here we consider the problems of estimating the sequences describing time-varying vectors. In so doing, we assume that the sequence is random; therefore, the problem can be formulated within the Bayesian approach as a problem of calculating optimal RMS estimates.

In solving the problem, we also assume that the random sequence is described using a linear shaping filter and we have the measurements linearly depending on the estimated sequence. The main attention is paid to the recursive algorithms, the essence of which is explained in section 5.5. The famous discrete Kalman filter, providing an elegant solution to the problem under discussion, is described.

7.1. Nonrecursive statement and solution of the optimal linear estimation of random sequences

In spite of the fact that the main attention is given to recursive estimation of random sequences, let us show that the problem considered in this chapter is related to the problem of constant parameter vector estimation, and thus, it can be reduced to the latter. Thus, first, we discuss the nonrecursive (batch) statement and the algorithm for solution of the estimation problem.

Consider the following simplest problem of estimating random sequences.

Assume that there are two random scalar sequences x_i and y_i, $i = 1, 2...$. Their statistical properties are set using correlation and cross correlation functions

$$k_x(\nu, \mu) = E\left\{(x_\nu - \bar{x}_\nu)(x_\mu - \bar{x}_\mu)\right\}; \tag{7.1}$$

$$k_y(\nu, \mu) = E\left\{(y_\nu - \bar{y}_\nu)(y_\mu - \bar{y}_\mu)\right\}; \tag{7.2}$$

$$k_{xy}(\nu, \mu) = E\left\{(x_\nu - \bar{x}_\nu)(y_\mu - \bar{y}_\mu)\right\}, \quad \nu, \mu = 1, 2... \tag{7.3}$$

Assume that their mathematical expectations $\bar{x}_j$ and $\bar{y}_j$, $j = \overline{1.i}$ are known.

Let us introduce vector $Y_i = (y_1, ..., y_i)^{\mathrm{T}}$ that includes all measurements y_j $j = \overline{1.i}$ accumulated by the current time *i*.

We need to estimate the values of sequence x_j at some time *j*, $j = \overline{1.i}$ using the measured values of sequence y_j, $j = \overline{1.i}$ accumulated by the current time *i*.

It follows from the given statement that the values of the random sequence should be estimated using the values of another sequence correlated with it.

Different relations between i and j in this statement correspond to different

variants of the estimation problem [19, 24]:

$j = i$ is **a filtering problem**, in the case that the discrete time for which the estimate is sought coincides with the current time (measurement arrival time);

$j < i$ is **an interpolation or smoothing problem**, in the case that the estimate is sought for an earlier time as compared with the current time;

$j > i$ is **an extrapolation or prediction problem**, in the case that the estimate is sought for the future time.

These problems are explained in Fig. 7.1. Filtering and prediction problems are usually solved **online**, when the estimate for the current time should be obtained or the estimated parameter should be predicted at the current time. Smoothing problems are solved in the so-called **off-line mode**, when the estimates are generated after all the measurements have been obtained.

To reveal interconnection between the problems of random sequence and constant vector estimation, introduce a composite vector $\left(x_j^{\mathrm{T}}, Y_i^{\mathrm{T}}\right)^{\mathrm{T}}$ and look at first at nonrecursive estimation algorithms using the whole measurement set $Y_i = (y_1, ..., y_i)^{\mathrm{T}}$ to calculate the estimates of x_j. Introduce $\hat{x}_{j/i}(Y_i)$, which is the estimate x_j at time j calculated by the measurements accumulated by the current time i, i.e.

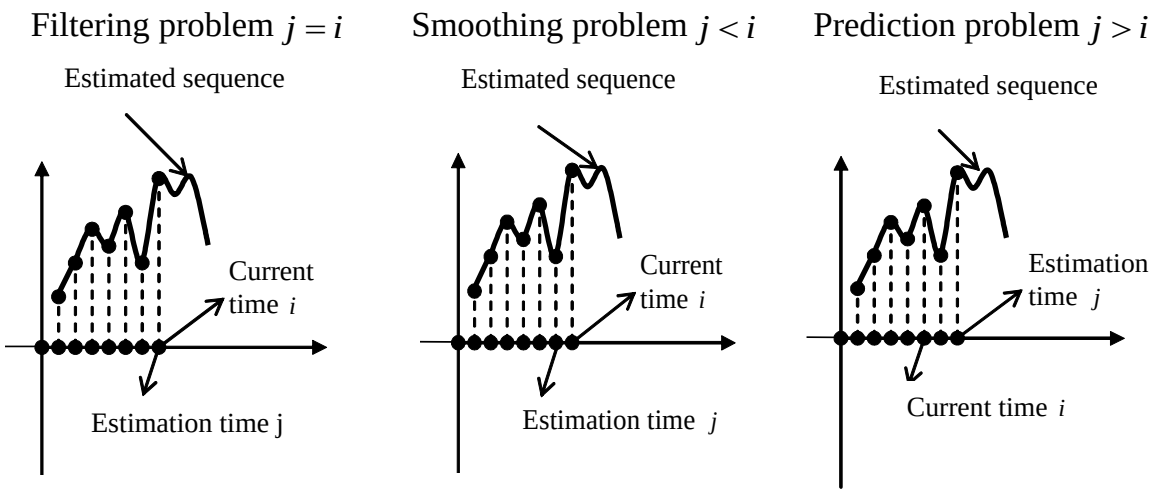


Fig. 7.1. Various types of estimation problems

The statements of the constant vector estimation problem detailed in chapter 2 were concretized depending on the level of the a priori statistical information used. Here, two first moments of vector $\left(x_j^{\mathrm{T}}, Y_i^{\mathrm{T}}\right)^{\mathrm{T}}$ are supposed to be known; thus, similarly to the technique used in 4.5, we can formulate the problem of finding unbiased estimates $\hat{x}_{j/i}(Y_i)$, minimizing the RMS criterion

$$J_{j/i}^{\mathrm{B}} = E\left\{\left(x_j - \hat{x}_{j/i}(Y_i)\right)^2\right\} \tag{7.4}$$

in the class of estimates linearly dependent on the measurements.

These estimates are called **RMS optimal linear estimates of a random sequence or linear unbiased estimates with minimum variance.**

Actually, this is a Bayesian statement of the random sequence estimation problem with a limitation on the class of the estimates used in minimization of the selected criterion. Using the results obtained in 4.5, where the solution algorithm is provided, with account for (4.24), (4.25), we can write the following equation for the sought linear estimate:

$$\hat{x}_j(Y_i) = \bar{x}_j + K_{j/i}(Y_i - \bar{Y}_i), \tag{7.5}$$

where $K_{j/i}$ is a $1 \times i$ row matrix satisfying **the Wiener-Hopf equation for discrete time**

$$K_{j/i} P^{Y_i} = P^{x_j Y_i}, \tag{7.6}$$

where

$$P^{x_j Y_i} = E\left\{(x_j - \bar{x}_j)(y_\mu - \bar{y}_\mu)\right\},\ \mu = \overline{1.i} \tag{7.7}$$

is a row matrix determining the correlation between x_j and values $y_1,....y_i$;

$$P^{Y_i} = E\left\{(y_\nu - \bar{y}_\nu)(y_\mu - \bar{y}_\mu)\right\}, \nu, \mu = \overline{1.i} \tag{7.8}$$

is a $m \times m$ covariance matrix for the whole measurement set.

A posteriori variance for the estimate (7.5) error is given by

$$P^{lin}_{j/i} = P^{x_j} - K_{j/i} P^{Y_i x_i}, \tag{7.9}$$

where

$$P^{x_j} = E\left\{(x_j - \bar{x}_j)^2\right\} \tag{7.10}$$

is the variance of the estimated sequence.

The elements of matrices $P^{x_j Y_i}$, P^{Y_i} and P^{x_j} can be easily found using (7.1)–(7.3).

If matrix P^{Y_i} is nonsingular, the formulas for the estimates and their variances can be written as follows:

$$K_{j/i} = P^{x_i Y_i}\left(P^{Y_i}\right)^{-1}; \tag{7.11}$$

$$P^{lin}_{j/i} = P^{x_j} - P^{x_j Y_i}(P^{Y_i})^{-1} P^{Y_i x_j}. \tag{7.12}$$

As noted in 4.3 (formula (4.8)), the error of the estimate, minimizing the RMS criterion of type (7.4) in the class of estimates linearly dependent on the measurements, is orthogonal to the whole measurement set or their arbitrary combination. In this case, this formula can be written in either equivalent form

$$E\left\{\left(x_j - \hat{x}_j(Y_i)\right)Y_i^{\mathrm{T}}\right\} = 0, \tag{7.13}$$

i.e.,

$$E\left\{x_j Y_i^{\mathrm{T}}\right\} = E\left\{\hat{x}_j(Y_i) Y_i^{\mathrm{T}}\right\}, \tag{7.14}$$

or

$$E\left\{\left(x_j - \hat{x}_j(Y_i)\right)\hat{x}_j(Y_i)\right\} = 0, \tag{7.15}$$

i.e.,

$$E\left\{x_j\hat{x}_j(Y_i)\right\} = E\left\{\hat{x}_j(Y_i)\hat{x}_j(Y_i)\right\}. \tag{7.16}$$

Note that the given solution based on the Wiener-Hopf equation is true for any type of estimation: filtering, smoothing, and prediction.

The results considered above will be fully true also for the case if the sequences are vector sequences, and criterion $J^B_{j/i} = E\left\{(x_j - \hat{x}_{j/i}(Y_j))^T(x_j - \hat{x}_{j/i}(Y_j))\right\}$ is used instead of (7.4).

Assuming that x_i and y_i (i=1, 2..) are n-and m-dimensional sequences with mathematical expectations $\overline{x}_j, \overline{y}_j$ ($j = \overline{1.i}$), for linear optimal estimates, minimizing this criterion, the above formulas will be true. Then, matrices $P^{x_jY_i}$, P^{Y_i}, and P^{x_j} will have dimensions $n\times[i\times m]$, $n\times[i\times m]$, $[i\times m]\times[i\times m]$, and $n\times n$, respectively.

It can be easily noticed that after introduction of vectors $\left(x_j^{\mathrm{T}}, Y_i^{\mathrm{T}}\right)^{\mathrm{T}}$, $j = \overline{1.i}$, the problems of random sequence estimation of x_j for each j in the presented statement are reduced to a set of problems of estimating constant vectors x_j, $j = \overline{1.i}$, considered in chapters 2–4. In this case, when $j = i$, a solution of the filtering problem will be formed; when $j < i$, we will derive a solution of the smoothing problem, and for $j > i$, a solution of the prediction problem. It is very important that in the case when i changes, for example, i increases, in order to obtain a solution, we need to use the whole set of accumulated measurements $Y_i = (y_1,...,y_i)^{\mathrm{T}}$. It is in this sense that the solution obtained is nonrecursive.

It should be emphasized that this problem statement does not imply setting a functional dependence between the measured sequences and those being estimated, and a priori information is given in the form of mathematical expectations, correlation and cross correlation functions (7.1)–(7.3). Actually, it is a generalization of the linear regression problem (considered at the end of section 4.5) for random sequences solved with only their first and second moments available. The obtained solution corresponds to the so-called nonrecursive procedure, where estimates (7.5) are calculated each time using the whole set of available measurements (measurement batch).

♦ **Example 7.1.** Suppose that x_i is a stationary zero-mean random sequence with correlation function $k(i-j) = \sigma^2 e^{-\alpha|i-j|}$, and the measurements of this sequence over the interval $j = \overline{1.i}$ can be represented as

$$y_i = x_i + v_i, \tag{7.17}$$

where v_i is a sequence independent of x_i, which is a zero-mean discrete white noise with variance r^2. The estimate of the sequence at some time j should be found.

For this example, we have

$$P^{x_j} = \sigma^2; \tag{7.18}$$

$$P^{Y_i} = \left\{E\left(x_\nu x_\mu\right) + E\left(v_\nu v_\mu\right)\right\} = \left\{\sigma^2 e^{-\alpha|\nu-\mu|} + r^2\delta_{\nu\mu}\right\}, \quad \nu,\mu = \overline{1.i}, \tag{7.19}$$

$$P^{x_j Y_i} = \left\{E\left(x_j x_\mu\right)\right\} = \left\{\sigma^2 e^{-\alpha|j-\mu|}\right\}, \qquad \mu = \overline{1.i}. \tag{7.20}$$

With account for the above, the following equation for matrix $K_{j/i}$ can be written:

$$K_{j/i} = P^{x_i Y_i}\left(P^{Y_i}\right)^{-1} = \sigma^2\left(e^{-\alpha|j-1|}, e^{-\alpha|j-2|}, \ldots e^{-\alpha|j-i|}\right)\times$$

$$\times\begin{bmatrix} \sigma^2+r^2 & \sigma^2 e^{-\alpha|1|} & & & \sigma^2 e^{-\alpha|1-i|} \\ \sigma^2 e^{-\alpha|1|} & \sigma^2+r^2 & & & \\ & & & & \\ & & & \sigma^2+r^2 & \sigma^2 e^{-\alpha|1|} \\ \sigma^2 e^{-\alpha|1-i|} & & & \sigma^2 e^{-\alpha|1|} & \sigma^2+r^2 \end{bmatrix}^{-1}.$$

Using (7.5), (7.9), and the last equation, we can specify the formula for the estimate in question and its a posteriori error variance. ♦

The given formulas allow determination of estimates and their accuracy characteristics; however, it is difficult to use them in applied problems. In particular, it is connected with the fact that dimensionality of matrix P^{Y_i} grows with the increase in the number of measurements, which makes the matrix inversion complicated. This is the reason is why, further, we focus mainly on recursive filtering problems, where algorithms convenient for implementation can be designed.

7.2. Recursive optimal linear filtering of random sequences. Problem statement

Formulate the problem of designing **recursive algorithms for calculating RMS optimal linear estimates of random sequences,** i.e., the problem of designing **linear recursive optimal algorithms**. Recursive algorithms were considered in section 5.5 as applied to the constant vector estimation problem. Their idea is that the sought estimate for the given measurement set i is generated by successive processing of the next i-th measurement and the estimate obtained at the previous processing step using the measurement set $Y_{i-1} = \left(y_1, \ldots, y_{i-1}\right)^{\mathrm{T}}$. The covariance matrix of the current step is also calculated using the covariance matrix for the previous step. As applied to the considered problem, recursive algorithms can be obtained if the filtering problem is

being solved, i.e., $i = j$. Then the random sequence is assumed to be set using a linear shaping filter of type (6.4), and the measurements linearly depend on the estimated sequence.

In the general case, the problem can be formulated as follows.

Let an n-dimensional random sequence be described by a shaping filter

$$x_i = \Phi_i x_{i-1} + \Gamma_i w_i, \tag{7.21}$$

and let there be m-dimensional measurements related to the sequence as follows:

$$y_i = H_i x_i + v_i, \tag{7.22}$$

where w_i is a p-dimensional vector of generating noise; v_i is an m-dimensional vector of measurement errors; Φ_i, H_i, Γ_i are the matrices of dimensionalities $n \times n$, $m \times n$, $n \times p$. Then, w_i and v_i are discrete zero-mean white noise:

$$E\{w_i w_j^{\mathrm{T}}\} = \delta_{ij} Q_i; \qquad E\{v_i v_j^{\mathrm{T}}\} = \delta_{ij} R_i. \tag{7.23}$$

Initial vector x_0 is considered to be a zero-mean vector with covariance matrix P_0, and vectors x_0, w_i, v_i are considered to be noncorrelated, i.e.,

$$E\left\{w_i v_j^T\right\} = 0,\ E\left\{x_0 v_j^T\right\} = 0,\ E\left\{x_0 w_j^T\right\} = 0. \tag{7.24}$$

Zero-mean nature of random vectors is assumed, without loss of generality, to simplify the obtained formulas.

It is required, using the measurements accumulated by the current time i $Y_i = (y_1^{\mathrm{T}}, y_2^{\mathrm{T}}, \ldots y_i^{\mathrm{T}})^{\mathrm{T}}$, to find a recursive algorithm for calculating RMS optimal unbiased linear estimates of sequence (7.21), minimizing the criterion

$$J_i^B = E\left\{\left(x_i - \hat{x}_{i/i}(Y_i)\right)^{\mathrm{T}}\left(x_i - \hat{x}_{i/i}(Y_i)\right)\right\}, \tag{7.25}$$

and recursive algorithm of calculating estimate error covariance matrices

$$\varepsilon_{i/i}(Y_i) = x_i - \hat{x}_{i/i}(Y_i), \tag{7.26}$$

determined as

$$P_i = E\left\{\left(x_i - \hat{x}_{i/i}(Y_i)\right)\left(x_i - \hat{x}_{i/i}(Y_i)\right)^{\mathrm{T}}\right\}. \tag{7.27}$$

According to the terminology introduced in 7.1, one can speak of the **state vector filtering problem** described, using (7.21), by measurements (7.22) or of the **filtering problem formulated in the state-space context.**

Further, we denote estimate $\hat{x}_{i/i}(Y_i)$ and its error by $\hat{x}_{i/i}(Y_i) \equiv \hat{x}_i(Y_i) \equiv \hat{x}_i$, $\varepsilon_{i/i}(Y_i) \equiv \varepsilon_i(Y_i) = \varepsilon_i$.

It can be easily seen that the constant vector estimation problem considered in chapters 2–4 is a particular case of the sequence filtering problem formulated here. Actually, we obtain this problem if in (7.21) we take $w_i \equiv 0$, $\Gamma_i \equiv 0$, and $\Phi_i \equiv E$.

The following distinctive features of the given problem statement, as compared to the one considered earlier, can be outlined as follows:

- the problem of Markov sequence estimation is being solved;
- the shaping filter for the Markov sequence is linear;
- measurement errors are introduced, and the functional dependence of measurements on the estimated sequence and measurement errors is determined; this dependence is also linear.

These features form the background for obtaining recursive linear algorithms in the form of a discrete Kalman filter. Such algorithms are convenient from the computational viewpoint. The Kalman filter equations are given in the next section.

It should be emphasized that in this statement, as in the previous section, no information is introduced on the PDF for the random sequence, and the solution is to be found in the class of linear estimates.

7.3. Discrete Kalman filter for random sequences

To design an optimal recursive algorithm for the above-formulated problem, we assume that the current estimate $\hat{x}_i$ is calculated using the subsequent measurement y_i, estimate $\hat{x}_{i-1}$, and its error covariance matrix P_{i-1} for the previous step (Fig. 7.2).

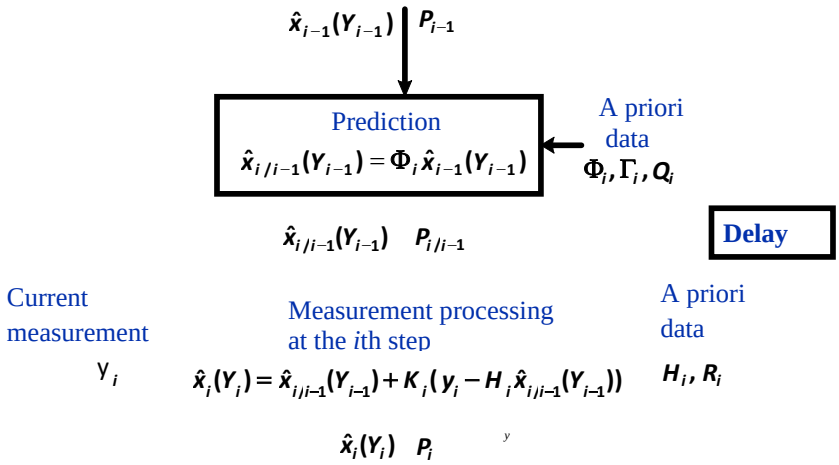


Fig. 7.2. Kalman filter block diagram for discrete time

In section 5.5, a recursive algorithm for the constant vector estimation problem is derived. It is emphasized that at each step, the estimate obtained using previous measurements is actually corrected by adding a summand being a product of the gain factor and the difference between the current measurement and its prediction. It is reasonable to assume that the structure of the recursive algorithm for a random sequence set using the shaping filter (7.21) will remain unchanged. The peculiarity of the algorithm structure, as applied to the changing sequence, manifests itself in the way of how to find the prediction estimate of the sequence and the prediction of the current

measurement, using the optimum estimate $\hat{x}_{i-1}$, corresponding to Y_{i-1}. Clearly, it should be done with account for the dynamics equation (7.21) for the sequence being estimated.

Taking into consideration the above-said, it can be shown that the recursive algorithm for the calculation of the RMS optimal linear unbiased estimates, minimizing the RMS criterion, can be obtained using a set of formulas divided into two blocks (see Fig. 7.2).

In the first block (prediction block)**,** the linear optimal estimate of the state vector is calculated at time i by measurements Y_{i-1}, i.e., the **prediction estimate** $\hat{x}_{i/i-1}$ and its **prediction error covariance matrix** are calculated using the optimal estimate $\hat{x}_{i-1}$ and covariance matrix P_{i-1} at the previous step:

$$\hat{x}_{i/i-1} = \Phi_i \hat{x}_{i-1}; \tag{7.28}$$

$$P_{i/i-1} = \Phi_i P_{i-1} \Phi_i^{\mathrm{T}} + \Gamma_i Q_i \Gamma_i^{\mathrm{T}}, \tag{7.29}$$

where

$$P_{i/i-1} = E\left\{ \left(x_i - \hat{x}_{i/i-1}(Y_{i-1})\right)\left(x_i - \hat{x}_{i/i-1}(Y_{i-1})\right)^{\mathrm{T}} \right\} = E\left\{ \left(\varepsilon_{i/i-1}\right)\left(\varepsilon_{i/i-1}\right)^{\mathrm{T}} \right\};$$

$$\varepsilon_{i/i-1} = x_i - \hat{x}_{i/i-1}(Y_{i-1}).$$

In the second block, the sought current estimate $\hat{x}_i$ and its error covariance matrix P_i are calculated using the current measurement y_i and the results obtained in the prediction block:

$$\hat{x}_i = \hat{x}_{i/i-1} + K_i(y_i - H_i \hat{x}_{i/i-1}); \tag{7.30}$$

$$K_i = P_{i/i-1} H_i^{\mathrm{T}} (H_i P_{i/i-1} H_i^{\mathrm{T}} + R_i)^{-1}; \tag{7.31}$$

$$P_i = P_{i/i-1} - P_{i/i-1} H_i^{\mathrm{T}} (H_i P_{i/i-1} H_i^{\mathrm{T}} + R_i)^{-1} H_i P_{i/i-1} = (E_n - K_i H_i) P_{i/i-1}. \tag{7.32}$$

These formulas determine the well-known **Kalman filter for discrete time (discrete Kalman filter)**. The recursive nature of this algorithm is obvious because to find the predicted estimate and its error covariance matrix, apart from a priori data,—matrices Φ_i, Γ_i, and Q_i, we use only the estimate and its error covariance matrix obtained at the previous step; at the same time, to calculate the next estimate and its covariance matrix, we use only the results obtained in the prediction block, the current measurement, and a priori data: matrices H_i and R_i. Despite the fact that only the current measurement is used, the resultant estimate is optimal over the whole measurement set (it minimizes criterion (7.25)), and in the block diagram, the estimates depend on the relevant measurement sets.

Equations (7.30) are usually referred to as the **state estimate observational update,** or simply the **state estimate update,** and (7.32), as the **error covariance update** [15, 24]. Matrix K_i is called the **Kalman gain matrix** or just the **Kalman gain**.

As in 4.4, it can be shown that the following formulas are convenient in some applications will be also true for matrices K_i, P_i:

$$K_i = P_i H_i^{\mathrm{T}} R_i^{-1}; \tag{7.33}$$

$$P_i = (P_{i/i-1}^{-1} + H_i^{\mathrm{T}} R_i^{-1} H_i)^{-1}. \tag{7.34}$$

Calculation of the gain and the covariance matrix according to (7.31), (7.32) is preferable if $m << n$, while it is more practical to use (7.33), (7.34) with $n << m$ and diagonal matrix R_i.

It should be mentioned that error covariance matrices P_i do not depend on measurements and are determined only by observation matrices H_i, covariance matrices Q_i and R_i, characterizing the properties of generating and measurement noise.

Thus, all the computations connected with the calculation of the covariance matrix and the gain can actually be done in advance. The block implementing these computations is sometimes referred to as a **covariance block (channel)**, and the one computing estimates (7.30), as **an estimation block (channel)**. It is significant that the Kalman filter is linear with respect to the measurements, and the gain depends only on matrices Q_i, R_i and H_i; therefore, it is independent of the measurements.

The Kalman filter equations were first provided and proven in the well-known work by R.E. Kalman [17]. The proof is based on the use of the orthogonality property.

After that publication many different approached have been developed proving the optimality of Kalman filter estimates.

It should be once more emphasized that here we consider the problem of obtaining the estimates optimal in the linear class, with no assumptions made on the PDF of random sequences. If PDF is assumed to be Gaussian, the above given algorithm will provide optimal Bayesian estimates, i.e., the estimates minimizing the RMS criterion (7.4) without limitations on the class of the estimates used. Thus, one of the proofs can be based on obtaining a recursive algorithm for calculating optimal Bayesian estimates of the problem under consideration, assuming that the initial vector, generating and measurement noises are Gaussian.

It should be also noted that the Kalman filter algorithm not only defines a convenient procedure for calculating the estimates, which solves the **problem of designing an algorithm for estimation of a random sequence**, but also the procedure for computing the covariance matrix characterizing the current accuracy of the estimation algorithm, which is important for **analyzing the estimation accuracy of a random sequence.**

Particularly, diagonal elements determine calculated estimate error variances, which in their turn determine calculated standard deviations of estimate errors for all components of the state vector.

The formulated problem and the presented formulas are given in Table 7.1
[8, 9, 13].

Table 7.1

Equations for discrete Kalman filter algorithm

Statement of the filtering problem	
State vector equation	$x_i = \Phi_i x_{i-1} + \Gamma_i w_i$
Measurements	$y_i = H_i x_i + v_i$
Initial conditions	$\overline{x}_0 = 0,\ P_0$
Generating (system) noise	$\overline{w}_i = 0,\ \ E\{w_i w_j^{\mathrm{T}}\} = \delta_{ij} Q_i;$
Measurement noise	$\overline{v}_i = 0,\ \ E\{v_i v_j^{\mathrm{T}}\} = \delta_{ij} R_i;$
Cross correlation	$E\left\{x_0 w_i^{\mathrm{T}}\right\} = 0;\ E\left\{w_i v_i^{\mathrm{T}}\right\} = 0;\ E\left\{x_0 v_i^{\mathrm{T}}\right\} = 0$
Matrices	$\Phi_i, -\ n \times n,\ \Gamma_i$ - $n \times p,\ Q_i$ - $p \times p$ $H_i - m \times n,\ R_i - m \times m$
Minimized criterion $J^B = E_{x_i, Y_i}\left\{\left(x_i - \hat{x}_i(Y_i)\right)^{\mathrm{T}}\left(x_i - \hat{x}_i(Y_i)\right)\right\}$	
Solution of the filtering problem	
Prediction	$\hat{x}_{i/i-1} = \Phi_i \hat{x}_{i-1}$
Prediction error covariance matrix	$P_{i/i-1} = \Phi_i P_{i-1} \Phi_i^{\mathrm{T}} + \Gamma_i Q_i \Gamma_i^{\mathrm{T}}$
Estimate	$\hat{x}_i = \hat{x}_{i/i-1} + K_i(y_i - H_i \hat{x}_{i/i-1})$
Gain	$K_i = P_{i/i-1} H_i^{\mathrm{T}} (H_i P_{i/i-1} H_i^{\mathrm{T}} + R_i)^{-1}$ variant *1* $K_i = P_i H_i^{\mathrm{T}} R_i^{-1}$ variant *2*
Estimate error covariance matrix	$P_i = (E_n - K_i H_i) P_{i/i-1}$ variant *1* $P_i = (P_{i/i-1}^{-1} + H_i^{\mathrm{T}} R_i^{-1} H_i)^{-1}$ variant *2*

In this statement, the Kalman filter is used to calculate the estimates optimal in the linear class, so that one can say that the Kalman filter is used to calculate the **potential accuracy of the random sequence (7.21) estimation by measurements (7.22) using linear estimates.**

Specify the provided Kalman filter equations using the following simple examples.

♦ **Example 7.2.** Obtain the Kalman filter equations for the problem of estimating scalar parameter $x_i = x_{i-1} = x$ by scalar measurements $y_i = x + v_i$, where $x \equiv x_0$ is a zero-mean random value with variance σ_0^2; v_i is a zero-mean discrete white noise independent of x, i.e., $E\{v_i v_j\} = \delta_{ij} r_i^2$.

Here, $\Phi = H = 1$, $\Gamma = Q \equiv 0$. While the parameter being estimated does not change, the prediction block is significantly simplified:

$$\hat{x}_{i/i-1} = \hat{x}_{i-1};$$
$$P_{i/i-1} = P_{i-1}.$$

Thus, the formula for the estimate takes the form:

$$\hat{x}_i = \hat{x}_{i-1} + K_i(y_i - \hat{x}_{i-1}).$$

The gain and the error variance of the optimal estimate can be calculated with the use of (7.33), (7.34):

$$K_i = \frac{P_{i-1}}{P_{i-1} + r_i^2};$$
$$P_i = (P_{i-1}^{-1} + R_i^{-1})^{-1} = \frac{P_{i-1} r_i^2}{P_{i-1} + r_i^2}, P_0 = \sigma_0^2.$$

If the variances of all measurements are assumed to be identical, i.e., $r_i^2 = r^2$, it can be easily verified that

$$P_i = \frac{P_0 r^2}{iP_0 + r^2}.$$

It follows from here that if a priori variance is significantly larger than the measurement variance, i.e., $P_0 >> r^2$, the equations for the gain and a posteriori variance can be determined as $K_i = \frac{1}{i}$, $P_i = \frac{r^2}{i}$, and thus, $\hat{x}_i = \hat{x}_{i-1} + \frac{1}{i}(y_i - \hat{x}_{i-1})$.

As could be expected, the last equation in this case is a recursive formula for calculating the arithmetic mean i.e., $\hat{x}_i = \frac{1}{i}\sum_{j=1}^{i} y_j$.

Figure 7.3 shows the filtering error samples and the tripled calculated standard deviations in the form $\pm 3(\sigma_i = \sqrt{P_i})$ for the case $P_0 = \sigma_0^2 = 1$, $r^2 = 1$.

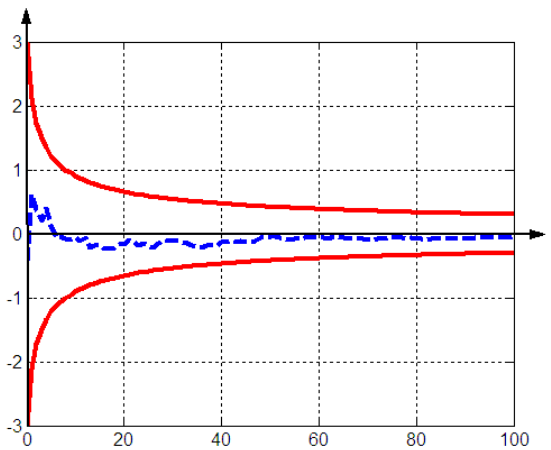


Fig. 7.3. Filtering errors and tripled calculated standard deviations in the estimation of the constant value

As follows from the plot, the standard deviation of the estimation error unlimitedly decreases, in this case, going down to about 0.1 over the interval considered here.

♦ **Example 7.3.** It is required to solve the problem of filtering the Wiener sequence $x_i = x_{i-1} + w_i$ by measurements $y_i = x_i + v_i$, where x_0 is a zero-mean random variable with variance σ_0^2; w_i, v_i are zero-mean discrete white noises independent of each other and of x_0 such that $E\{v_i v_j\} = \delta_{ij} r^2$, $E\{w_i w_j\} = \delta_{ij} q^2$.

In this case, $\Phi = H = 1$, $\Gamma = 1$. The parameter being estimated is varying; however, the equations for the estimates are the same as in the previous example, i.e.,

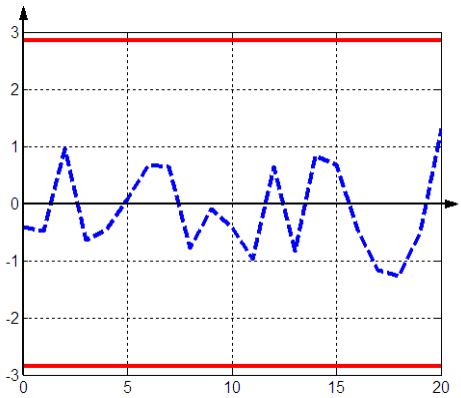
$$\hat{x}_{i/i-1} = \hat{x}_{i-1}; \qquad \hat{x}_i = \hat{x}_{i-1} + K_i(y_i - \hat{x}_{i-1}).$$

Take the following formulas for the prediction and estimation error variances:

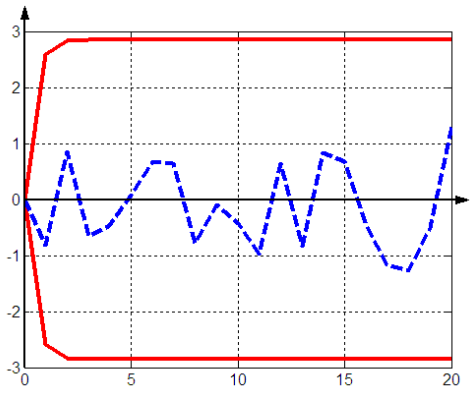
$$P_{i/i-1} = P_{i-1} + q^2;$$

$$P_i = P_{i/i-1}\left(1 - \frac{P_{i/i-1}}{P_{i/i-1} + r^2}\right) \quad \text{or} \quad P_i = \left(P_{i-1} + q^2\right)\left(\frac{r^2}{P_{i-1} + q^2 + r^2}\right).$$

$a)\ P_0 = \sigma_0^2 = P_\infty = 1$



$b)\ P_\infty > P_0 = \sigma_0^2 = 0$



$c)\ (P_0 = 3^2) > P_\infty = 1$

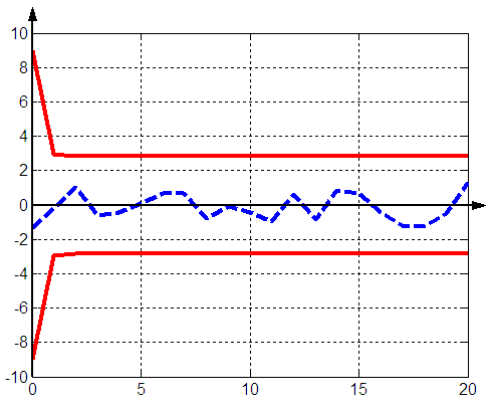


Fig. 7.4. Wiener sequence filtering errors and tripled standard deviations:
a – stationary case, i.e., error variance is constant; *b* – error variance increases and reaches a steady-state value; *c* – error variance decreases and reaches a steady-state value

Figure 7.4 presents the results of solving the problem of filtering a Wiener sequence over 20 steps with $q = 1$, $r^2 = 4$ as filtering errors and tripled standard deviations ($\pm 3(\sigma_i = \sqrt{P_i})$ with different values $P_0 = \sigma_0^2$. ♦

♦ **Example 7.4.** Obtain Kalman filter equations for the problem of filtering a scalar stationary exponentially correlated sequence with a correlation function $k(i-j) = \sigma^2 e^{-\alpha|i-j|}$ considered in 7.1.

With account for Example 6.6, Eqs. (7.21), (7.22) in this case are realized as follows: $x_i = \Phi x_{i-1} + w_i$; $y_i = x_i + v_i$, where $\Phi = e^{-\alpha}$, x_0 is a zero-mean random variable with variance $P_0 = \sigma^2$; w_i, v_i are zero-mean white noises independent of x_0 with variances $q^2 = \sigma^2(1 - e^{-2\alpha})$ and r^2.

The equations for the prediction block can be written as follows:

$$\hat{x}_{i/i-1} = \Phi \hat{x}_{i-1}; \qquad P_{i/i-1} = \Phi^2 P_{i-1} + q^2,$$

and the equations for estimates and their error variances as

$$\hat{x}_i = \Phi \hat{x}_{i-1} + K_i(y_i - \Phi \hat{x}_{i-1})$$

$$P_i = \left(\frac{1}{P_{i/i-1}} + \frac{1}{r^2} \right)^{-1} = \left(\frac{1}{\Phi^2 P_{i-1} + q^2} + \frac{1}{r^2} \right)^{-1}, \text{ where } K_i = \frac{P_i}{r^2}.$$

Figure 7.5 shows an example of filtering errors of an exponentially correlated sequence and tripled calculated standard deviations in the form $\pm 3(\sigma_i = \sqrt{P_i})$, with $\alpha = 0.1$, $\sigma = 1$ m, $q^2 = 1(1 - e^{-0,2}) = 0.2$ m^2 and two values of measurement error variances: $r^2 = 1$ m^2 (*a*) and $r^2 = (0.1)^2$ m^2 (*b*).

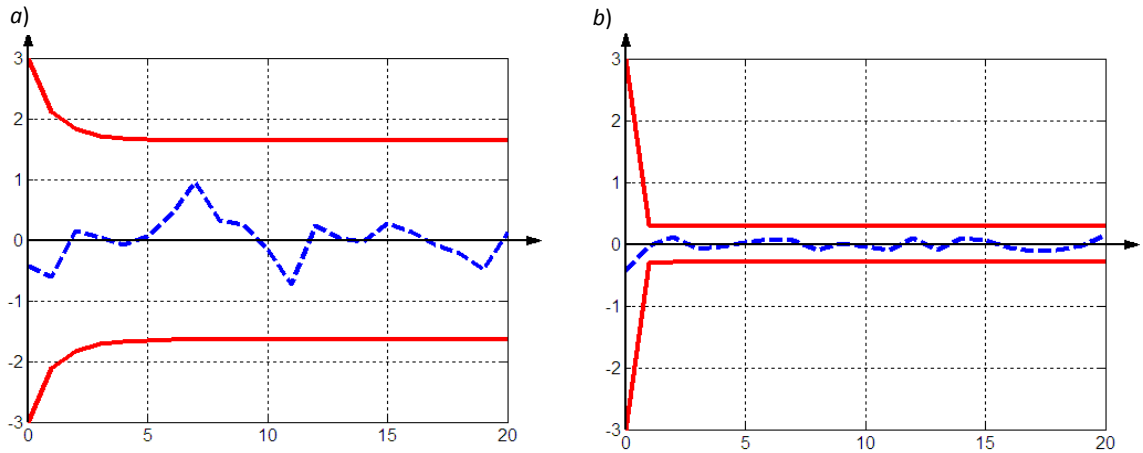


Fig. 7.5. Filtering errors and tripled calculated standard deviations with different levels of measurement errors ♦

As can be seen from the plots, the variances reach some steady-state values, which strongly depend on the level of measurement errors and generating noise. The reasons and conditions for reaching the steady-state value are discussed below.

7.4. Kalman filter error equations. Innovation sequence

Subtracting Equation (7.30) from the equation of shaping filter (7.21), equations for the Kalman filter **prediction and estimation errors** can be easily obtained:

$$\varepsilon_{i/i-1} = \Phi_i x_{i-1} + \Gamma_i w_i - \Phi_i \hat{x}_{i-1} = \Phi_i \varepsilon_{i-1} + \Gamma_i w_i ; \tag{7.35}$$

$$\begin{aligned} \varepsilon_i &= \Phi_i x_{i-1} + \Gamma_i w_i - \Phi_i \hat{x}_{i-1} - K_i (H_i x_i - H_i \hat{x}_{i/i-1}) + K_i v_i = \\ &= \Phi_i \varepsilon_{i-1} + \Gamma_i w_i - K_i H_i (\Phi_i \varepsilon_{i-1} + \Gamma_i w_i) + K_i v_i = \\ &= (E_n - K_i H_i)\Phi_i \varepsilon_{i-1} + (E_n - K_i H_i)\Gamma_i w_i + K_i v_i . \end{aligned} \tag{7.36}$$

From the last formula it follows, in particular, that the linear optimal estimate error is a Markov sequence, and the following recursive relationship is true for it:

$$\varepsilon_i = (E_n - K_i H_i)\Phi_i \varepsilon_{i-1} + (\Gamma_i - K_i H_i) w_i + K_i v_i , \tag{7.37}$$

which contains both generating noise and measurement noise in the right-hand part.

Equations (7.29), (7.32), or, which is the same, (7.34), will be true for error covariance matrices (7.35), (7.36), which can be easily verified (see problem 7.1), using the equation for the gain (7.31).

Clearly, the orthogonality property (7.15) will be satisfied for the estimates calculated using the Kalman filter

$$E\left\{ \left(x_i - \hat{x}_i(Y_i) \right) \hat{x}_i(Y_i) \right\} = 0 . \tag{7.38}$$

It should be noted that

$$E\left\{ \varepsilon_i(Y_i) x_i^{\mathrm{T}} \right\} = E_{x_i Y_i} \left\{ (x_i - \hat{x}_i^{\mathrm{T}}(Y_i)) x_i^{\mathrm{T}} \right\} = P_i , \tag{7.39}$$

(see problem 7.3).

In analyzing the properties of Kalman filter estimates, **measurement residual** is of great importance, which is defined here as

$$\mu_i = y_i - H_i \hat{x}_{i/i-1}(Y_{i-1}) . \tag{7.40}$$

This sequence possesses an important feature, namely, it is a discrete zero-mean white noise, i.e.,

$$E\left\{ \mu_i \mu_j^{\mathrm{T}} \right\} = \delta_{ij} L_i , \tag{7.41}$$

where

$$L_i = H_i P_{i/i-1} H_i^{\mathrm{T}} + R_i . \tag{7.42}$$

Formula (7.42) determines the residual covariance matrix, which can be easily proven (see problem 7.2). Let us prove that (7.41) is true. Since $E_{y_i/Y_{i-1}}\{ y_i \} = H_i \hat{x}_{i/i-1}(Y_{i-1}),$ then with $j < i$,

$$E_{Y_i}\left\{\left(y_i - H_i\hat{x}_{i/i-1}(Y_{i-1})\right)y_j^{\mathrm{T}}\right\} = E_{Y_{i-1}}E_{y_i/Y_{i-1}}\left\{\mu_i y_j^{\mathrm{T}}\right\} = 0,$$

i.e.,

$$E\left\{\mu_i y_j^{\mathrm{T}}\right\} = 0, \text{ for } j < i. \tag{7.43}$$

Similarly, $E\left\{\mu_i\mu_j^{\mathrm{T}}\right\} = 0$, for $j < i$. If $j > i$, this equality can be easily obtained by verifying that $E\left\{\mu_i y_j^{\mathrm{T}}\right\} = 0$. Thus, the last equality is true for any nonidentical j and i, and, therefore, (7.41) is also true. It follows from the proven statement that random sequence y_i is transformed into white noise using (7.40). This operation is sometimes called **whitening.**

It follows from (7.40), (7.43) that the residual at the current time is orthogonal to the set of previous measurements $Y_{i-1} = \left(y_1,...,y_{i-1}\right)^{\mathrm{T}}$, and the current measurement can be represented as a sum

$$y_i = H_i\hat{x}_{i/i-1}(Y_{i-1}) + \mu_i.$$

Note that the equation for the Kalman filter estimate can be written as a sum of two summands:

$$\hat{x}_i = \hat{x}_{i/i-1} + K_i\mu_i(Y_{i-1}), \tag{7.44}$$

the first summand is a linear combination of the previous measurement set $Y_{i-1} = \left(y_1,...,y_{i-1}\right)^{\mathrm{T}}$, and the second is a linear combination of the residual vector components $\mu_i(Y_{i-1})$ orthogonal to Y_{i-1}. Since this vector contains new data that are not present in Y_{i-1}, the data are updated due to the use of μ_i, which explains the term used for the residual (7.40): **an innovation sequence** [24].

♦ **Example 7.5.** Specify the given equations for estimation errors applied to the problem of filtering the constant value and the Wiener sequence.

For the constant value:

$$\varepsilon_{i/i-1} = \varepsilon_{i-1};$$

$$\varepsilon_i = (1-K_i)\varepsilon_{i-1} + K_i v_i;$$

$$\mu_i = y_i - \hat{x}_{i/i-1}(Y_{i-1}) = \varepsilon_{i-1} + v_i;$$

$$L_i = P_{i-1} + r^2.$$

For the Wiener sequence:

$$\varepsilon_{i/i-1} = \varepsilon_{i-1} + w_i;$$

$$\varepsilon_i = (1-K_i)\varepsilon_{i-1} + (1-K_i)w_i + K_i v_i;$$

$$\mu_i = y_i - \hat{x}_{i/i-1}(Y_{i-1}) = \varepsilon_{i/i-1} + v_i = \varepsilon_{i-1} + w_i + v_i;$$

$$L_i = P_{i-1} + q^2 + r^2 .$$

It should be noted that the measurement residual generated as a difference between the measured values and the current estimate

$$\tilde{\mu}_i = y_i - \hat{x}_i(Y_i) = \varepsilon_i \ + v_i = (1-K_i)\varepsilon_{i-1} + (1-K_i)w_i + (1+K_i)v_i ,$$

is not white noise, unlike the residual which is the difference between the measured values and the prediction estimate. ♦

7.5. Dynamics of the covariance matrix and steady-state mode in the filtering problem

When studying errors of the Kalman filter, it is important to investigate how their covariance matrices vary with time. In scalar examples above, it was noted that a filtering error variance can reach some steady-state value. Let us discuss this question in more detail.

Analysis of Equation (7.32) for the covariance matrix shows that it has two summands. The first summand

$$P_{i/i-1} = \Phi_i P_{i-1} \Phi_i^{\mathrm{T}} + \Gamma_i Q_i \Gamma_i^{\mathrm{T}} \tag{7.45}$$

shows that the error covariance matrix varies during calculation of the prediction error, which is partly conditioned by the generating (system) noise. To analyze the character of this variation, introduce matrix $\Delta P_i^{(1)} = P_{i/i-1} - P_{i-1}$ and present it as $P_{i/i-1} = P_{i-1} + \Delta P_i^{(1)}$. From the results of (6.5) and Equation (7.45) it follows that the prediction error covariance matrix $P_{i/i-1}$ can both increase and decrease as compared with the error covariance matrix P_{i-1} at the previous step. Decrease is possible if the covariance (variance) matrix of the sequence being estimated decreases. However, in applied problems, the error covariance matrix of the sequence being estimated usually grows. It means, in particular, that estimate error variances of each component defined by diagonal elements can only increase or at least remain unchanged. This behavior is due to the fact that the state vector value changes according to the equation of the shaping filter (6.11). Its right-hand part contains generating noise, which increases uncertainty in the knowledge of the state vector at each step. Even if the state vector was accurately known at some time, for example, at $P_{i-1} = 0$, during the prediction, covariance matrix $P_{i/i-1} = \Gamma_i Q_i \Gamma_i^{\mathrm{T}}$ grows and becomes nonzero since $\Gamma_i Q_i \Gamma_i^{\mathrm{T}} \geq 0$, with $Q > 0$. It is discussed in more detail in example 7.6.

The second summand in the equation for the covariance matrix, which shows the effect of the next (current) measurement, can be represented as follows:

$$\Delta P_i^{(2)} = P_{i/i-1} - P_i = P_{i/i-1} H_i^{\mathrm{T}} (H_i P_{i/i-1} H_i^{\mathrm{T}} + R_i)^{-1} H_i P_{i/i-1} .$$

Since we have a nonnegatively definite matrix in the right-hand part, then $P_{i/i-1} - P_i \geq 0$ and, therefore, $P_i \leq P_{i/i-1}$, i.e., the error covariance matrix can only

decrease or remain unchanged as compared with the prediction error covariance matrix. This relationship seems quite logical since the use of new measurement data cannot degrade the estimation accuracy.

Therefore, the following can be written for the covariance matrix

$$P_i = P_{i-1} + \Delta P_i^{(1)} - \Delta P_i^{(2)}.$$

Clearly, if summands $\Delta P_i^{(1)}$ and $\Delta P_i^{(2)}$ are equal, the filtering error covariance matrix will not depend on time, i.e., it will be constant. Then we can speak of the **steady-state mode of the filtering problem**. This mode can be used, for example, to solve filtering problems, where all matrices used in Kalman filter equations (Φ, Γ, H, Q, R) are constant. We have already noted the existence of the steady-state solution when the simulation results of the filtering problems for the Wiener and exponentially-correlated sequences are considered in examples 7.3, 7.4. Clearly, for the steady-state mode to be feasible, there must be a solution to the following equation:

$$P_\infty^f = \Phi P_\infty^f \Phi^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}} - (\Phi P_\infty^f \Phi^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}}) H^{\mathrm{T}} \times \left(H(\Phi P_\infty^f \Phi^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}}) H^{\mathrm{T}} + R \right)^{-1} \times \quad (7.46)$$
$$\times H(\Phi P_\infty^f \Phi^{\mathrm{T}} + \Gamma Q \Gamma^{\mathrm{T}}).$$

Here, P_∞^f determines the filtering error covariance matrix for the steady-state mode. If condition (7.46) is met, the prediction error covariance matrix and the gain also remain unchanged:

$$P_\infty^{pr} \equiv \Phi P_\infty \Phi^T + \Gamma Q \Gamma^T;$$
$$K_\infty = P_\infty^f H^{\mathrm{T}} R^{-1}.$$

Thus, the Kalman filter turns into a steady-state difference equation:

$$\hat{x}_i = \Phi \hat{x}_{i-1} + K_\infty (y_i - H\, \Phi \hat{x}_{i-1}) = (E_n - K_\infty H) \Phi \hat{x}_{i-1} + K_\infty y_i,$$

which is the **Wiener filter for discrete time.**

♦ **Example 7.6.** Analyze possible dynamics of error variance and obtain the Wiener filter for the problem considered in 7.3, i.e., the problem of estimating the Wiener sequence by its measurements with white-noise errors.

Since $\Phi = 1$, $P_{i/i-1} = P_{i-1} + q^2$, at each prediction step, the error variance increases by a value equal to the variance of generating noise q^2, i.e., $\Delta P_i^{(1)} = q^2$ (Fig. 7.6).

As the next measurement is processed, as shown in Fig. 7.6, the error variance decreases since

$$\Delta P_i^{(2)} = P_{i/i-1} - P_i = P_{i/i-1} \left(\frac{P_{i/i-1}}{P_{i/i-1} + r^2} \right) \geq 0.$$

If $\Delta P_i^{(2)} < q^2$, at the next step, the error variance will exceed the variance at the previous step, i.e., $P_i > P_{i-1}$, and will grow as the number of measurements increases. If $\Delta P_i^{(2)} > q^2$, the error variance will decrease with the growth of the number of measurements: $P_i < P_{i-1}$.

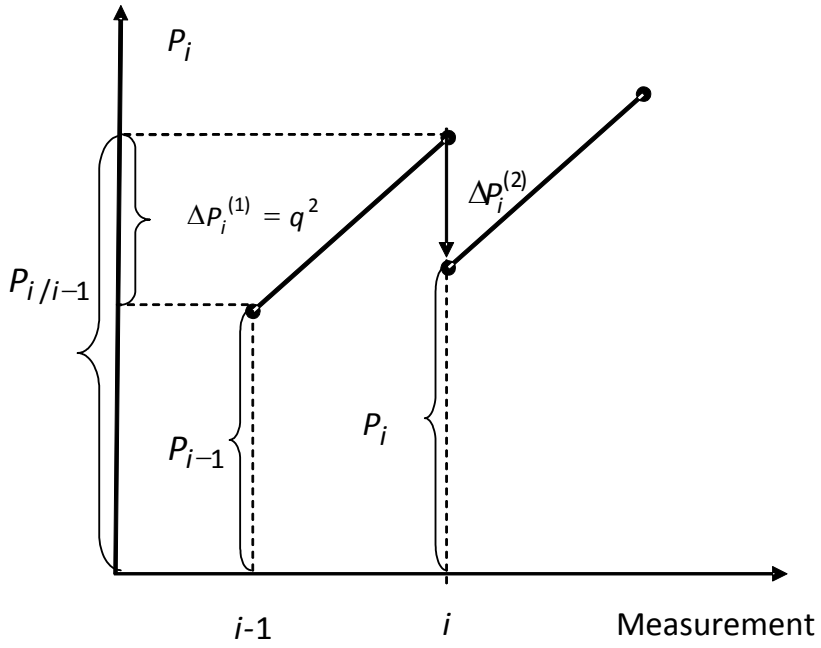


Fig. 7.6. Dynamics of filtering and prediction error variances in estimating the Wiener sequence

With $\Delta P_i^{(1)} = \Delta P_i^{(2)}$ the variances at the current and previous steps will be the same, and with given q^2 and r^2, their values can be found by solving the equation

$$\frac{(P_{i-1} + q^2)^2}{P_{i-1} + q^2 + r^2} = q^2 .$$

It can be easily verified that the solution of this equation will coincide with the solution of the equation of type (7.46), which, in this case, is written as

$$\left(P_\infty^f + q^2\right)\left(\frac{r^2}{P_\infty^\phi + q^2 + r^2}\right) - P_\infty^f = 0$$

or

$$\left(P_\infty^f\right)^2 + P_\infty^f q^2 - q^2 r^2 = 0 .$$

Having solved the equation, we obtain the following for the steady-state variance:

$$P_\infty^f = -\frac{q^2}{2} \pm \sqrt{\frac{q^4}{4} + q^2 r^{`2}} = -\frac{q^2}{2} \pm \frac{q^2}{2}\sqrt{1 + \frac{4r^{`2}}{q^2}} .$$

Note that the considered equation has two solutions. Both of them should be analyzed if they are positive. If one solution is negative, only the positive one is used because the variance naturally cannot be negative. If the variance at initial time is taken to be $\sigma_0^2 = P_\infty$, it will be change starting from the first measurement; if $\sigma_0^2 > P_\infty$, the variance will decrease tending to the steady-state value $\sigma_i^2 \underset{i\to\infty}{\to} P_\infty$. If $\sigma_0^2 < P_\infty$, the variance will increase approaching the same value from below. If the filtering error

variance is a steady-state value, the gain and the prediction error variance will also tend to steady-state values

$$K_\infty = \frac{P_\infty^f}{r^2}; \qquad P_\infty^{ex} = P_\infty^f + q^2.$$

Thus, assuming $q^2 = 1$, $r^2 = 2$, we obtain $P_\infty = 1$ and $K_\infty = \frac{1}{2}$, and the Wiener filter is given by

$$\hat{x}_i = 0.5\hat{x}_{i-1} + 0.5y_i = \frac{\hat{x}_{i-1} + y_i}{2},$$

i.e., the estimate at the current step is an arithmetic mean between the estimate at the previous step and the current measurement.

As shown in problem 7.6, if $q << r$, the following can be written: $P_\infty \approx rq$; $P_\infty^{pr} \approx q(r+q)$; $K_\infty \approx \frac{q}{r}$. If $q >> r$, $P_\infty \approx r^2$, $P_\infty^{pr} \approx q^2$, $K_\infty \approx 1$. ♦

Equation (7.46) (see problem 7.7) can be easily specified for the steady-state mode of the problem of filtering an exponentially correlated sequence by measurements with white-noise errors.

7.6. Statement and general solution of the problem of recursive optimal filtering of random sequences using nonlinear measurements

Consider a more general statement (as compared to the previous section) of recursive optimal Bayesian filtering of random sequences. First, remove the limitation on the linear nature of estimates used in minimization of criterion (7.25), second, assume that the measurements can nonlinearly depend on the parameters being estimated.

In the general case, the problem can be formulated as follows.

We have an n-dimensional random sequence in the form of a shaping filter

$$x_i = \Phi_i x_{i-1} + \Gamma_i w_i \tag{7.47}$$

and m-dimensional measurements

$$y_i = s_i(x_i) + v_i, \tag{7.48}$$

where w_i is a p-dimensional vector of generating noise; v_i is an m-dimensional measurement error vector; Φ_i, Γ_i are the known matrices of dimensionalities $m \times n$, $n \times p$; $s_i(x_i) = (s_{i1}(x_i),...,s_{im}(x_i))^{\mathrm{T}}$ is a known m-dimensional function generally nonlinear with respect to argument x.

Sequences w_i and v_i are discrete zero-mean white noises whose values are independent of each other at different times, and initial conditions vector x_0 is assumed to be a zero-mean vector with covariance matrix P_0. Vectors x_0, w_i, v_i are also considered to be independent of each other.

PDFs for random vector x_0 and sequences w_i and v_i are assumed to be known and set by the corresponding functions $p_{x_0}(x_0)$, $p_{w_i}(w_i)$, and $p_{v_i}(v_i)$.

It is required to derive a recursive algorithm for calculating RMS optimal estimates of sequence (7.1), minimizing the criterion

$$J_i^{\mathrm{B}} = E\left\{\left(x_i - \hat{x}_i(Y_i)\right)^{\mathrm{T}}\left(x_i - \hat{x}_i(Y_i)\right)\right\}, \tag{7.49}$$

and their accuracy characteristics as estimation error covariance matrices (7.26), using measurements $Y_i = (y_1^{\mathrm{T}}, y_2^{\mathrm{T}}, ...y_i^{\mathrm{T}})^{\mathrm{T}}$. In so doing, no restrictions are imposed on the class of the estimates used.

When the measurements are linear and (7.49) is minimized, only for the class of the linear estimates, the problem solution is determined by (7.28)–(7.32). These formulas are derived in 7.3 based on the formulas for linear constant vector estimation problem. In this case, in order to get the sought algorithms, we'll rely upon the results obtained in 4.2, according to which the sought estimate is defined as

$$\hat{x}_i(Y_i) = \int x_i\, p(x_i / Y_i)dx_i\,, \tag{7.50}$$

where $p(x_i / Y_i)$ is a posteriori PDF, or just a posteriori density.

Accuracy characteristics in the form of conditional and unconditional a posteriori error covariance matrices of **optimal estimates of random sequence** are given by

$$P_i\ (Y_i) = \int (x_i - \hat{x}_i(Y_i))(x_i - \hat{x}_i(Y_i))^{\mathrm{T}}\, p(x_i / Y_i)dx_i\,; \tag{7.51}$$

$$P_i\ = \int\int (x_i - \hat{x}_i(Y_i))(x_i - \hat{x}_i(Y_i))^{\mathrm{T}}\, p(x_i, Y_i)dx_i dY_i\,. \tag{7.52}$$

Here, unconditional a posteriori covariance matrix P_i characterizes **the potential accuracy of the solution of the formulated optimal filtering problem** on the average over all measurements, and matrix $P_i\ (Y_i)$ characterizes the calculated accuracy for the current measurement set. It is for this matrix that we are going to obtain the recursive algorithm.

We shall call the estimate (7.50) **an RMS optimal Bayesian estimate or just an optimal estimate of random sequence.** It will possess all the properties of the optimal estimates detailed in 4.3.

Here, similarly to 4.2, it will be reasonable to emphasize that along with setting the rule for optimal estimate calculation (7.50), the rule (7.51) is also set for calculating **the current a posteriori error covariance matrix** $P_i\ (Y_i)$ characterizing the calculated estimation accuracy for a certain measurement set. Thus, identically to the case of the constant vector estimation, by **designing the optimal estimation algorithms for random sequences**, we mean the problem of designing a procedure providing the calculation of estimate (7.50) and its conditional a posteriori covariance matrix (7.51). Call this procedure an **optimal algorithm**. By **analyzing the accuracy of estimating** the **random sequences,** we mean the **problem of the calculation and analysis of unconditional a posteriori covariance matrix** P_i **.**

Outline the major specific features of this statement as compared with the statement given in 7.2.

1. No limitations on the class of the estimates used are imposed in minimization of the selected criterion.
2. Measurements are assumed to depend nonlinearly on the estimated parameters.
3. PDFs for the vector of initial conditions, generating and measurement noise are considered to be known, and not only their first two moments.
4. Instead of uncorrelatedness conditions (7.23), (7.24) of random vectors at different times, conditions of their independence are assumed to be fulfilled.

It follows from (7.50), (7.51) that as in the case of estimating the vector of constant parameters, to solve the formulated problem, we need to know a posteriori density $p(x_i / Y_i)$, finding which is the main contents and the main difficulty of the nonlinear filtering problem.

The recursive optimal filtering problem is greatly simplified if a posteriori density is Gaussian. Clearly, a posteriori density will be Gaussian if the measurements linearly depend on the estimated sequence, i.e., $s_i(x_i) = H_i x_i$, and the vectors of initial conditions, generating and measurement noise are Gaussian, i.e.:

$$p(x_0) = N(x_0; 0, P_0);\ p(w_i) = N(w_i; 0, Q_i);\ p(v_i) = N(v_i; 0, R_i).$$

Thus, it is obvious that if we additionally assume the Gaussian character of initial conditions, generating and measurement noise in the problem considered in 7.2, a posteriori density will also be Gaussian. Moreover, it can be shown that this problem can be solved using the Kalman filter equations (7.28)–(7.32).

It follows that the Kalman filter equations given in section 7.3 in the Gaussian case allow obtaining RMS optimal Bayesian estimates of random sequence (7.21) by measurements (7.22), i.e., the estimates minimizing the RMS criterion without any limitations on the class of the estimates used.

Therefore, for the considered particular case, the Kalman filter estimates possess all the properties of optimal estimates (see section 4.3).

In particular, conditional and unconditional error covariance matrices of optimal estimates will coincide, i.e., $P_i = P_i(Y_i)$, and if we select an arbitrary, not necessarily linear, algorithm for finding estimates $\tilde{x}(Y_i)$ and find their unconditional error covariance matrix $\tilde{P}_i$, according to features 3, 4 given in 4.3, the following inequalities will always be true:

$$\tilde{P}_i - P_i \geq 0,$$
$$\det(\tilde{P}_i) \geq \det(P_i).$$

Property 5 is also important, implying that obtaining the optimal estimate of the state vector as a whole ensures finding of the optimal estimate of the vector being an arbitrary linear transformation of the vector being estimated.

Therefore, estimate error covariance matrix (7.32) calculated in the Kalman filter

characterizes the **potential accuracy of estimating a random Gaussian sequence formed by Eqs. (7.21).**

Exercises

Exercise 7.1. Assuming that in the filtering problem considered in section 7.2, the error covariance matrix at the previous step $P_{i-1} = E\left\{\varepsilon_{i-1}\varepsilon_{i-1}^{\mathrm{T}}\right\}$ is known, show that filtering equations (7.29), (7.34) are true for prediction error covariance matrices and estimate errors at the current step calculated as $\hat{x}_{i/i-1} = \Phi_i \hat{x}_{i-1}$; $\hat{x}_i = \hat{x}_{i/i-1} + K_i(y_i - H_i\hat{x}_{i/i-1})$, where matrix K_i is given by (7.31).

Exercise 7.2. Show that the covariance matrix for residual $\mu_i = y_i - H_i\hat{x}_{i/i-1}(Y_{i-1})$ is given by $L_i = H_i P_{i/i-1} H_i^{\mathrm{T}} + R_i$.

Exercise 7.3. Prove that $E\left\{\varepsilon_i(Y_i)x_i^{\mathrm{T}}\right\} = P_i$.

Exercise 7.4. It is required to get the optimal estimate of the Wiener sequence set given by $x_i = x_{i-1} + w_i$ by using scalar measurements $y_i = x + v_i$, where w_i, v_i are zero-mean discrete white noises independent of x_0 $E\{v_i v_j\} = \delta_{ij} r_i^2$, $E\{v_i v_j\} = \delta_{ij} q_i^2$; x_0 is a zero-mean random value with variance σ_0^2. Then, $E\{v_i w_j\} = \delta_{ij} b_i^2$, i.e., the noises depend on each other.

Specify the formulas of the discrete Kalman filter.

Exercise 7.5. Write the error equation for the steady-state mode of the problem from example 7.6.

Exercise 7.6. Find the value of the variance of the filtering error, the prediction error and the gain in steady-state mode in the Wiener sequence estimation problem from example 7.6 for the case when $q << r$ and $q >> r$.

Exercise 7.7. Specify Equation (7.46) for the steady-state value of the error variance of filtering the scalar exponentially correlated sequence by using the measurements against the white noise background. Find its approximate solution and equations for the variance of the prediction error, the gain, and the Wiener filter if $q >> r$. Relate the equation for the steady-state value of the filtering error variance with the similar equation in example 7.6.

Test questions

1. Formulate the problem statement of obtaining RMS optimal linear estimates of a random sequence using the measurements of another sequence correlated with it. Write down its solution using the discrete Wiener-Hopf equation.
   Explain the specific features of filtering, smoothing, and prediction problems.
2. Formulate the linear problem of recursive estimation of a random sequence described using a shaping filter.

3. Which main blocks does the discrete Kalman filter algorithm include? Explain why this algorithm is a recursive one. Specify these blocks by the example of a constant scalar estimation problem.
4. Explain what is meant by prediction and filtering errors. Write the equation for these errors. Why do filtering errors form a Markovian sequence?
5. What are the conditions for the existence of steady-state mode in the filtering problem?
6. What is the Wiener filter? Explain how it is related to the Kalman filter. Provide an example.
7. Provide the statement and the general solution to the problem of recursive optimal Bayesian filtering of random sequences.
8. Explain the difference between the problems of recursive optimal Bayesian filtering and recursive optimal linear filtering of random sequences.
9. Under which additional conditions, as compared with the conditions in 7.2, will the Kalman filter be optimal without imposing limitations for the class of the estimates used?

8. Filtering of random processes. Kalman-Bucy filter

In practice, applied filtering problems are usually formulated in continuous form using differential equations. In this case, we have to deal with random processes. Consider this problem within the approach proposed by R. Kalman and R. Bucy.

8.1. Random processes and methods of their description

A random process $\mathbf{x}(t)$ in a scalar case is a function of time t, whose value with any fixed t is a random variable. The main characteristics and classes of random processes are introduced similarly to random sequences. In particular, the correlation function for a random process is given by

$$k_x(t_1,t_2)=\int\int(x(t_1)-\overline{x}(t_1))(x(t_2)-\overline{x}(t_2))^{\mathrm{T}}\,p(x_1,t_1,x_2,t_2)dx_1dx_2\,, \qquad (8.1)$$

where $p(x_1,t_1,x_2,t_2)$ is a joint PDF for random variables $x(t_1)$ and $x(t_2)$.

Hereinafter, both random processes $\mathbf{x}(t)$ and their samples $x(t)$ are denoted by $x(t)$, similarly with the random sequences [16].

For stationary processes, as follows from section 6.2, the correlation function depends on the difference of arguments.

♦ **Example 8.1.** As an example of the correlation function of a stationary (time-invariant) process, consider function

$$k_x(\tau)=\sigma_x^2e^{-\alpha|\tau|}. \qquad (8.2)$$

The process with such a correlation function is called an **exponentially correlated** process and is a continuous analog of the exponentially correlated sequence given in example 6.6 [25]. Here, $k_x(0)=\sigma_x^2$ is the process variance, and $\tau_{\mathrm{c}}=1/\alpha$ is the **correlation interval**.

Plots of function (8.2) with a unit variance and two values of correlation intervals $\tau_c=1\mathrm{s}$ and $\tau_c=0.1\mathrm{s}$ are shown in Fig. 8.1.

To describe the properties of stationary processes, along with the correlation function, the **power spectral density (PSD)** is used, which is a Fourier transform of the correlation function:

$$S_x(\omega)=\int_{-\infty}^{\infty}k_x(\tau)\exp(-j\omega\tau)d\tau\,. \qquad (8.3)$$

Inverse representation is also true:

$$k_x(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_x(\omega)\exp(j\omega\tau)d\omega\,. \qquad (8.4)$$

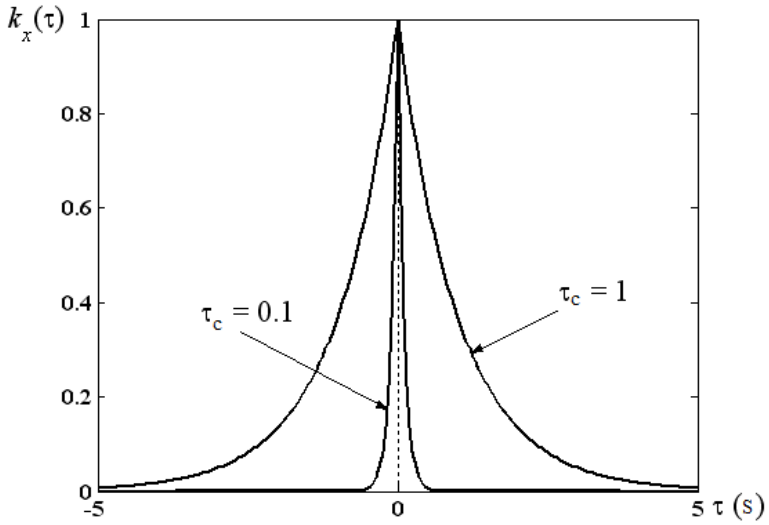


Fig. 8.1. Correlation function $k_x(\tau) = \sigma_x^2 e^{-\alpha|\tau|}$ with two correlation intervals ♦

Due to even nature of functions $S_x(\omega)$, $k_x(\tau)$, the formulas can be written as

$$S_x(\omega) = \int_{-\infty}^{\infty} k_x(\tau)\cos\omega\tau d\tau = 2\int_0^{\infty} k_x(\tau)\cos\omega\tau d\tau;$$

$$k_x(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)\cos\omega\tau d\omega = \frac{1}{\pi}\int_0^{\infty} S_x(\omega)\cos\omega\tau d\omega.$$

Obviously, from the last formula, the following is derived:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)d\omega = k_x(0) = \sigma_x^2, \tag{8.5}$$

meaning that the area enclosed by $S_x(\omega)$ and the abscissa axis determines the process variance σ_x^2 accurate to a constant coefficient.

♦ **Example 8.2.** Obtain a formula for the PSD of exponentially correlated process with correlation function (8.2) and analyze its behavior. The spectral density of this function, as follows from Table A3.1, is given by

$$S_x(\omega) = \frac{2\sigma_x^2\alpha}{\alpha^2 + \omega^2}. \tag{8.6}$$

PSD plots for two correlation intervals $\tau_c = 1/\alpha$ are presented in Fig. 8.2.

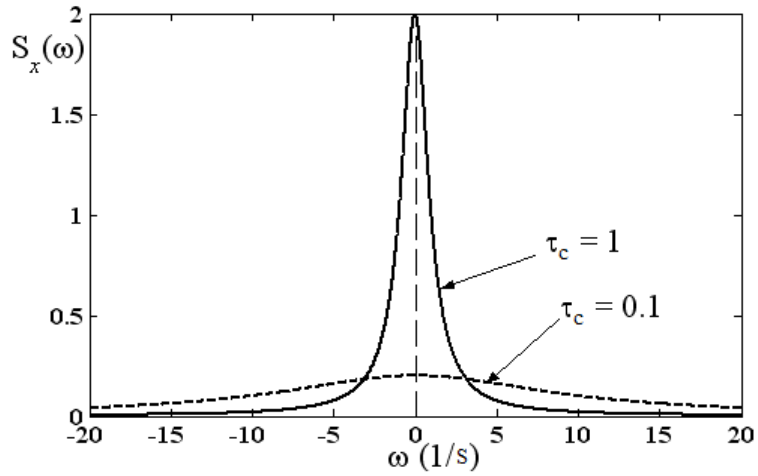


Fig. 8.2. PSD of the exponentially correlated process with various correlation intervals

The frequency area with nonzero PSD can be estimated by magnitude α, while, as ω varies from 0 to α, the PSD only decreases twofold from $S_x(0) = \frac{2\sigma_x^2}{\alpha}$ to $S_x(\alpha) = \frac{\sigma_x^2}{\alpha}$, critically going down with $\omega > \alpha$. This can be distinctly seen from the PSD plot for positive frequencies in log scale. The plots with $\sigma = 1$ and two values α = $1 \cdot s^{-1}$ and α = $10 \cdot s^{-1}$ are presented in Fig. 8.3.

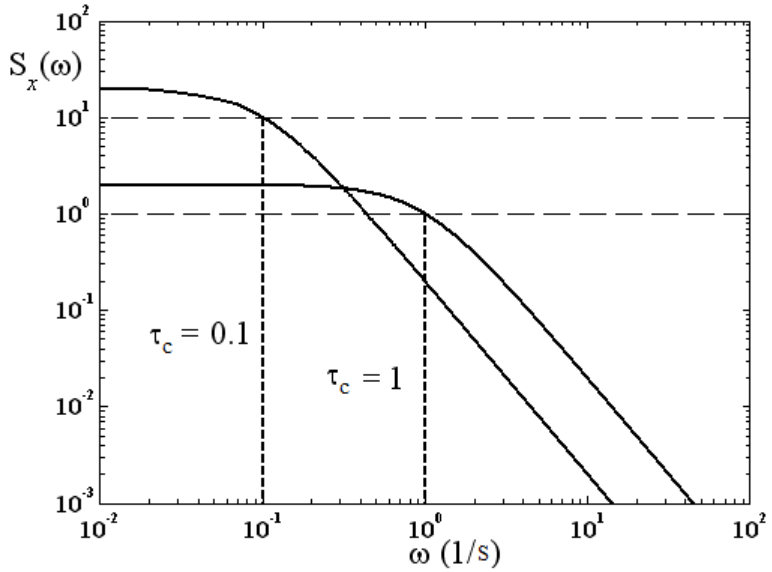


Fig. 8.3. PSD of the exponentially correlated process with various correlation intervals in log scale ♦

The process with constant PSD for all ω, i.e., $S_x(\omega) = Q$, is called **white noise**, and Q is the **PSD of the white noise**. The correlation function for white noise has the form

$$k_x(\tau) = Q\delta(\tau), \qquad (8.7)$$

which follows from the following representation for delta function $\delta(\tau)$ [21]:

$$\frac{1}{2\pi}\int e^{j\omega\tau} d\omega = \delta(\tau).$$

It follows from (8.7) that continuous white noise, unlike discrete noise, has infinite variance. Variables $x(t_1)$ and $x(t_2)$ at any different time points $t_1 \neq t_2$ are noncorrelated, just as for a random sequence.

Note that the dimensionality of white-noise PSD is the same as that of its variance multiplied by the time unit, i.e.,

$$[Q] = [\text{units}^2][\text{time}].$$

The above-said agrees with the formula $k(\tau) = Q\delta(\tau)$: there is a parameter measured in [units^2] in the left-hand side, and a parameter measured in $[Q]$[time]$^{-1}$ in the right-hand side. It follows from the fact that the delta function with the time argument has a dimensionality inverse of time $[\delta(t)] = [time]^{-1}$ since, according to the nature of the delta function, $\int_{-\infty}^{\infty} \delta(\tau) d\tau = 1$.

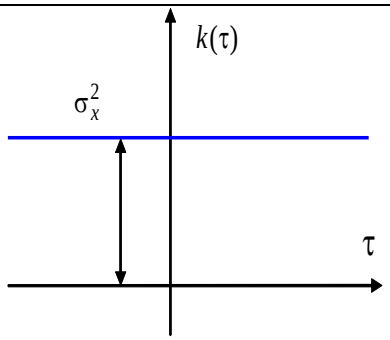
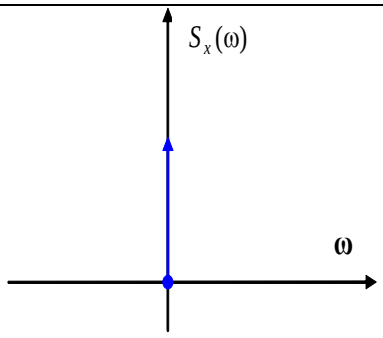
It can be noted that white noise has properties inverse, in some sense, to those of a process presenting a time-invariant random variable (random bias, or random constant) with constant correlation function and PSD being a delta function:

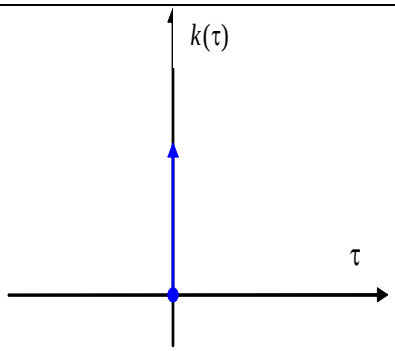
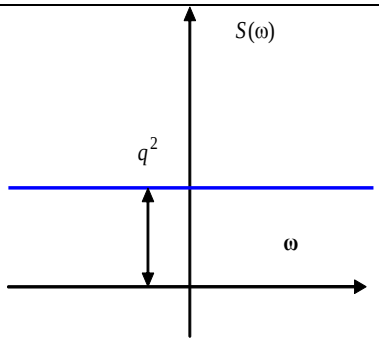
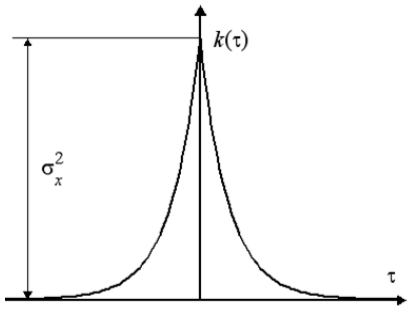
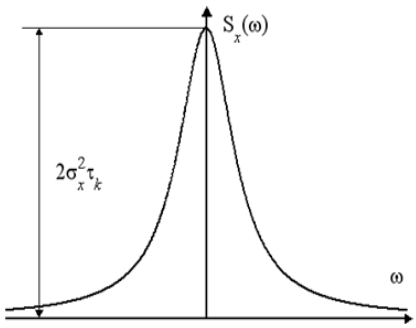
$$S_x(\omega) = 2\pi\sigma_x^2\delta(\omega).$$

Sample plots of correlation functions and corresponding PSDs of the random constant, white noise and the exponentially correlated process are given in Table 8.1.

Table 8.1

Correlation functions and PSDs of simple processes

Process	Correlation function	PSD
Random constant with variance σ_x^2	 $k(\tau) = \sigma_x^2$	 $S_x(\omega) = 2\pi\sigma_x^2\delta(\omega)$

White noise with PSD q^2	 $k_x(\tau)=\sigma_x^2\delta(\tau)$	 $S_x(\omega)=q^2$
Exponentially correlated process	 $k_x(\tau)=\sigma_x^2 e^{-\alpha\lvert\tau\rvert}$	 $S_x(\omega)=\dfrac{2\sigma_x^2\alpha}{\alpha^2+\omega^2}$

8.2. Shaping filter

Consider an *n*-dimensional random process $x(t)$

$$\dot{x}(t)=F(t)x(t)+G(t)w(t), \tag{8.8}$$

where $F(t)$, $G(t)$, $Q(t)$ are the known $n\times n$-, $n\times p$- and $p\times p$-matrices. Suppose that at initial time point, $x(0)$ is a random vector with the known mathematical expectation $\bar{x}(0)$ and covariance matrix $P(0)$, $w(t)$ is a zero-mean *p*-dimensional white noise independent of $x(0)$:

$$E(w(t+\tau)w^T(t))=Q(t)\delta(t-\tau).$$

As with the sequences, Equation (8.8) used to describe the processes is called a **shaping filter**; white noise in the right-hand part is called **generating (forcing)** white noise or the **system noise**; matrix $F(t)$ is a **dynamics matrix**; matrix $G(t)$ is a **generating noise matrix,** and vector $x(t)$ is a **state vector.**

Use formula (A3.8) from Appendix 3 and write the solutions to Equation (8.8) in the form

$$x(t)=\Phi(t,t_1)x(t_1)+\int_{t_1}^{t}\Phi(t,\tau)G(\tau)w(\tau)\mathrm{d}\tau, \tag{8.9}$$

where $\Phi(t,t_1)$ is a fundamental matrix for the equation $\dot{x}(t)=F(t)x(t)$.

It can be shown that the mathematical expectation, covariance matrix and correlation function are given by (exercise 8.1):

$$\bar{x}(t) = \Phi(t, t_0)\bar{x}(t_0); \tag{8.10}$$

$$P(t) = \Phi(t, t_0)P(t_0)\Phi^T(t, t_0) + \int_{t_0}^{t} \Phi(t, \tau)G(\tau)Q(\tau)G^T(\tau)\Phi^T(t, \tau)d\tau; \tag{8.11}$$

$$k(t_2, t_1) = \begin{cases} \Phi(t_2, t_1)P(t_1), & t_2 > t_1, \\ P(t_2)\Phi^T(t_1, t_2), & t_2 \le t_1. \end{cases} \tag{8.12}$$

Taking (A3.19), (A3.20) into consideration, it can be easily verified that covariance matrix (8.11) is a solution to the differential equation

$$\dot{P}(t) = F(t)P(t) + P(t)F^T(t) + G(t)Q(t)G^T(t). \tag{8.13}$$

Stationary equations with constant matrices F, Q and G

$$\dot{x}(t) = Fx(t) + Gw(t); \tag{8.14}$$

$$\dot{P} = FP + PF^T + GQG^T \tag{8.15}$$

are a particular case of Equations (8.8), (8.13).

Discuss the conditions under which the output of stationary system (8.14) is a stationary process. Obviously, that first of all, mathematical expectation $x(t)$ at the initial time should be zero, i.e., $\bar{x}(t_0) = 0$, otherwise, the mathematical expectation of process $x(t)$, as follows from (8.10), will depend on time. To make the covariance matrix (8.13) independent of time, the conditions should be met, wherein there is a steady-state solution to this equation. It means that there exists matrix P_∞, such that $P = P_\infty$, and

$$\dot{P}_\infty = FP_\infty + P_\infty F^T + GQG^T = 0. \tag{8.16}$$

If covariance matrix $P(0)$ for vector $x(0)$ is selected to coincide with the solution of this equation $P(0) = P_\infty$, process $x(t)$ generated by Equation (8.14) becomes a stationary process since $P(t) \equiv P(0)$, moreover, as follows from (8.12), the correlation function

$$k(\tau) = \Phi(\tau)P_\infty, k(-\tau) = P_\infty\Phi^T(\tau) \tag{8.17}$$

will depend only on τ. It also follows from (8.17) that

$$k(\tau) = k^T(-\tau) = \Phi(\tau)P_\infty. \tag{8.18}$$

It should be noted that for a stationary system, the fundamental matrix is defined as a matrix exponential, i.e.,

$$\Phi(\tau) = e^{F\tau} = \sum_{\nu=0}^{\infty} F^\nu \tau^\nu / \nu!.$$

Thus, the conditions for occurrence of **a stationary process at the output of a**

stationary system to which white noise is inputted are as follows: zero-mean of the process at the initial time point, existence of solution to Equation (8.16), and selection of the initial covariance matrix coinciding with this solution.

If the steady-state solution to Equation (8.16) exists, but the initial covariance matrix does not coincide with P_∞, then, since $P(t) \to P_\infty$ with time, the process can be considered stationary after the transient mode at $t \to \infty$ is completed.

N o t e 1. If we additionally assume that $x(0)$ and the generating noise are Gaussian, i.e.,

$$f(x(0)) = N(x(0); \bar{x}(0), P(0)); \tag{8.19}$$

$$f(w(t)) = N(w(t); 0, Q(t)), \tag{8.20}$$

process $x(t)$ will also be Gaussian. It is explained by the fact that linear transformations of Gaussian vectors provide a Gaussian vector.

N o t e 2. Using (8.9), it can be easily verified that process $x(t)$ is a Markov process. Actually, if we fix the time points $t_1 > t_2 > t_3$, then, as follows from (8.9), the process at t_3 with fixed values at t_1 and t_2 depends only on t_2 and does not depend on t_1. Then, it is significant that $w(t)$ is white noise, which, in the statistical sense, is independent of initial conditions $x(0)$.

♦ **Example 8.3.** Consider a shaping filter

$$\dot{x}(t) = -\alpha x(t) + \sqrt{2\sigma_x^2 \alpha}\, w(t). \tag{8.21}$$

Obtain the correlation function for this process and specify Equation (8.15).

Since $F = -\alpha$, $q_w = \sqrt{2\sigma_x^2 \alpha}$, Equation (8.14) takes the form

$$\dot{P} = -\alpha P - \alpha P + q_w^2.$$

While $\Phi(t, t_0) = e^{-\alpha(t-t_0)}$, solution to this equation can be represented as

$$P(t) = P(t_0) e^{-2\alpha(t-t_0)} + 2\sigma_x^2 \alpha \int_{t_0}^{t} e^{-2\alpha(t-\tau)} d\tau.$$

Clearly, (8.16) is reduced to Eq. $2\alpha P_\infty = 2\sigma_x^2 \alpha$ which has the following solution: $P_\infty = \sigma_x^2$. Thus, with $P(0) = \sigma_x^2$, the process will be stationary, and its correlation function will take the form

$$k(\tau) = k(-\tau) = \Phi(\tau) P = \sigma_x^2 e^{-\alpha\tau}. \quad ♦$$

8.3. Statement and general solution of the optimal linear filtering problem. Kalman-Bucy filter

Provide the mathematical formulation and general solution to the optimal linear filtering problem considered within the state-space approach. Recall that this solution to the filtering problem was derived by R. Kalman for discrete time, and then jointly with

R. Bucy, for continuous time as well [1, 18, 27]. This gave the filter its name – the Kalman-Bucy filter.

Formulate the mathematical statement of the continuous optimal linear filtering problem similarly to that given for the the discrete case [18, 19, 25, 28].

Let there be an *n*-dimensional Markov process

$$\dot{x}(t) = F(t)x(t) + G(t)w(t),\ x(t_0) = x_0 \tag{8.22}$$

and *m*-dimensional measurements

$$y(t) = H(t)x(t) + v(t), \tag{8.23}$$

where $F(t)$, $G(t)$, $H(t)$ are generally known $n \times n$, $n \times p$, $m \times n$ time-dependent matrices; x_0 is zero-mean initial conditions vector with covariance matrix P_0; $w(t), v(t)$ are zero-mean white noises independent of each other and of initial conditions x_0:

$$E\left\{x_0 w^T(t)\right\} = 0;\ E\left\{w(t)v^T(t)\right\} = 0;\ E\left\{x_0 v^T(t)\right\} = 0; \tag{8.24}$$

$$E\{w(t)w^T(\tau)\} = Q(t)\delta(t-\tau),\ Q(t) \geq 0; \tag{8.25}$$

$$E\{v(t)v^T(\tau)\} = R(t)\delta(t-\tau),\ R(t) > 0. \tag{8.26}$$

The filtering problem in state space **for continuous time** is formulated as follows. Using measurements (8.23) $Y(t) = \left\{y(\tau) : \tau \in [0,t]\right\}$ accumulated over the interval $[0,t]$ by time t, obtain a linear RMS optimal estimate of vector $x(t)$, which minimizes the criterion

$$r^b(t) = E\left\{(x(t) - \hat{x}(t))^T (x(t) - \hat{x}(t))\right\}. \tag{8.27}$$

It can be shown that the estimate and its error covariance matrix are given by [18, 19, 25]:

$$\dot{\hat{x}}(t) = F(t)\hat{x}(t) + K(t)(y(t) - H(t)\hat{x}(t)); \tag{8.28}$$

$$K(t) = P(t)H(t)^T R^{-1}(t); \tag{8.29}$$

$$\dot{P}(t) = P(t)F(t)^T + F(t)P(t) - P(t)H(t)^T R^{-1}(t)H(t)P(t) + G(t)Q(t)G^T(t). \tag{8.30}$$

These equations define the well-known **continuous Kalman-Bucy filter**.

It is important to notice that the measurement residual $y(t) - H(t)\hat{x}(t)$ called an innovation, as in discrete case, is a white noise with the PSD matrix equal to *R*.

Using (8.8), (8.28), the filtering error equation can be easily obtained (exercise 8.3).

Covariance matrix (8.30) is called a posteriori covariance matrix.

It follows from (8.28) that the estimation formula can be written as

$$\dot{\hat{x}}(t) = (F(t) - K(t)H(t))\hat{x}(t) + K(t)y(t), \tag{8.31}$$

i.e., the Kalman filter is a linear generally nonstationary dynamic system, whose

properties are defined by matrix $(F(t)-K(t)H(t))$, and to which the measurements, weighted with the use of matrix $K(t)$, are inputted.

The block diagram of generation of the process being estimated and the measurements is presented in Fig. 8.4, and the block diagram of the Kalman-Bucy filter, in Fig. 8.5.

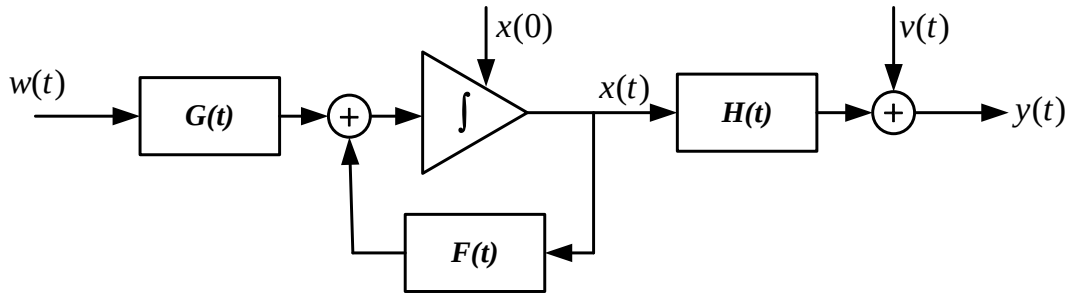


Fig. 8.4. Block diagram of generation of the process being estimated and its measurements

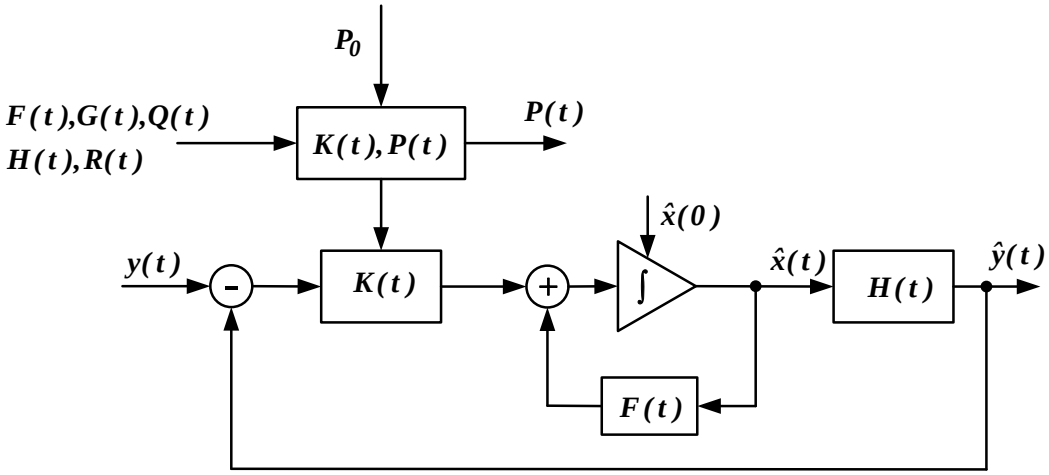


Fig. 8.5. Block diagram of the Kalman-Bucy filter

Comparison of these diagrams shows that the Kalman-Bucy filter contains a block similar to the shaping filter with added single negative feedback, and instead of generating white noise, measurement residuals $\mu(t)=y(t)-H(t)\hat{x}(t)$ weighted with the use of matrix $K(t)$ are inputted.

As in the discrete case, in this algorithm, two blocks (channels) can be distinguished: an **estimation block**, corresponding to the linear differential equation for optimal estimates (8.28), and a **covariance block** corresponding to **nonlinear Riccati differential equation** (8.30). The latter, like the gain matrix $K(t)$, does not depend on measurements, rather, it depends only on matrices $F(t),G(t),Q(t)$ and $H(t),R(t)$ determining the properties of the processes being estimated and measured. Both equations should be solved to obtain the estimates; besides, the covariance matrix

equation can be solved in advance since it is independent from the measurements.

The above given formulas and the conditions under which the problem is solved are summarized in Table 8.2 [8, 13, 19, 20].

Table 8.2

Statement of the filtering problem and the continuous Kalman-Bucy algorithm

Initial data	
State vector equation	$\dot{x} = F(t)x + G(t)w(t)$
Measurements	$y = H(t)x + v(t)$
Initial conditions	x_0 is a zero-mean random vector with covariance matrix P_0
Generating noise	$M\{w(t)w^{\mathrm{T}}(\tau)\} = Q(t)\delta(t-\tau)$, $Q(t) \geq 0$
Measurement noise	$M\{v(t)v^{\mathrm{T}}(\tau)\} = R(t)\delta(t-\tau)$, $R(t) > 0$
Cross correlation	$M\left\{x_0 w^{\mathrm{T}}(t)\right\} = 0$; $M\left\{w(t)v^{\mathrm{T}}(t)\right\} = 0$ $M\left\{x_0 v^{\mathrm{T}}(t)\right\} = 0$
Matrices	$F(t) - n \times n$, $G(t) - n \times p$ $H(t) - m \times n$
Statement of the filtering problem and minimized criterion Using measurements $Y(t) = \{y(\tau) : \tau \in [0,t]\}$, obtain estimate $\hat{x}(t)$ which minimizes the criterion $r^{\mathrm{B}}(t) = M\left\{(x(t) - \hat{x}(t))^{\mathrm{T}}(x(t) - \hat{x}(t))\right\}$ and error covariance matrix $P(t) = M\left\{(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^{\mathrm{T}}\right\}$, $r^{\mathrm{B}}(t) = SpP(t)$	
Solution of the filtering problem	
Initial conditions	$\hat{x}(0) = 0$, $P(0) = P_0$
Estimate	$\dot{\hat{x}}(t) = F(t)\hat{x}(t) + K(t)(y(t) - H(t)\hat{x}(t))$
Gain factor	$K(t) = P(t)H^{\mathrm{T}}(t)R^{-1}(t)$
Error covariance matrix	$\dot{P}(t) = P(t)F^{T}(t) + F(t)P(t) -$ $-P(t)H^{T}(t)R^{-1}(t)H(t)P(t) + G(t)Q(t)G^{T}(t)$

Notably, the PSD matrix of white noise measurement errors should be non-singular, i.e., $R(t) > 0$, because we need an inverse matrix of $R(t)$ to calculate the covariance matrix.

♦ **Example 8.4.** Solve the problem of filtering of a constant scalar

$$\dot{x} = 0$$

by using the measurements with white noise errors, i.e.,

$$y(t) = x + v(t),$$

where x is a zero-mean random value with variance σ_x^2, $v(t)$ is a zero-mean white noise noncorrelated with x with the correlation function

$$M\{v(t)v^T(\tau)\} = r^2\delta(t-\tau).$$

Here, $F = 0,\ G = 0,\ H(t) = 1,\ P_0 = \sigma_0^2$, and Equations (8.28)–(8.30) are reduced to

$$\dot{\hat{x}}(t) = K(t)(y(t) - \hat{x}(t));$$

$$K(t) = P(t)/r^2;\ \dot{P}(t) = P^2(t)/r^2.$$

Direct substitution yields

$$P(t) = \frac{\sigma_x^2 r^2}{r^2 + \sigma_x^2 t} \text{ and } K(t) = \frac{\sigma_x^2}{r^2 + \sigma_x^2 t}.$$

If the inequality $\frac{r^2}{t} << \sigma_x^2$ is true, meaning a low effect of a priori information, approximate formulas $P(t) = \frac{r^2}{t}, K(t) = \frac{1}{t}$ can be written, from which we obtain the estimation equation

$$\dot{\hat{x}}(t) = \frac{1}{t}(y(t) - \hat{x}(t)).$$

Since the fundamental matrix in this case is given by

$$\Phi(t, t_0) = \frac{t_0}{t},$$

the general solution to the linear nonstationary estimation equation for this problem can be written as

$$\hat{x}(t) = \frac{t_0}{t}x_0 + \int_0^t \frac{\tau}{t} y(\tau)\frac{1}{\tau}d\tau \to \frac{1}{t}\int_0^t y(\tau)d\tau.$$

It follows herefrom that with the assumptions made in this example, as in a similar discrete problem, the algorithm for obtaining the optimal estimate is reduced to the calculation of the mean value of all incoming measurements. ♦

Note 1. Along with the estimates**,** the Kalman-Bucy filter generates the covariance matrix $P(t)$, whose diagonal elements characterize the current estimation accuracy, which is extremely important in navigation data processing problems.

Note 2. While the general solution of nonstationary differential equation (8.31) is determined by (A3.8), the Kalman-Bucy filter estimates linearly depend on the measurements, and thus, are RMS optimal linear estimates.

Note that in this problem statement, as in the previous section, the form of PDF for

generating noise, white noise measurement errors, and initial conditions is not specified. If they are additionally assumed Gaussian, the estimates obtained by the above formulas will be optimal, in the RMS sense, without adding the term *linear*; in other words, they will minimize criterion (8.27) without an assumption on the linearity of the estimation algorithm. In this case, we speak about the Bayesian optimal filtering problem.

Note 3. If the filtering problem is solved for a stationary system with constant matrices F, G, Q, R, Equation (8.22) may have a steady-state solution, i.e.,

$$P_\infty F^T + FP_\infty - P_\infty H^T R^{-1} HP_\infty + GQG^T = 0.$$

In this case, while $K = K_\infty = P_\infty H^T R^{-1}$, the estimation equation also becomes stationary, and the Kalman filter will coincide with the Wiener filter.

Note 4. In some cases, to solve Equation (8.30) without generating noise, i.e., with

$$\dot{P}(t) = P(t)F^T(t) + F(t)P(t) - P(t)H^T(t)R(t)^{-1}H(t)P(t),$$

it is convenient to proceed to the equation for the inverse covariance matrix using formula $P(t)P^{-1}(t) = E$. It can be easily seen that in this case

$$\frac{dP^{-1}(t)}{dt} = -F^T(t)P^{-1}(t) - P^{-1}(t)F(t) + H^T(t)R(t)^{-1}H(t),$$

or, with account for (A4.20),

$$P^{-1}(t) = \Phi^T(t_0,t)P_0^{-1}\Phi(t_0,t) + \int_{t_0}^{t} \Phi^T(\tau,t_0)H^T R^{-1} H\Phi\ (\tau,t_0)d\tau.$$

8.4. Interrelation of continuous and discrete filtering algorithms

As was mentioned at the beginning of this chapter, applied filtering problems are formulated in continuous form. However, for computers, we need a discrete statement. To obtain this statement, first, we should design an n-dimensional random sequence (for sampling interval Δt)

$$x_i = \Phi_i x_{i-1} + \Gamma_i w_i \tag{8.32}$$

with discrete p-dimensional zero-mean white noise w_i, with the mathematical expectation and the covariance matrix at discrete times, coinciding with similar parameters for the continuous process (8.8), i.e.,

$$m_i \equiv Mx_i = Mx(t_i) \equiv \bar{x}(t_i), \tag{8.33}$$

$$P_i \equiv M\left\{(x_i - m_i)(x_i - m_i)^T\right\} = M\left\{\left(x(t_i) - \bar{x}(t_i)\right)\left(x(t_i) - \bar{x}(t_i)\right)^T\right\}. \tag{8.34}$$

These equalities are called **conditions of stochastic (statistic) equivalence,** meaning that statistical characteristics (8.33), (8.34) of a continuous random process and its discrete analog are the same for the same time points [8].

Hence, it follows that we need to find such matrices Φ_i, Γ_i and the covariance matrix of discrete white noise Q_i which will guarantee that conditions (8.33), (8.34) are met. The resultant sequence is called a **sequence stochastically equivalent** to process (8.8).

Taking (8.10) into consideration, note that if matrix Φ_i in (8.32) is selected in the form

$$\Phi_i = \Phi(t_{i-1} + \Delta t, t_{i-1}), \tag{8.35}$$

condition (8.32) will be met.

Using (8.11), write

$$\begin{aligned} P(t_i) &= \Phi(t_i, t_{i-1}) P(t_{i-1}) \Phi^T(t_i, t_{i-1}) + \\ &+ \int_{t_{i-1}}^{t_i} \Phi(t_i, \tau) G(t_{i-1} + \tau) Q(t_{i-1} + \tau) G^T(t_{i-1} + \tau) \Phi^T(t_i, \tau) d\tau. \end{aligned} \tag{8.36}$$

With account for (6.16), it can easily be seen that in calculating matrix Φ_i using (8.35), to make the covariance matrices of continuous process and its discrete analog equal, i.e., $P_i = P(t_i)$, matrices Γ_i and Q_i should be selected such that

$$\Gamma_i Q_i \Gamma_i^T = Q_{di}^*, \tag{8.37}$$

where

$$Q_{di}^* = \int_{t_{i-1}}^{t_{i-1}+\Delta t} \Phi(t_{i-1} + \Delta t, \tau) G(\tau) Q(\tau) G^T(\tau) \Phi^T(t_{i-1} + \Delta t, \tau) d\tau . \tag{8.38}$$

The dimensionality of matrix Q_i corresponds to that of vector w_i and is equal to p, so the dimensionality of matrix Γ_i must be $n \times p$.

Thus, to get a discrete representation of the process described by Equation (8.8), formulas (8.35), (8.37), (8.38) can be used.

Assume that the continuous system is stationary, i.e., matrices F, G, Q are time-invariant. In this case, matrix Φ_i is determined as a matrix exponent

$$\Phi_i = \Phi_i(\Delta t) = e^{F\Delta t} \approx \sum_{\nu=0}^{k} F^\nu \Delta t^\nu / \nu!, \tag{8. 39}$$

and (8.37), (8.38) take the form

$$\Gamma Q_d \Gamma^T = Q_d^*, \tag{8.40}$$

where

$$Q_d^* = \int_0^{\Delta t} e^{F\tau} G Q G^T (e^{F\tau})^T d\tau . \tag{8.41}$$

Using some $\tau^* \in [0, \Delta t]$ and the mean value theorem, we can write for (8.41)

$$Q_d^* = e^{F\tau^*} G Q G^T \left(e^{F\tau^*}\right)^T \Delta t . \tag{8.42}$$

If we assume that $e^{F\tau^*} \approx E$ because of smallness of Δt,

$$Q_d^* \approx GQG^T \Delta t. \tag{8.43}$$

Using (8.37), (8.38), write

$$\Gamma Q_d \Gamma^T \approx GQG^T \Delta t. \tag{8.44}$$

If

$$\Gamma = G\Delta t, \tag{8.45}$$

we obtain the following for the covariance matrix of discrete white noise:

$$Q_d = \frac{1}{\Delta t} Q. \tag{8.46}$$

If we introduce white noise as a mean value of continuous white noise over the sampling interval, i.e.,

$$w_i = \frac{1}{\Delta t} \int_0^{\Delta t} w(t) dt, \tag{8.47}$$

its covariance matrix will be given by $M(w_i w_i^T) = \frac{1}{\Delta t} Q$ (see exercise 8.2), i.e., if the matrices are selected using (8.45), (8.46), the covariance matrix of discrete noise coincides with that of the continuous noise averaged over the sampling interval Δt.

It should be borne in mind that the given formulas are approximate. This especially refers to matrices Q_i and Γ_i. To obtain more accurate formulas, matrix Q_d^* should be calculated according to (8.38) or (8.41).

To complete the formulation of the discrete form of the filtering problem, we need to find matrices H_i and R_i, which can be used to write down the measurements. To determine these matrices, we assume that the measurement set over the interval $t \in [t_{i-1}, t_{i-1} + \Delta t]$ is replaced by one averaged measurement, i.e.,

$$y_i = \frac{1}{\Delta t} \int_{t_i + \Delta t}^{t_i} (H(t)x(t) + v(t)) dt. \tag{8.48}$$

In this case, matrices H_i and R_i are given by

$$H(t) \approx H(t_i) = H_i; \tag{8.49}$$

$$R(t) \approx R(t_i), \quad t \in [t_{i-1}, t_{i-1} + \Delta t]; \tag{8.50}$$

$$R_i \approx \frac{1}{\Delta t} R(t_i). \tag{8.51}$$

The last formula follows from the fact that

$$v_i = \frac{1}{\Delta t} \int_{t_i - \Delta t}^{t_i} v(t) dt.$$

Clearly, to receive a discrete analog of a continuous problem, a correct sampling interval Δt should be chosen. Firstly, this interval should be selected based on the

condition that matrices $F(t), H(t)$ and the matrices in (8.38) change insignificantly over the sampling interval $t \in [t_{i-1}, t_{i-1} + \Delta t]$. In addition, it is necessary to take into consideration possible variability of the process conditioned by internal dynamics defined by matrix $F(t) \approx F(t_i) = F_i$ assumed to be constant for $t \in [t_{i-1}, t_{i-1} + \Delta t]$. This variability is known to depend on natural frequencies, which, in turn, depend on the matrix eigenvalues. That is why when selecting the sampling interval, it is desirable to ensure the validity of expression

$$\Delta t \le k / |\lambda|_{\max}, \tag{8.52}$$

where k is taken to be 0.1–0.2, and $|\lambda|_{\max}$ is the maximum module of dynamics matrix eigenvalues.

♦ **Example 8.5.** Obtain the discrete variant of the filtering problem for an exponentially correlated process described by the equation

$$\dot{x}(t) = -\alpha x(t) + \sqrt{2\sigma_x^2 \alpha}\, w(t),$$

by measurements

$$y(t) = x(t) + v(t),$$

where σ_x^2 is the process variance; $\tau_c = 1/\alpha$ is the correlation interval; $w(t)$ is the white noise of unit PSD; $v(t)$ is the measurement white noise independent of $w(t)$ and $x(0)$ with PSD r^2.

Using (8.35), (8.37), (8.38), we can write:

$$\Phi = e^{-\alpha \Delta t},\ \Gamma = \sqrt{2\sigma_x^2 \alpha} \int_0^{\Delta t} e^{-\alpha \tau} d\tau = \frac{\sqrt{2\sigma_x^2 \alpha}}{\alpha}(1 - e^{-\alpha \Delta t}),\ Q_d = \frac{1}{\Delta t},\ r_d^2 = \frac{r^2}{\Delta t}.$$

Thus, the discrete problem is formulated as follows. Estimate the sequence

$$x_i = e^{-\alpha \Delta t} x_{i-1} + \frac{\sqrt{2\sigma_x^2 \alpha}}{\alpha}(1 - e^{-\alpha \Delta t}) w_i,$$

by using measurements

$$y_i = x + v_i,$$

where $y_i = y(t_i) = \int_{t_i}^{t_i + \Delta t} y(\tau) d\tau$, $v_i = v(t_i) = \int_{t_i}^{t_i + \Delta t} v(\tau) d\tau$.

With consideration for (8.52) and the fact that $\lambda = \tau_k$, it is recommended to account for the condition $\dfrac{\Delta t}{\tau_k} \le 0.1$ in selecting the sampling interval. ♦

Exercises

Exercise 8.1. Show that the mathematical expectation, the covariance matrix and the correlation function of the process defined by the shaping filter (8.8) will be

determined by (8.10)–(8.12).

Exercise 8.2. Let $w(t)$ be a zero-mean white noise with PSD Q. Introduce a random sequence

$$w_i = \frac{1}{\Delta t}\int_{t_i}^{t_i+\Delta t} w(t)dt\,,$$

where Δt is the sampling interval.

Find the correlation function for this sequence.

Exercise 8.3. Having the equations for the process

$$\dot{x}(t) = F(t)x(t) + G(t)w(t)\,.$$

and estimation equations

$$\dot{\hat{x}}(t) = F(t)\hat{x}(t) + K(t)(y(t) - H(t)\hat{x}(t))\,,$$

write the equations for filtering errors. Is the process describing the filtering errors a stationary and Markov one?

Exercise 8.4. Write the equations for the estimate and a posteriori variance for the Kalman-Bucy filter in the filtering of the Wiener process described by $\dot{x}(t) = \sqrt{2\sigma_x^2\alpha}\,w(t)$ using measurements $y(t) = x(t) + v(t)$, where $v(t)$ is measurement white noise with PSD r^2 independent of $w(t)$ and $x(0)$. Obtain the solutions to these equations.

Exercise 8.5. Write the equations for the estimate and a posteriori variance for the Kalman-Bucy filter in the filtering of the Markov process (8.21) using measurements $y(t) = x(t) + v(t)$, where $v(t)$ is measurement white noise with PSD r^2 independent of $w(t)$ and $x(0)$. Obtain the solutions to these equations.

Exercise 8.6. Obtain the formulas for a posteriori variance for the continuous problem of estimating a zero-mean random value with variance σ_0^2 by measurements $y(t) = x + v(t)$, where $v(t)$ is the white noise with PSD r^2, independent of x. Obtain a similar formula for discrete analog of the continuous problem and compare these formulas.

Exercise 8.7. Formulate the problem of integrated processing of aircraft height measurements by the measurements

$$y^{\mathrm{SNS}}(t) = h(t) + \Delta h^{\mathrm{SNS}}(t)\,;$$
$$y^{\mathrm{BA}}(t) = h(t) + \Delta h^{\mathrm{BA}}(t)$$

from the satellite and barometric systems as a problem of optimal Kalman filtering in the form (8.1), (8.2) assuming that errors $\Delta h^l(t)$, $l = SNS, BA$ can be described by noncorrelated random processes, each being a sum of white noise $v^l(t)$ with known PSD r_l^2 and exponentially correlated first-order Markov process $\delta^l(t)$ with known

variance σ_l^2 and correlation interval $\tau^l = 1/\alpha^l$, i.e., $\Delta h^l(t) = \delta^l(t) + v^l(t)$.

Consider two variants of the problem.

Variant 1. Let the aircraft height be presented as a sum of the first-order polynomial with unknown coefficients determining the initial height $h(0)$, constant vertical velocity $V_h(0)$, and double integrated white noise of given PSD q_h^2.

Variant 2. There is no a priori data on height variation, and the problem is solved in the invariant statement.

Note. To formulate these problems, use invariant and noninvariant processing schemes.

Exercise 8.8. Write the Kalman-Bucy filter algorithm in the problem of estimating the vehicle position and speed if

$$\dot{X} = V;$$

$$\dot{V} = 0,$$

with only position measurements available, i.e.,

$$y = X + v,$$

where $v(t)$ is a zero-mean white noise with PSD r^2 noncorrelated with X.

Obtain the formulas for error variance of position and speed estimation.

Note. To derive the solution for the covariance matrix, it is recommended to proceed to the equation for inverse matrix (see Note 4 at the end of section 8.3).

Exercise 8.9. Show that problem 8.3 can be formulated as the problem of estimating two constant values. Obtain the formula for the covariance matrix for this formulation.

Test questions

1. Give the definitions of mathematical expectation, variance, and correlation function for the random process.
2. What kind of process is called a stationary process? What is the PSD of the process? How can the process variance be found using only its PSD?
3. Give the definitions of the shaping filter for a random process. Provide formulas for the mathematical expectation and the covariance matrix for the random process described by the shaping filter.
4. Formulate the problem of optimal linear filtering.
5. Formulate the problem of optimal Bayesian filtering.
6. Write the formulas for the continuous Kalman-Bucy filter and give your comments.
7. Compare the block diagrams of the shaping filter and the Kalman filter.
8. Explain the procedure for proceeding from the continuous filtering problem to its discrete variant.

Appendices

Appendix 1

A1. Introduction to matrix analysis

This appendix is the introduction to matrix analysis, in particular, the main matrix operations, often used in applied estimation, are considered. The materials of this appendix is borrowed from [2, 11, 12].

A **matrix** is a rectangular array of numbers, symbols, or expressions, arranged in n rows and m columns. An example of matrix A is

$$A=\begin{bmatrix} a_{11} & . & . & . & a_{1m} \\ . & . & . & . & . \\ a_{i1}. & . & a_{ij} & . & a_{im} \\ . & . & . & . & . \\ a_{n1} & . & . & . & a_{nm} \end{bmatrix}. \qquad \text{(A1.1)}$$

The following notations are used for matrices:

$$A=\left\{a_{ij}\right\};\ A=\left\{A(i,j)\right\};\ A=\left\{A[i,j]\right\},\ i=\overline{1.n},\ j=\overline{1.m}.$$

A **column vector**

$$a=\begin{bmatrix} a_1 \\ . \\ . \\ a_n \end{bmatrix}, \qquad \text{(A1.2)}$$

and a row vector

$$a=[a_1,.....a_n] \qquad \text{(A1.3)}$$

are the special kind of $n\times 1$ and $1\times n$ matrix.

A square matrix. A square matrix is a matrix with an identical number of rows and columns of dimension $n\times n$.

A diagonal matrix. A square matrix D of dimension n in which all the elements outside the main diagonal are zero is called a diagonal matrix.

$$D=\begin{bmatrix} d_{11} & 0 & . & 0 & 0 \\ 0 & d_{22} & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & d_{n-1n-1} & 0 \\ 0 & 0 & . & 0 & d_{nn} \end{bmatrix}. \qquad \text{(A1.4)}$$

The diagonal matrix is often defined as $D=diag\{d_{ii}\}, \quad i=\overline{1.n}$.

An identity matrix. An identity matrix is a diagonal matrix in which all the elements on the main diagonal are equal to 1. Such matrix is usually denoted by E, or E_n, where n is the dimensions of this matrix, i.e.,

$$E=\begin{bmatrix} 1 & 0 & . & 0 & 0 \\ 0 & 1 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & 1 & 0 \\ 0 & 0 & . & 0 & 1 \end{bmatrix}. \qquad \text{(A1.5)}$$

Transpose of a matrix. Let $A=\{a_{ij}\}, i=\overline{1.n}, j=\overline{1.m}$ be an $n\times m$ matrix, then transpose A is an $m\times n$ matrix, which is defined as

$$A^{\mathrm{T}}=\{a_{ji}\}, j=\overline{1.m}, i=\overline{1.n}. \qquad \text{(A1.6)}$$

The rows of such a matrix are the columns of A, but the rows of A are the columns of A^{T}.

An **upper triangular matrix** is a matrix where non-zero elements are on and above the main diagonal

$$A=\begin{bmatrix} * & * & . & * & * \\ 0 & * & . & * & * \\ . & . & . & . & . \\ 0 & 0 & . & * & * \\ 0 & 0 & . & 0 & * \end{bmatrix}. \qquad \text{(A1.7)}$$

A **lower triangular matrix** has non-zero elements on and below the main diagonal.

Symmetric matrix. A square matrix of dimension *n* is a symmetric matrix if

$$A=A^{\mathrm{T}}. \qquad \text{(A1.8)}$$

Block matrix. A block matrix is the following one:

$$A=\left[\begin{array}{c|c|c|c|c} A_{11} & . & . & . & A_{1m} \\ \hline . & . & . & . & . \\ \hline A_{i1}. & . & A_{ij} & . & A_{im} \\ \hline . & . & . & . & . \\ \hline A_{n1} & . & . & . & A_{nm} \end{array}\right], \tag{A1.9}$$

where A_{ij}, $i=\overline{1.n}, j=\overline{1.m}$ are matrices, i.e., the matrix formed from blocks or submatrix is called a block matrix.

A **block diagonal matrix** is a matrix in which the blocks on the main diagonal are non-zero.

A trace of a matrix. The trace of a square matrix A is defined to be the sum of elements on the main diagonal, i.e.,

$$TrA=\sum_{i=1}^{n} a_{ii}\,. \tag{A1.10}$$

Sometimes it is denoted as

$$SpA=\sum_{i=1}^{n} a_{ii}\,. \tag{A1.11}$$

This notation comes from German ‘Spur’.

Matrix addition. Let A,B,C be the matrices of $n\times m$ dimension. The elements of C matrix, corresponding to the sum A and B ,

$$C=A+B, \tag{A1.12}$$

are defined as

$$c_{ij}=a_{ij}+b_{ij}, i=\overline{1.n},\ j=\overline{1.m}. \tag{A1.13}$$

Multiplication of matrix A by scalar α is the multiplication of each element of the matrix by the scalar, i.e.,

$$\alpha A=\left\{\alpha a_{ij}\right\}. \tag{A1.14}$$

Matrix multiplication. Let A and B be $n\times m$ and $m\times l$ matrices, then the elements of $n\times l$ matrix C, corresponding to the product AB,

$$C=AB, \tag{A1.15}$$

are defined as $c_{ij}=\sum_{k=1}^{m} a_{ik}b_{kj}, i=\overline{1.n},\ j=\overline{1.l}$.

Therefore, the matrix multiplication is correct if the number of columns in the first matrix is equal to the number of rows in the second, i.e.,

$$C=AB, \tag{A1.16}$$

and it has the following structure:

$$\underbrace{n\left\{\overbrace{\begin{bmatrix} c_{11} & . & . & c_{1m} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ c_{n1} & . & . & c_{nm} \end{bmatrix}}^{C}\right.}_{m} = \underbrace{n\left\{\overbrace{\begin{bmatrix} a_{11} & . & . & . & a_{l1} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{n1} & . & .. & . & a_{nl} \end{bmatrix}}^{A}\right.}_{l} \underbrace{l\left\{\overbrace{\begin{bmatrix} b_{11} & . & . & . & b_{1m} \\ . & . & . & . & . \\ . & . & . & . & . \\ b_{l1} & . & . & . & b_{lm} \end{bmatrix}}^{B}\right.}_{m}.$$

In this case, we say that the matrices have consistent dimensions.

As a result of multiplication of $1\times n$ row a^{T} by $n\times 1$ column b, we have a scalar. At the same time, we have a square matrix from multiplication of column b and row a:

$$ba^{\mathrm{T}} = \begin{bmatrix} b_1a_1 & . & . & b_1a_n \\ . & . & . & . \\ . & . & . & . \\ b_na_1 & . & . & b_na_n \end{bmatrix}. \qquad \text{(A1.17)}$$

A vector norm is the value

$$\|a\| = \sqrt{\sum_{i=1}^{n} a_i^2} = \sqrt{a^{\mathrm{T}}a}\,. \qquad \text{(A1.18)}$$

In general, matrix multiplication is non-commutative, i.e.,

$$AB \neq BA.$$

If $AB = BA$, the matrices are called **commutative**.

The following equality $a^{\mathrm{T}}a = Sp(aa^{\mathrm{T}})$ is true.

Conventional rules of matrix multiplication can be used to **multiply block matrices**, i.e.,

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{12}+A_{12}B_{22} \\ A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{12}+A_{22}B_{22} \end{bmatrix}.$$

The **determinant** of the square matrix $A = \{a_{ij}\}$ of dimension *n* can be calculated using the determinants for the matrix of dimension $n-1$ as

$$\det(A) = \sum_{j=1}^{n} a_{1j}A_{1j}, \qquad \text{(A1.19)}$$

where A_{1j} are **cofactors**, i.e., the determinants for the matrix of dimension $n-1$ obtained by deleting the first row and the *j*-th column from matrix *A* and multiplying by $(-1)^{1+j}$.

A similar representation can be obtained for an arbitrary *i*-th row:

$$\det(A) = \sum_{j=1}^{n} a_{ij} A_{ij},$$

where A_{ij} are the determinants of order $n-1$ obtained by deleting the *i*-th row and the *j*-th column and multiplying by $(-1)^{i+j}$.

For example, in the case of a two-dimensional matrix, it is easy to find that

$$\det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{21}a_{12}.$$

The determinant of the matrix is denoted as

$$\det(A) = |A|.$$

A square matrix is called **non-singular** if its determinant is not zero, i.e.,

$$\det(A) = |A| \neq 0.$$

Otherwise, we have a **singular** matrix.

Let us introduce the following notation of the determinants formed of elements of a rectangular $n \times m$ matrix:

$$A\begin{pmatrix} i_1\ i_2 \ldots i_p \\ k_1 k_2 \ldots k_p \end{pmatrix} = \begin{vmatrix} a_{i_1k_1} & a_{i_1k_2} & \ldots & a_{i_1k_p} \\ a_{i_2k_1} & a_{i_2k_2} & \ldots & a_{i_2k_p} \\ . & . & . & . \\ a_{i_pk_1} & a_{i_pk_2} & \ldots & a_{i_pk_p} \end{vmatrix}.$$

Such a determinant is called a **minor** of order *p* if $1 \le i_1 < i_2 < \ldots < i_p \le n$ and $1 \le k_1 < k_2 < \ldots < k_p \le m$.

If $i_1 = k_1, i_2 = k_2, .. i_p = k_p$, such minors are called **principal** minors.

The **rank of the matrix** is the largest order of any non-zero minor in the matrix. If r is a rank of rectangular *n*×*m* matrix, then, obviously, $r < n, m$.

An inverse matrix. Let $A = \{a_{ij}\}$ be a non-singular *n*× *n* matrix. An inverse A^{-1} matrix is the matrix satisfying

$$AA^{-1} = E. \qquad \text{(A1.20)}$$

Let us denote the inverse matrix as $A^{-1} = B = \{b_{ij}\}$.

For the elements of the inverse matrix, the following equality is valid [2]:

$$b_{ij} = \frac{A_{ji}}{\det(A)},\ \det(A) = a_{11}A_{11} + .. + a_{1n}A_{1n}, \tag{A1.21}$$

where A_{ji} are cofactors.

For example, in the case of a two-dimensional matrix, it is easy to find

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

The inverse matrix satisfies the following equality: $A^{-1}A = E$, i.e., the matrix and the inverse matrix are commutative.

Orthogonal matrix. A square matrix A is orthogonal if

$$AA^{\mathrm{T}} = E. \tag{A1.22}$$

A **characteristic polynomial** of a square matrix A is the polynomial

$$p(\lambda) = \det(A - \lambda E). \tag{A1.23}$$

For example, in the case of a two-dimensional matrix, this polynomial looks as follows:

$$p(\lambda) = \det(A - \lambda E) = \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21}.$$

The **characteristic equation** is defined as

$$p(\lambda) = \det(A - \lambda E) = 0.$$

Eigenvalues of a matrix. Eigenvalues of a square matrix *A* of dimension *n* are defined as the *n* roots of the characteristic equation

$$p(\lambda) = \det(A - \lambda E) = 0. \tag{A1.24}$$

A set of all eigenvalues is called a **matrix spectrum.**

An **eigenvector** of a square matrix A is vector *x*, such that

$$Ax = \lambda x, \tag{A1.25}$$

i.e., $(A - \lambda E)x = 0$, where λ is the eigenvalue.

This equation means that the multiplication of the matrix by the eigenvector does not change its direction. Since the characteristic equation of the matrix of dimension *n* has *n* roots, every root (eigenvalue) has the eigenvector, respectively. Thereby, the matrix has *n* eigenvalues and eigenvectors. It should be noted that some eigenvalues may coincide, and but the same eigenvalues may correspond to different eigenvectors. The problem of finding them is known as an **eigenvalue problem**.

Similarity transformation. Let A and C be $n \times n$ square matrices, and C is non-degenerate. Then matrix B is defined as

$$B = CAC^{-1}. \tag{A1.26}$$

Matrices B and A are called **similar matrices,** and matrix C is **similarity transformation.** It is easy to see that similar matrices have the same characteristic polynomial.

Indeed,

$$\det(B-\lambda E)=\det(CAC^{-1}-\lambda CC^{-1})=\det(A-\lambda E)\det(C)\det(C^{-1})=$$

$$=\det(A-\lambda E)\det(CC^{-1})=\det(A-\lambda E).$$

It follows that similar matrices have the same eigenvalues.

It can be easily verified that the determinant and the trace of the square matrix A could be represented using eigenvalues, i.e.,

$$TrA=\sum_{i=1}^{n}\lambda_i, \tag{A1.27}$$

$$\det(A)=\prod_{i=1}^{n}\lambda_i. \tag{A1.28}$$

Therefore, similar matrices have the same determinants and traces.

Diagonalization of symmetric matrices. Symmetric matrix A can always be diagonalized by using an orthogonal matrix, i.e., there always exists an orthogonal matrix, such that

$$T^{\mathrm{T}}T=E,\ T^{\mathrm{T}}=\left[t_1\quad ...\quad t_j\quad ...\quad t_n\right], \tag{A1.29}$$

where

$$TAT^{\mathrm{T}}=\left\{\lambda_j\delta_{ij}\right\},\quad j=\overline{1.n}, \tag{A1.30}$$

and $\lambda_j, t_j,\ j=\overline{1.n}$ are the eigenvalues and eigenvectors of matrix A, i.e.,

$$A\,t_j=\lambda_j t_j, \tag{A1.31}$$

moreover, $t_i^{\mathrm{T}}t_j=\delta_{ij}$. Here, the value δ_{ij} is the **Kronecker delta**

$$\delta_{ij}=\begin{cases}1,\ i=j,\\ 0,\ i\neq j.\end{cases} \tag{A1.32}$$

The above definition implies that there is an additional condition for norms of eigenvalues, i.e., its norms must be equal to 1.

Note that if a matrix is not symmetric, it is not always possible to convert it into a diagonal matrix. However, an arbitrary square matrix can be converted to other matrices of special (normal or canonical) types, such as Jordan matrix, Frobenius matrix, etc.

A **quadratic form**. Let *A* be a square matrix and *x* is a vector of dimension *n*. The **quadratic form** is determined as

$$y=x^{\mathrm{T}}Ax. \tag{A1.33}$$

The quadratic form is called **positive definite** if for all nonzero vectors *x*, the quadratic form is positive, i.e.,

$$y=x^{\mathrm{T}}Ax>0,\ \text{for}\ x\neq 0. \tag{A1.34}$$

In this case, matrix A is called **positive definite**.

If the strict inequality is replaced by $\geq$, then the quadratic form and the corresponding matrix are called **nonnegative definite**.

Reversing the inequality sign, we get a **negative definite** quadratic form and a negative definite matrix.

Matrix inequality. If the following inequality holds

$$x^{\mathrm{T}}Ax > x^{\mathrm{T}}Bx > 0 \qquad \textbf{(}\text{A1.35)}$$

or $x^{\mathrm{T}}(A-B)x > 0$ and the vector x is not zero, then we can say that matrix $A>B$. Similarly, you can introduce other types of inequalities.

One can show that when some inequality holds, the same inequality is valid for its eigenvalue. So, if all eigenvalues of the square matrix A are positive or non-negative (negative or non-positive), i.e., $\lambda_j > 0$ or $\lambda_j \geq 0$ ($\lambda_j < 0$ or $\lambda_j \leq 0$), $j = \overline{1.n}$, then the corresponding inequalities are valid for the matrix, i.e., $A > 0$ or $A \geq 0$, ($A < 0$ or $A \leq 0$).).

Functions of matrices. Certain functions of square matrices can be given by power series. M matrix A^m of degree m is defined as

$$A^m = \underbrace{A \cdot A \ldots \cdot A}_{m}. \qquad \text{(A1.36)}$$

From the above definition, we can obtain the **exponential of matrix**, defined as

$$e^A = \exp(A) = \sum_{i=0}^{\infty} \frac{A^i}{i!}. \qquad \text{(A1.37)}$$

Functions $f(A)$ of symmetric matrices can be obtained in different ways. Firstly, matrix A must be diagonalized by the orthogonal matrix, i.e., we get the following representation:

$$TAT^{\mathrm{T}} = \Lambda = diag\left\{\lambda_j\right\}, \quad j = \overline{1.n},$$

from which it follows that $A = T^{\mathrm{T}}\Lambda T$.

Next, $f(A)$ is defined as

$$f(A) = T^{\mathrm{T}} f(\Lambda) T, \qquad \text{(A1.38)}$$

where $f(\Lambda) = diag\left\{f(\lambda_j)\right\}, \quad j = \overline{1.n}$.

The derivative and integral of a matrix, whose elements depend, for example, on time, i.e., $A(t) = \left\{a_{ij}(t)\right\}, i = \overline{1.n}, j = \overline{1.m}$, is matrix $\dot{A}(t) = \left\{\dot{a}_{ij}(t)\right\}, i = \overline{1.n}, j = \overline{1.m}$ or $\int_0^t A(\tau)d\tau = \left\{\int_0^t a_{ij}(\tau)d\tau\right\}, i = \overline{1.n}, j = \overline{1.m}$, whose elements are the derivatives or integrals of the original matrix.

The **derivative of the scalar-valued function** $s(x)$ **of the vector argument** is calculated by the formula:

$$\frac{ds(x)}{dx^{\mathrm{T}}} = \left[\frac{\partial s(x)}{\partial x_1} \ldots \frac{\partial s(x)}{\partial x_n} \right]. \tag{A1.39}$$

Notation $\frac{ds(x)}{dx}$ means that

$$\frac{ds(x)}{dx} = \begin{bmatrix} \frac{\partial s(x)}{\partial x_1} \\ \ldots\ldots\ldots \\ \frac{\partial s(x)}{\partial x_n} \end{bmatrix}, \tag{A1.40}$$

thereby,

$$\frac{ds(x)}{dx} = \left[\frac{ds(x)}{dx^{\mathrm{T}}} \right]^{\mathrm{T}}. \tag{A1.41}$$

This implies that

$$\frac{d^2 s(x)}{dx dx^{\mathrm{T}}} = \frac{d}{dx}\left[\frac{ds(x)}{dx^{\mathrm{T}}} \right] = \begin{bmatrix} \frac{\partial^2 s(x)}{\partial x_1 \partial x_1} \ldots \frac{\partial^2 s(x)}{\partial x_1 \partial x_n} \\ \ldots\ldots\ldots\ldots\ldots \\ \frac{\partial^2 s(x)}{\partial x_n \partial x_1} \ldots \frac{\partial^2 s(x)}{\partial x_n \partial x_n} \end{bmatrix}, \quad \frac{d^2 s(x)}{dx dx^{\mathrm{T}}} = \left[\frac{d^2 s(x)}{dx^{\mathrm{T}} dx} \right]^{\mathrm{T}}. \tag{A1.42}$$

The derivative of an *m*-vector-valued function of a vector argument $x = (x_1, \ldots x_n)^{\mathrm{T}}$ is defined as

$$s(x) = \begin{bmatrix} s_1(x_1, \ldots x_n) \\ . \\ . \\ s_m(x_1, \ldots x_n) \end{bmatrix}. \tag{A1.43}$$

Thus,

$$\frac{ds(x)}{dx^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial s_1(x)}{\partial x_1} \ldots \frac{\partial s_1(x)}{\partial x_n} \\ \ldots\ldots\ldots\ldots\ldots \\ \frac{\partial s_m(x)}{\partial x_1} \ldots \frac{\partial s_m(x)}{\partial x_n} \end{bmatrix}, \text{ and } \frac{ds^{\mathrm{T}}(x)}{dx} = \begin{bmatrix} \frac{\partial s_1(x)}{\partial x_1} \ldots \frac{\partial s_m(x)}{\partial x_1} \\ \ldots\ldots\ldots\ldots\ldots \\ \frac{\partial s_1(x)}{\partial x_n} \ldots \frac{\partial s_m(x)}{\partial x_n} \end{bmatrix}, \tag{A1.44}$$

$$\text{i.e.,} \quad \frac{ds^{\mathrm{T}}(x)}{dx} = \left[\frac{ds(x)}{dx^{\mathrm{T}}} \right]^{\mathrm{T}}.$$

In general, if A is a square symmetric matrix, then $\frac{d}{dx}[g^{\mathrm{T}}(x)Ag(x)] = 2\frac{dg^{\mathrm{T}}(x)}{dx}Ag(x)$.

The derivative of scalar function $s(A)$ **of a matrix** (**matrix gradient**) is determined as

$$\frac{ds(A)}{dA} = \left\{\frac{ds}{da_{ij}}\right\}.$$

For example, for square matrix A, the following formulas are valid:

$$\frac{ds(Sp(A))}{dA} = E\,,\ \frac{ds(Sp(BAC))}{dA} = B^{\mathrm{T}}C^{\mathrm{T}};$$

$$\frac{ds(Sp(ABA^{\mathrm{T}}))}{dA} = A(B + B^{\mathrm{T}}). \qquad \text{(A1.45)}$$

Block matrix inversion formula. Suppose

$$P = \begin{bmatrix} P^x & P^{xy} \\ (P^{xy})^{\mathrm{T}} & P^y \end{bmatrix}, \qquad \text{(A1.46)}$$

where P^x, P^y, P^{xy} are $n\times n$, $m\times m$ and $n\times m$ matrices; moreover, there are inverse matrices for P^x, P^y. In this case, inverse matrix P^{-1} is defined as

$$P^{-1} = \begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}, \qquad \text{(A1.47)}$$

where

$$A = \left[P^x - P^{xy}\left(P^y\right)^{-1}P^{yx}\right]^{-1},$$

$$B = -\left[P^x - P^{xy}\left(P^y\right)^{-1}P^{yx}\right]^{-1}P^{xy}\left(P^y\right)^{-1},$$

$$C = \left[P^y - P^{yx}\left(P^x\right)^{-1}P^{xy}\right]^{-1}.$$

The following relations are valid:

$$A = \left(P^x\right)^{-1} + \left(P^x\right)^{-1}P^{xy}CP^{yx}\left(P^x\right)^{-1},$$

$$B = -AP^{xy}\left(P^y\right)^{-1} = -\left(P^x\right)^{-1}P^{xy}C\,,$$

$$C = \left(P^y\right)^{-1} + \left(P^y\right)^{-1}P^{yx}AP^{xy}\left(P^y\right)^{-1}.$$

There are some useful matrix relations in Table A1.1.

In particular, the so-called **matrix inversion lemma** is

$$\left[P^{-1} + H^{\mathrm{T}}R^{-1}H\right]^{-1} = P - PH^{\mathrm{T}}\left(HPH^{\mathrm{T}} + R\right)^{-1}HP\,, \qquad \text{(A1.48)}$$

where P, R are the square non-singular matrices of dimension n and m, and H is an $m \times n$ matrix.

It follows from the lemma that

$$\left[P^{-1} + R^{-1}\right]^{-1} = P - P\left(P + R\right)^{-1} P, \qquad \text{(A1.49)}$$

and

$$\left[r^2 E + q^2 I\right]^{-1} = \frac{1}{r^2}\left(E - \frac{q^2}{nq^2 + r^2} I\right). \qquad \text{(A1.50)}$$

The last formula is easy to get if

$$P^{-1} = r^2 E,\ H = [1,1...1],\ R^{-1} = q^2.$$

Table A1.1

Some useful matrix formulas

Relation	Note
$(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}}$ (A1.51)	A, B – matrices of consistent dimensions
$Tr(ABC) = Tr(BCA) = Tr(CAB)$ (A1.52)	A, B, C – arbitrary matrices of consistent dimensions
$Tr(A) = \sum_{i=1}^{n} \lambda_i$ (A1.53)	A – square matrix, λ_i – eigenvalue of A
$\det(AB) = \det(A)\det(B)$ (A1.54)	A, B – square n-dimensional matrices
$\det(A) = \prod_{i=1}^{n} \lambda_i$ (A1.55)	A – square matrix, λ_i – eigenvalue of A
$(AB)^{-1} = B^{-1} A^{-1}$ (A1.56)	A, B – square non-singular n-dimensional matrices
$\left[P^{-1} + H^T R^{-1} H\right]^{-1} = = P - PH^T\left(HPH^T + R\right)^{-1} HP$ (A1.57)	P, R – square non-singular matrices of dimensions n and m, H – $m \times n$ matrix
$\left[P^{-1} + R^{-1}\right]^{-1} = P - P\left(P + R\right)^{-1} P$ (A1.58)	P, R – square non-singular n-dimensional matrices
$\left[r^2 E + q^2 I\right]^{-1} = \frac{1}{r^2}\left(E - \frac{q^2}{nq^2 + r^2} I\right)$ (A1.59)	I – a square n-dimensional matrix of units; r^2, q^2 – positive values

Table A1.1 (continued)

$\frac{\partial}{\partial x^{\mathrm{T}}}(Ax)=A,\ \frac{\partial}{\partial x}(x^{\mathrm{T}}A^{\mathrm{T}})=A^{\mathrm{T}}$ (A1.60)	A – square symmetric matrix, $x=(x_1,...x_n)^{\mathrm{T}}$ – n-vector
$\frac{\partial}{\partial x}(x^{\mathrm{T}}Ax)=2Ax,$ $\frac{\partial}{\partial x^{\mathrm{T}}}(x^{\mathrm{T}}Ax)=2x^{\mathrm{T}}A$ (A1.61)	
$\frac{\partial^2}{\partial x\partial x^{\mathrm{T}}}(x^{\mathrm{T}}Ax)=2A$ (A1.62)	
$\frac{\partial}{\partial x}[g^{\mathrm{T}}(x)Ag(x)]=2\frac{\partial g^{\mathrm{T}}(x)}{\partial x}Ag(x),$ in particular, $\frac{\partial}{\partial x}[(y-Hx)^{\mathrm{T}}A(y-Hx)]=$ $=2H^{\mathrm{T}}A(y-Hx)]$ (A1.63)	A – square symmetric matrix, $g(x)=(g_1(x_1,...x_n),...g_m(x_1,...x_n))^{\mathrm{T}}$ – m-vector valued function, x, y – n- and m-vectors, H– $m\times n$-matrix
$x^{\mathrm{T}}Ax-2x^{\mathrm{T}}z=$ $=(x-A^{-1}z)A(x-A^{-1}z)-z^{\mathrm{T}}A^{-1}z$ (A1.64)	A – square non-singular symmetric matrix, x, z – n-vectors
Using $P=\begin{bmatrix} P^x & P^{xy} \\ (P^{xy})^{\mathrm{T}} & P^y \end{bmatrix}$ find $P^{-1}=\begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}$ (A1.65)	$C=\left[P^y-P^{yx}\left(P^x\right)^{-1}P^{xy}\right]^{-1}$ $B=-\left[P^x-P^{xy}\left(P^y\right)^{-1}P^{yx}\right]^{-1}P^{xy}\left(P^y\right)$ $A=\left[P^x-P^{xy}\left(P^y\right)^{-1}P^{yx}\right]^{-1}$

Appendix 2

A2. Random variables and vectors

A2.1. Random variables

A random variable (RV) is a variable whose value is not known in advance and you can only introduce a numeric measure (probability) of the fact (event) that this value belongs to a prescribed domain on real axes.

We shall assume a random **variable** as given if the function enabling to determine the probability of any possible event is defined, i.e., we can calculate the probability that a random variable will belong to some interval or sets of intervals on real axes.

Such a function defining stochastic properties of a random variable is known as a **probability distribution function** or **cumulative distribution function (CDF),** which is a scalar function $F_{\mathbf{x}}(x)$ of a real-valued argument x. This function defines the probability that the random variable $\mathbf{x}$ belongs to the open interval $(-\infty, x)$ i.e., the probability that $\mathbf{x} < x$.

Thus,

$$F_{\mathbf{x}}(x) = \mathbf{Pr}(\mathbf{x} : \mathbf{x} < x). \tag{A2.1}$$

Sometimes term **a probability distribution** or **distribution function** can be used instead of a probability distribution function if it is clear from the context.

The **CDF** (A2.1) is a non-negative, non-decreasing, left-continuous function satisfying the following evident equalities:

$$F_{\mathbf{x}}(-\infty) = \mathbf{Pr}(\mathbf{x} : \mathbf{x} < -\infty) = 0; \tag{A2.2}$$

$$F_{\mathbf{x}}(\infty) = \mathbf{Pr}(\mathbf{x} : \mathbf{x} < \infty) = 1. \tag{A2.3}$$

In addition to the probability distribution function, the properties of random variables can be described by a **probability density function (PDF)**, defined as

$$p_{\mathbf{x}}(x) = \frac{dF_{\mathbf{x}}(x)}{dx}. \tag{A2.4}$$

The lower subscript of the PDF indicates the random variable to which it corresponds, and hereafter, it can be omitted if it does not lead to misunderstanding.

Integrating both parts of (A2.4) from $-\infty$ to x and taking into account the (A2.2), we obtain

$$F_{\mathbf{x}}(x) = \int_{-\infty}^{x} p_{\mathbf{x}}(u)du. \tag{A2.5}$$

The probability density function is a non-negative ($f_{\mathbf{x}}(x) \geq 0$) function satisfying the **normalizing condition**

$$\int_{-\infty}^{\infty} p_{\mathbf{x}}(u)du = 1. \qquad \text{(A2.6)}$$

For the probability of an event, in the case when $x_1 \le \mathbf{x} < x_2$, the following evident equations are valid:

$$\mathbf{Pr}(x_1 \le \mathbf{x} < x_2) = \mathbf{Pr}(\mathbf{x} : \mathbf{x} < x_2) - \mathbf{Pr}(\mathbf{x} : \mathbf{x} < x_1) = F_{\mathbf{x}}(x_2) - F_{\mathbf{x}}(x_1) \qquad \text{(A2.7)}$$

or

$$\mathbf{Pr}(x_1 \le \mathbf{x} < x_2) = F_{\mathbf{x}}(x_2) - F_{\mathbf{x}}(x_1) = \int_{x_1}^{x_2} p_{\mathbf{x}}(u)du\,. \qquad \text{(A2.8)}$$

Using formula (A2.4) , we can write

$$p_{\mathbf{x}}(x) = \lim_{dx \to 0} \frac{F_{\mathbf{x}}(x+dx) - F_{\mathbf{x}}(x)}{dx} = \lim_{dx \to 0} \frac{\mathbf{Pr}(x \le \mathbf{x} < x+dx)}{dx};$$

thus, for small dx

$$\mathbf{Pr}(x \le \mathbf{x} < x+dx) = F_{\mathbf{x}}(x+dx) - F_{\mathbf{x}}(x) \approx p_{\mathbf{x}}(x)dx\,. \qquad \text{(A2.9)}$$

A2.2. Stochastic characteristics of random variables

In addition to CDF and PDF, stochastic (statistical) properties of random variables can be described by a set of stochastic numerical characteristics. The main ones are a **mathematic expectation (mean)**, **moments**, **variance**, **root-mean-square value, root-mean-square deviation** (RMSD), also known as a **standard deviation or standard**. Speaking about an error, we use the term a **root-mean-square error (RMSE).** The relations between these characteristics and PDF are determined by equations presented in Table A2.1.

Table A2.1

The main stochastic characteristics of a random variable

Characteristic	Definition
Mathematic expectation (mean)	$E_{\mathbf{x}}(x) = \bar{x} = \int x p_{\mathbf{x}}(x)dx$ (A2.10)
Moment of *n* order	$E_{\mathbf{x}}(x^n) = \int x^n p_{\mathbf{x}}(x)dx$ (A2.11)
Central moment of *n* order	$E_{\mathbf{x}}(x-\bar{x})^n = \int (x-\bar{x})^n p_{\mathbf{x}}(x)dx$ (A2.12)
Variance	$D = \int (x-\bar{x})^2 p_{\mathbf{x}}(x)dx$ (A2.13)
Root-mean-square deviation (RMSD)	$\sigma = \left(\int (x-\bar{x})^2 p_{\mathbf{x}}(x)dx \right)^{1/2}$ (A2.14)
Root-mean-square value (RMSV)	$\sqrt{E_{\mathbf{x}}(x^2)} = \sqrt{\int x^2 p_{\mathbf{x}}(x)dx}$ (A2.15)

*Notation***.** In the above equations, the limits of integration are assumed infinite.

Further, when the limits are not denoted, they are considered to be infinite.

A random variable with a zero expected value is called a **zero-mean random variable.**

Considering (A2.14), (A2.16), we can obtain the following useful formula for the variance of an RV:

$$D = E_{\mathbf{x}}\left(x^2\right) - \bar{x}^2. \qquad \text{(A2.16)}$$

The variance of an RV determines the measure of PDF concentration in the the vicinity of the expected value. This fact is reflected in **P.L. Chebyshev's inequality.** For RV $\mathbf{x}$, with the mean value $\bar{x}$ and variance D, where $\varepsilon > 0$, we can write the following:

$$\mathbf{Pr}(|\mathbf{x} - \bar{x}| \geq \varepsilon) \leq \frac{D}{\varepsilon^2}.$$

The validity of this inequality follows from the definition of the variance.

Indeed,

$$D = \int_{-\infty}^{\infty} (x - \bar{x})^2 p_{\mathbf{x}}(x) dx \geq \varepsilon^2 \int_{|x-\bar{x}| \geq \varepsilon} p(x) dx = \varepsilon^2 \mathbf{Pr}(|\mathbf{x} - \bar{x}| \geq \varepsilon).$$

This implies that if the variance decreases, the probability that a random variable falls outside the domain $(\bar{x} - \varepsilon \leq \mathbf{x} \leq \bar{x} + \varepsilon)$ decreases too.

To solve the applied problems, an important characteristic of a random variable and the corresponding CDF and PDF is a **quantile.**

A2.3. Gaussian random variables

Gaussian RVs are most widely used in solving applied problems. A **Gaussian or normal random variable** is a variable for which CDF and PDF can be written as follows:

$$F_{\mathbf{x}}(x) = \frac{1}{(2\pi)^{1/2}\sigma} \int_{-\infty}^{x} \exp\left\{-\frac{(t-\bar{x})^2}{2\sigma^2}\right\} dt; \qquad \text{(A2.17)}$$

$$p_{\mathbf{x}}(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{-\frac{(x-\bar{x})^2}{2\sigma^2}\right\}.$$

These functions are called a **Gaussian (normal) probability distribution and a Gaussian (normal) probability density function**.

Further, for Gaussian PDF, we will use the following notation:

$$p_{\mathbf{x}}(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{-\frac{(x-\bar{x})^2}{2\sigma^2}\right\} = N(x; \bar{x}, \sigma^2). \qquad \text{(A2.18)}$$

Gaussian CDF and PDF and their dependences on expected values and RMSD are

shown in Fig. A2.1.

It follows that when the area, in which PDF is essential nonzero, decreases, the variance decreases too. It can be shown that

$$\lim_{\sigma\to 0}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\bar{x})^2}{2\sigma^2}\right\}=\delta(x-\bar{x}),$$

where $\delta(\bullet)$ is the delta-function.

Note that for a Gaussian PDF, the median, the mean value and the mode are equal.

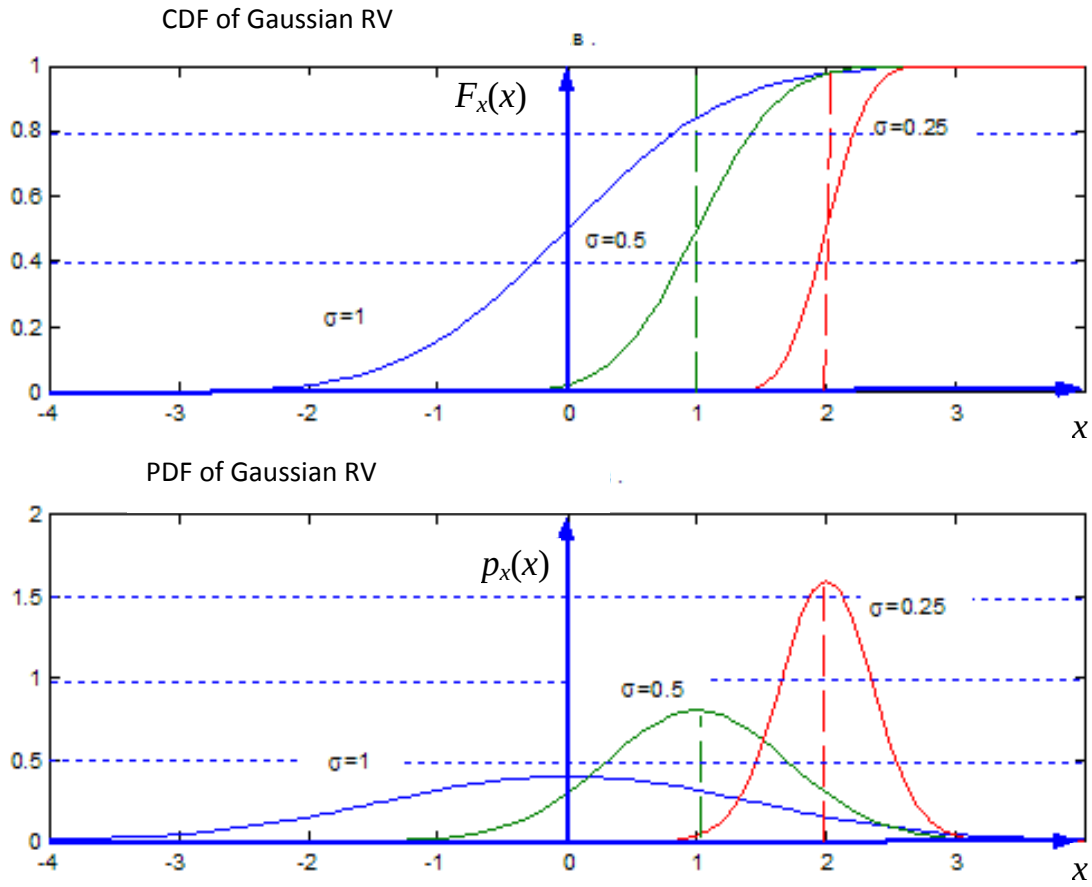


Fig. A2.1. Graphs of CDF and PDF of Gaussian RV at different expected values ($\bar{x}=0,\bar{x}=1,\bar{x}=2$) and RMSD $\sigma=1;\sigma=0.5;\quad\sigma=0.25$

For the Gaussian CDF corresponding to the zero-mean RV, the following formula is valid:

$$F_{\mathbf{x}}(x)=1-F_{\mathbf{x}}(-x).$$

The odd central moments of the Gaussian random variable are zero, i.e.,

$$\int(x-\bar{x})^{2k-1}p_{\mathbf{x}}(x)dx=0,$$

and the following formula is valid for the even moments:

$$\int(x-\bar{x})^{2k}p_{\mathbf{x}}(x)dx=1\times3\times..(2k-1)\sigma^{2k},k=1,2....\qquad\text{(A2.19)}$$

From the above formulas follows that parameters $\bar{x}$ and σ^2 of the functions (A2.17), (A2.18) are the mathematical expectation and the variance of the Gaussian random variable.

The probability values $\mathbf{Pr}\left[\bar{x}-k\sigma \le \mathbf{x} < \bar{x}+k\sigma\right] = \mathbf{Pr}\left[\left|\mathbf{x}-\bar{x}\right| \le k\sigma\right]$ of the Gaussian RV at different k are presented in Table A2.2 and in Fig. A2.2.

Table A2.2

Probability $\mathbf{Pr}\left[\left|\mathbf{x}-\bar{x}\right| \le k\sigma\right]$ for the Gaussian random value for different k

k	1	2	3	4
$\mathbf{Pr}\left[\left\|\mathbf{x}-\bar{x}\right\| \le k\sigma\right]$	0.6827	0.9545	0.9973	0.9999

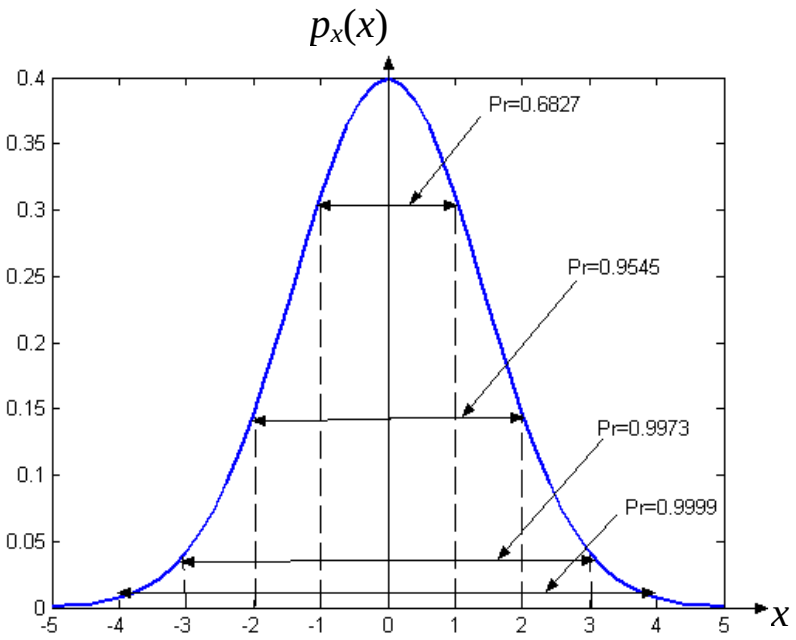


Fig. A2.2. Probability $\mathbf{Pr}\left[\left|\mathbf{x}-\bar{x}\right| \le k\sigma\right]$ for the Gaussian RV if $p_{\mathbf{x}}(x) = N(x;0,\sigma^2)$, $k = 1,2,3,4$

From Table A2.2 follows that for the modulus of the zero-mean Gaussian RV, i.e., $\left|\mathbf{x}-\bar{x}\right|$, the quantile of order 0.6827 equals to σ, and the probability that the value of the zero-mean Gaussian RV belongs to the interval $\pm 3\sigma$ is equal to 0.997. Usually, the value which is equal to 3σ is called a **three-sigma value** or a **three- sigma error** if an RV describes some measurement errors**.**

The fact that for a Gaussian RV

$$\mathbf{Pr}\left[\left|\mathbf{x}-\bar{x}\right| \le 3\sigma\right] = 0.997 \qquad \text{(A2.20)}$$

is called a **three-sigma rule.**

The following quantitative characteristics are often used for Gaussian RVs.

An average absolute deviation is defined as the expected value of $|\mathbf{x}-\bar{x}|$, i.e.,

$$E_{\mathbf{x}}\left\{|\mathbf{x}-\bar{x}|\right\}=\sigma E_{|\mathbf{u}|}|\mathbf{u}|=\sqrt{\frac{2}{\pi}}\sigma\approx 0.798\sigma. \qquad \text{(A2.21)}$$

A probable deviation (probable error) ε, which is a quantile of order $\frac{1}{2}$ for $|\mathbf{x}-\bar{x}|$, i.e.,

$$\mathbf{Pr}\left[|\mathbf{x}-\bar{x}|\le\varepsilon\right]=0{,}5. \qquad \text{(A2.22)}$$

In other words, it is such a value that probability of $|\mathbf{x}-\bar{x}|<\varepsilon$ is equal to probability of $|\mathbf{x}-\bar{x}|>\varepsilon$, i.e., it is a median for $|\mathbf{x}-\bar{x}|$. For a standard Gaussian RV with zero mean and $\sigma=1$, the probable deviation is given as $\varepsilon=0.674$, so that for the zero-mean Gaussian RV with variance σ^2, we obtain

$$\varepsilon\approx 0.674\sigma. \qquad \text{(A2.23)}$$

A2.4. Random vectors

A random vector is a vector each component of which is a random variable. For a random vector $\mathbf{x}=(\mathbf{x}_1,...\mathbf{x}_n)^{\mathrm{T}}$, its properties are fully given by the joint **CDF** or joint **PDF**, defined as follows:

$$F_{\mathbf{x}}(x)=\mathbf{Pr}(\mathbf{x}_1<x_1,...,\mathbf{x}_n<x_n); \qquad \text{(A2.24)}$$

$$p_{\mathbf{x}}(x)=\frac{\partial^n F_{\mathbf{x}}(x)}{\partial x_1...\partial x_n}; \qquad \text{(A2.25)}$$

$$F_{\mathbf{x}}(x)=\int_{-\infty}^{x_n}...\int_{-\infty}^{x_1}p_{\mathbf{x}}(u)du_1...du_n. \qquad \text{(A2.26)}$$

Formula (A2.24) determines the probability of an event in which inequality $\mathbf{x}_j<x_j$, $j=\overline{1.n.}$ holds for each component.

As in the one-dimensional case, a joint PDF is a nonnegative function which satisfies the **normalizing condition**

$$\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}p_{\mathbf{x}}(x)dx_1...dx_n=1. \qquad \text{(A2.27)}$$

In addition, a joint PDF satisfies the **consistency condition**, which, at *m*<*n*, is written as

$$p_{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_m}(x_1,x_2,...,x_m)=\int...\int p_{\mathbf{x}}(x_1,x_2,...,x_m,x_{m+1}...x_n)dx_{m+1}...dx_n, \qquad \text{(A2.28)}$$

and is a symmetric function of its arguments. The latter means that the PDF for vector $\mathbf{x} = (\mathbf{x}_1, ...\mathbf{x}_n)^{\mathrm{T}}$ does not depend on the sequence in which its components are arranged, in particular, $p_{\mathbf{x}_i,\mathbf{x}_j}(x_i, x_j) = p_{\mathbf{x}_j,\mathbf{x}_i}(x_j, x_i)$.

The probability for a random vector to belong to domain Ω, mathematical expectation $\bar{x}$ and **covariance matrix** P, which is a generalization of the notion of variance to the multidimensional case, has the form:

$$\mathbf{Pr}(\mathbf{x} \in \Omega) = \int_{\Omega} p_{\mathbf{x}}(x)dx; \tag{A2.29}$$

$$\bar{x} = \int x p_{\mathbf{x}}(x)dx = E_{\mathbf{x}}(x); \tag{A2.30}$$

$$P = \int (x - \bar{x})(x - \bar{x})^{\mathrm{T}} p_{\mathbf{x}}(x)dx = E_{\mathbf{x}}(xx^{\mathrm{T}}) - \bar{x}\bar{x}^{\mathrm{T}}. \tag{A2.31}$$

Here and in what follows, integrals are understood as multiples. If the integration domain is not indicated, as noted above, the limits for each component are assumed to lie in the range from $-\infty$ to $+\infty$. The diagonal elements of the covariance matrix determine the variances of the corresponding random vector components.

Mathematical expectation $E_{\mathbf{x}_i,\mathbf{x}_j}\left\{(x_i - \bar{x}_i)(x_j - \bar{x}_j)\right\}$ for two random variables $\mathbf{x}_i$ and $\mathbf{x}_j$ is called a **correlation coefficient**. Thus, nondiagonal elements

$$P_{ij} = E_{\mathbf{x}_i,\mathbf{x}_j}\left\{(x_i - \bar{x}_i)(x_j - \bar{x}_j)\right\} = \iint (x_i - \bar{x}_i)(x_j - \bar{x}_j) p_{\mathbf{x}_i,\mathbf{x}_j}(x_i, x_j)dx_i dx_j,\ i \neq j,\ i, j = \overline{1.n}$$

determine the correlation coefficients between different components.

By virtue of PDF symmetry, equality $P_{ij} = P_{ji}$ is valid, which means that the covariance matrix is symmetric, that is, $P = P^{\mathrm{T}}$. An important property of the covariance matrix is the fact that it is a nonnegative definite matrix, that is, the one for which $x^{\mathrm{T}}Px \geq 0$ for any $x \neq 0$.

If the mathematical expectation of a random vector is zero, then, as in the scalar case, such a vector is called a **zero-mean vector**.

If $E_{\mathbf{x}_i,\mathbf{x}_j}\left\{(x_i - \bar{x}_i)(x_j - \bar{x}_j)\right\} = 0$, then random variables are called **uncorrelated or orthogonal.** Hence it follows that for a random vector whose components are not correlated with each other, the covariance matrix has a diagonal form. If we have two random vectors, then we can introduce a **cross-correlation matrix** defined as

$$B = \int \int (x - \bar{x})(y - \bar{y})^{\mathrm{T}} p_{\mathbf{x},\mathbf{y}}(x, y)dxdy,$$

where $p_{\mathbf{x},\mathbf{y}}(x, y)$ is a joint PDF. If this matrix is equal to zero, then random vectors are defined as **noncorrelated or orthogonal**.

The notion of independence of random variables is also important. Random variables are called independent if the joint PDF is equal to the PDF product for each of these random variables, i.e.,

$$p_{\mathbf{x}}(x_1 ... x_n) = \prod_{i=1}^{n} p_{\mathbf{x}_i}(x_i).$$

A definition for **independent random vectors** is introduced in a similar way.

Independent random vectors are uncorrelated since

$$E_{\mathbf{x}_i,\mathbf{x}_j}\left\{(x_i - \bar{x}_i)(x_j - \bar{x}_j)\right\} = \iint (x_i - \bar{x}_i)(x_j - \bar{x}_j)\, p_{\mathbf{x}_i,\mathbf{x}_j}(x_i, x_j)dx_i dx_j =$$

$$= \int (x_i - \bar{x}_i)\, p_{\mathbf{x}_i}(x_i)dx_i \int (x_j - \bar{x}_j) p_{\mathbf{x}_j}(x_j)dx_j = 0.$$

The converse proposition is generally not valid.

The above notions defined for a two-dimensional vector $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^{\mathrm{T}}$ are presented in Table A2.3.

Table A2.3

The main notions and stochastic characteristics for a two-dimensional random vector

Notions and characteristics	Definition
Probability distribution function	$F_{\mathbf{x}}(x) = \mathbf{Pr}(\mathbf{x}_1 < x_1, \mathbf{x}_2 < x_2);$
Relationship between PDF and CDF	$p_{\mathbf{x}}(x) = \dfrac{\partial^2 F(x)}{\partial x_1 \partial x_2},\ F_{\mathbf{x}}(x) = \int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2} p_{\mathbf{x}}(u)du_2 du_1$
Normalizing condition	$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p_{\mathbf{x}}(x)dx_1 dx_2 = 1$
PDF symmetry	$p_{\mathbf{x}}(x_1, x_2) = p_{\mathbf{x}}(x_2, x_1)$
Consistency conditions	$p_{\mathbf{x}_1}(x_1) = \int p_{\mathbf{x}}(x_1, x_2)dx_2$ $p_{\mathbf{x}_2}(x_2) = \int p_{\mathbf{x}}(x_1, x_2)dx_1$
Independence	$p_{\mathbf{x}}(x_1, x_2) = p_{\mathbf{x}_1}(x_1)\, p_{\mathbf{x}_2}(x_2)$
Noncorrelatedness	$E_{\mathbf{x}}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) = 0$
Probability for a random vector to belong to domain Ω	$\Pr(\mathbf{x} \in \Omega) = \int\limits_{\Omega} p_{\mathbf{x}}(x)dx_1 dx_2$
Mathematical expectation	$\bar{x}_i = \int\int x_i p_{\mathbf{x}}(x_1, x_2)dx_1 dx_2 = \int x_i p_{\mathbf{x}_i}(x_i)dx_i,$ $i = 1, 2$
Correlation coefficients	$P_{12} = P_{21} = \iint (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)\, p_{\mathbf{x}}(x_1, x_2)dx_1 dx_2$
Variances of components	$\sigma_i^2 = \iint (x_i - \bar{x}_i)^2\, p_{\mathbf{x}_i}(x_i)dx_i,\ i = 1, 2$
Covariance matrix	$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix},\ P_{ii} = \sigma_i^2,\ i = 1, 2,\ P = P^{\mathrm{T}} \geq 0$

A2.5. Gaussian random vectors

A **Gaussian random vector** is a vector for which PDF is defined as

$$p_{\mathbf{x}}(x) = N(x;\bar{x},P) = \frac{1}{(2\pi)^{n/2}(\det P)^{1/2}} \exp\left\{-0.5(x-\bar{x})^{\mathrm{T}} P^{-1}(x-\bar{x})\right\}. \quad \text{(A2.32)}$$

In this formula, $\bar{x}$ and P are a mathematical expectation and a covariance matrix which, as in the one-dimensional case, fully define the Gaussian PDF.

Vectors are called **jointly Gaussian** if their joint distribution density is Gaussian. Note that situations are possible in which each vector or random variable is Gaussian individually, but their joint density is not.

If jointly Gaussian vectors **x** and **y** are not correlated, i.e., $E_{\mathbf{x},\mathbf{y}}\{(x-\bar{x})(y-\bar{y})^{\mathrm{T}}\} = 0$, they are independent, and thus,

$$p_{\mathbf{x},\mathbf{y}}(x,y) = p_{\mathbf{y}}(y)\, p_{\mathbf{x}}(x).$$

In particular, this is easy to verify by the example when **x** and **y** are scalars since

$$p_{\mathbf{x},\mathbf{y}}(x,y) = N\left(\begin{matrix} x \\ y \end{matrix}; \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} P_{11} & 0 \\ 0 & P_{22} \end{bmatrix}\right) = \frac{1}{(2\pi)\ P_{11}^{1/2} P_{22}^{1/2}} \exp\left\{-\frac{(x-\bar{x})^2}{2P_{11}} - \frac{(y-\bar{y})^2}{2P_{22}}\right\} =$$

$$= \frac{1}{(2\pi)^{1/2} P_{11}^{1/2}} \exp\left\{-\frac{(x-\bar{x})^2}{2P_{11}}\right\} \frac{1}{(2\pi)^{1/2} P_{22}^{1/2}} \exp\left\{-\frac{(y-\bar{y})^2}{2P_{22}}\right\}.$$

As noted earlier, in a general case, this proposition is not true.

Let us analyze the form of a two-dimensional Gaussian PDF. Assume that the covariance matrix is nondiagonal, i.e.,

$$P = \begin{bmatrix} \sigma_1^2 & r^* \\ r^* & \sigma_2^2 \end{bmatrix}. \quad \text{(A2.33)}$$

Introducing the **normalized correlation coefficient** in the form

$$r = \frac{r^*}{\sigma_1 \sigma_2}, \quad \text{(A2.34)}$$

it is easy to see that

$$P^{-1} = \frac{1}{(1-r^2)} \begin{bmatrix} \dfrac{1}{\sigma_1^2} & -\dfrac{r}{\sigma_1\sigma_2} \\ -\dfrac{r}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2} \end{bmatrix}. \quad \text{(A2.35)}$$

Therefore, the PDF for a two-dimensional Gaussian vector can be written as

$$N(x;0,P) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2rx_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right\}. \quad \text{(A2.36)}$$

For different values of c, equation

$$g(x_1, x_2) = \frac{x_1^2}{\sigma_1^2} - \frac{2r x_1 x_2}{\sigma_1 \sigma_2} + \frac{x_2^2}{\sigma_2^2} = c^2 \qquad \text{(A2.37)}$$

defines ellipses. It should be noted that the axes of these ellipses rotate relative to the vertical axis at a certain angle. As an example, Fig. A2.3 shows isolines for r = 0.75.

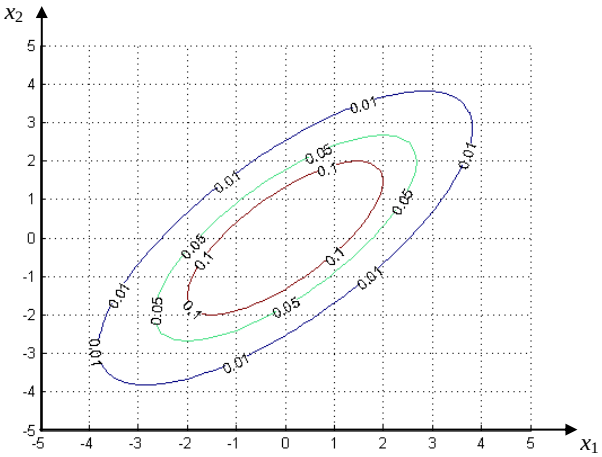


Fig. A2.3. PDF isolines for a zero-mean Gaussian vector for $\sigma_1 = \sigma_2 = 1$.

When navigation problems are solved on a plane, it is often assumed that the vehicle coordinates are a Gaussian random vector with a mathematical expectation at the point of its expected location. The uncertainty of the point location on the plane can be described with the use of the **equal probability ellipses** that were introduced above, in particular, the ellipse corresponding to equation (A2.32) for $c = 1$.

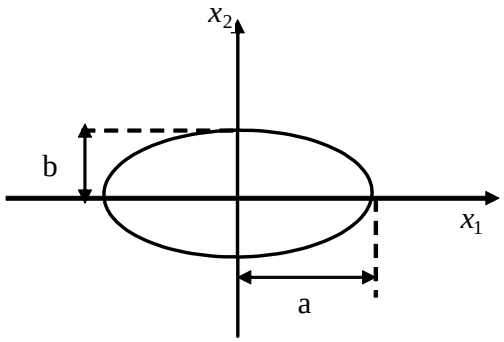


Fig. A2.4. Error ellipse for a two-dimensional Gaussian vector with independent components.

Since this ellipse intersects the axes at points coinciding with the values of the corresponding RMS, i.e., at $x_2 = 0, x_1 = \sigma_1$, and at $x_1 = 0, x_2 = \sigma_2$, it is called a **root-mean-square error ellipse**, or a **standard ellipse**. In navigation applications, it is described using the **ellipse parameters**: **major** a and **minor** b **semiaxes** and **directional angle** τ, which sets the orientation of the major semiaxis relative to the

axis x_2. These three parameters completely determine the covariance matrix of the two-dimensional Gaussian PDF [24]. Figure A2.4 shows a special case when $\sigma_2 = b$, $\sigma_1 = a$, $\tau = 90^\circ$, and, therefore,

$$P = \begin{bmatrix} a^2 & 0 \\ 0 & b^2 \end{bmatrix}, \tag{A2.38}$$

that is, the dimensions of the ellipse semiaxes determine the RMS values for each coordinate.

When estimating the accuracy of the vehicle position, it is very important to be able to characterize the uncertainty of position with a single value. It is for these purposes that the values of the probability for a point on the plane to appear in specific domain Ω are used. For a two-dimensional zero-mean Gaussian vector with PDF (A2.31), this probability is defined as

$$\mathbf{Pr}(\mathbf{x} \in \Omega) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \iint_{\Omega} \exp\left\{-\frac{1}{2(1-r^2)} g(x_1, x_2)\right\} dx_1 dx_2, \tag{A2.39}$$

where $g(x_1, x_2)$ is an equal probability ellipse (A2.32).

If domain Ω is represented by $g(x_1, x_2)$, then, passing on to polar coordinates, it can be shown that

$$\mathbf{Pr}(\mathbf{x} : g(x_1, x_2) \le c^2) = 1 - \exp\left\{-\frac{c^2}{2(1-r^2)}\right\}. \tag{A2.40}$$

For the case of independent random variables, at $\sigma_1 = \sigma_2 = \sigma$, the ellipse turns into a circle with a radius $R = c\sigma$ and, therefore, from (A2.35) it follows that the probability of finding a random vector within a circle with such a radius is determined by the Rayleigh distribution:

$$\mathbf{Pr}\left(\mathbf{x} : \frac{\sqrt{x_1^2 + x_2^2}}{\sigma^2} \le \frac{R^2}{\sigma^2}\right) = \mathbf{Pr}\left(\mathbf{x} : \sqrt{x_1^2 + x_2^2} \le R\right) = F(R) = 1 - \exp\left\{-\frac{R^2}{2\sigma^2}\right\}, \; R>0. \tag{A2.41}$$

The value R, which corresponds to 50% probability of finding the Gaussian random vector within a circle of a specified radius, i.e., the probability is 0.5, is called a **circular error probable (CEP)**, and correspondingly, the circle is called a **circle of equal probabilities**. In the case when the ellipse is a circle, i.e., with independent random variables and the same RMS $\sigma_1 = \sigma_2 = \sigma$, 50% probability of getting into the circle (Pr = 0.5) is achieved if its radius is equal to 1.177σ. For $R = 3.4\sigma$, **Pr** = 0.997. If this is not the case, then the radius of the circle at which 0.5 probability of getting into this circle is achieved, should be found using (A2.39).

Sometimes the notion of **a Distance Root-Mean-Square** error (DRMS) is used.

$$DRMS = \sqrt{\sigma_1^2 + \sigma_2^2}. \tag{A2.42}$$

Depending on the values of the covariance matrix or parameters of the ellipse, this value corresponds to 65–68% probability of getting into the circle with such a radius. Doubled radial RMS error (2DRMS) is also often used. It corresponds to the probability of getting into a circle with a radius equal to a doubled radial error. The exact value of the probability depends on the specific ratios of variances and the correlation coefficient, and its approximate value is determined as Pr = 0.95.

A3. Ordinary differential equations

An ordinary differential equation in the Cauchy form is an equation of the form [5, 8, 12, 25]:

$$\dot{x}(t) = F(t, x(t), u(t)), \tag{A3.1}$$

where $x(t) = (x_1(t),...x_n(t))^{\mathrm{T}}$ is an n-dimensional vector; $u(t) = (u_1(t)...u_p(t))^{\mathrm{T}}$ is a p-dimensional input vector, which can mean both control and disturbance; $F(.) = (F_1(.),...F_n(.))^{\mathrm{T}}$ is an n-dimensional, in the general case, nonlinear vector-function. Vector $x(t) = (x_1(t),...x_n(t))^{\mathrm{T}}$ is called a **phase vector** or **state vector**.

In the cases when $F(t, x(t), u(t)) = F(x(t), u(t))$ does not explicitly depend on time, the equation is called **stationary**. If this equation depends on time, it is called **nonstationary**.

When $u(t) \equiv 0$, the equation is called **homogeneous**; when $u(t) \neq 0$, it is called **inhomogeneous**.

Equation

$$\dot{x}(t) = F(t)x(t) + G(t)u(t), \tag{A3.2}$$

in which $F(t), G(t)$ are matrices of the corresponding dimensions, is called a **linear differential equation**.

If matrices $F(t), G(t)$ depend on time, the equation is called **a nonstationary linear differential equation**.

If matrices F, G do not depend on time, i.e.,

$$\dot{x}(t) = Fx(t) + Gu(t), \tag{A3.3}$$

the equation is called **a stationary linear differential equation**.

The solution of a differential equation is such a function of time $x(t)$, $x(t_0) = x_0$, the substitution of which into the original differential equation turns it into an identity. The value of function $x(t_0) = x_0$ at the initial moment is called the **initial condition**.

The fundamental, or transitional, matrix of the system of equations (A3.2) is a matrix that satisfies the equation

$$\frac{d\Phi(t, t_0)}{dt} = F(t)\Phi(t, t_0) \tag{A3.4}$$

with the **initial condition** of the form $\Phi(t_0, t_0) = E$.

The general solution to the differential equation (A3.2) satisfying the initial condition $x(t_0) = x_0$ has the form [5, 8]

$$x(t) = \Phi(t, t_0)x(t_0) + \int_{t_0}^{t} \Phi(t, \tau)G(\tau)u(\tau)\mathrm{d}\tau. \tag{A3.5}$$

The first summand represents the general solution of the linear homogeneous nonstationary differential equation

$$\dot{x}(t) = F(t)x(t).$$

The second summand is a **partial** solution of the differential equation (A3.2) at $x(t_0) = 0$. This solution is called **partial** because it corresponds to zero initial conditions and depends on a specific type of input action $u(t)$.

To show that (A3.5) is a solution to (A3.3), it is necessary to use the following rule [8]:

$$\frac{d}{dt}\int_0^t g(t,\tau)d\tau = g(t,t) + \int_0^t \frac{d}{dt} g(t,\tau)d\tau. \qquad \text{(A3.6)}$$

Differentiating (A3.8) and taking into account that

$$\dot{\Phi}(t,t_0) = F(t)\Phi(t,t_0),$$

we have

$$\dot{x}(t) = F(t)\Phi(t,t_0)x(t_0) + \Phi(t,t)\mathrm{G}(t)u(t) + \int_{t_0}^t F(t)\Phi(t,\tau)\mathrm{G}(\tau)u(\tau)\mathrm{d}\tau =$$

$$= F(t)x(t) + \mathrm{G}(t)u(t).$$

For stationary equations (A3.3), the fundamental matrix depends on the difference of the arguments and is a **matrix exponent** for F:

$$\Phi(t-t_0) = e^{F(t-t_0)} = \sum_{i=0}^{\infty} \frac{1}{i!} F^i (t-t_0)^i. \qquad \text{(A3.7)}$$

Since $Fe^{F(t-t_0)} = F\sum_{i=0}^{\infty} \frac{1}{i!} F^i (t-t_0)^i$, it is easy to see that $e^{F(t-t_0)}$ satisfies an equation of the type (A3.4).

Thus, the general solution of the stationary linear differential equation (A3.3) can be written as

$$x(t) = e^{F(t-t_0)}x(t_0) + \int_{t_0}^t e^{F(t-\tau)}\mathrm{G}u(\tau)\mathrm{d}\tau. \qquad \text{(A3.8)}$$

Consider a linear matrix differential equation

$$\dot{P}(t) = F(t)P(t) + P(t)F^T(t) + Q(t). \qquad \text{(A3.9)}$$

It is easy to verify that the general solution to this equation is matrix

$$P(t) = \Phi(t,t_0)P(t_0)\Phi^T(t,t_0) + \int_{t_0}^t \Phi(t,\tau)Q(\tau)\Phi^T(t,\tau)d\tau, \qquad \text{(A3.10)}$$

where $\Phi(t,t_0)$ is a fundamental matrix for equation $\dot{x}(t) = F(t)x(t)$.

Indeed, differentiating this relation, taking into account (A3.6), we derive

$$\dot{P}(t) = F(t)\Phi(t,t_0)P(t_0)\Phi^T(t,t_0) + \Phi(t,t_0)P(t_0)\Phi^T(t,t_0)F^T(t) +$$

$$Q(t) + \int_{t_0}^{t} F(t)\Phi(t,\tau)Q(\tau)\Phi^T(t,\tau)d\tau + \int_{t_0}^{t} \Phi(t,\tau)Q(\tau)\Phi^T(t,\tau)F(t)^T d\tau.$$

Taking into account (A3.10), we derive (A3.9). For the stationary case, (A3.10) has the form

$$P(t) = e^{F(t-t_0)}P(t_0)e^{F^T(t-t_0)} + \int_{t_0}^{t} e^{F(t-\tau)}Qe^{F^T(t-\tau)}d\tau\,. \qquad \text{(A3.11)}$$

References

1. **Autolas, A. (Ed.):** Mathematical System Theory, *The Influence of R.E.Kalman* (Springer-Verlag, 1991, 1st edn.)
2. **Bellman, P.** *Introduction to Matrix Analysis*, 2nd edn. SIAM, Dec. 1, 1997.
3. **Bergman, N.,** *Recursive Bayesian estimation: Navigation and tracking applications*, Linkoping University, Sweden, 1999.
4. **Brown, R.G., Hwang, P.Y.C.** *Introduction to Random Signals and Applied Kalman Filtering* (Wiley, 2012, 4th edn.)
5. **Bryson, A.E and Ho, Y.** *Applied Optimal Control: Optimization, Estimation and Control* (Revised Printing). John Wiley and Sons, New York, 1975.
6. **Doucet, A., Freitas, N., Gordon, N.,** *Sequential Monte Carlo Methods in Practice*, New York, NY, Springer New York, 2001.
7. **Emelyantsev, G.I. and Stepanov, A.P.,** *Integrirovannye inertsial'no-sputnikovye sistemy orientatsii i navigatsii* (Integrated Orientation and Navigation Systems), Peshekhonov V.G., Ed., St. Petersburg: Concern CSRI Elektropribor, 2016.
8. **Gelb, A., Kasper, J., Nash, R., Price, C., and Sutherland, A.,** (1994) *Applied Optimal Estimation*, M.I.T. Press, Cambridge, MA.
9. **Gibbs, Bruce P.** *Advanced Kalman Filtering, Least-Squares and Modeling: A Practical Handbook*, John Wiley&Sons, Inc., 2011.
10. **Global Positioning System**: *Theory and Applications*. Volume I, B. W. Parkinson and J. J. Spilker, Jr., (eds.), Washington, D.C.: AIAA, 1996.
11. **Golub Gene H. and Charles F. Van Loan**, *Matrix Computations*. Johns Hopkins University Press, 1996.
12. **Korn, G.A. and Korn, Th. M.**, *Mathematical handbook for scientists and engineers*, McGraw-Hill 1968.
13. **Grewal, M. S., and A. P. Andrews,** *Kalman Filtering: Theory and Practice*, 2nd edn., New York: Wiley, 2000
14. **Grewal, M.S., Andrews, A.P., Bartone, C.G**.: *Global Navigation Satellite Systems*, Inertial Navigation, and Integration (John Wiley & Sons, Inc. 2013, 3rd edn.)
15. **Groves Paul D.** *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*. Second Edition. Artech House GNSS Library, 2013.
16. **Kailath Thomas, Sayed Ali H., and Hassibi Babak** *Linear Estimation*, Prentice Hall, 2000.
17. **Kalman R.E.** A New Approach to Linear and Filtering Prediction Problems. *Trans. ASME, J. Basic Eng*., 1960, vol. 82 D.
18. **Kalman, R.E., Bucy, R.S.** New Results in Linear Filtering and Prediction Theory, *Trans. ASME, J. Basic Engineering*, Vol. 83, March 1961, pp. 95–108.

19. **Meditch, J.S.** *Stochastic Optimal Linear Estimation and Control*, McGraw Hill, New York, 1969.
20. **Sage E. and Mels J.** *Theory of Estimation and its Applications in Communications and Control*, McGraw Hill, New York, 1971.
21. **Sergienko, A.B.** *Tsifrovaya obrabotka signalov* (Digital Signal Processing), 2nd edn., St. Petersburg, 2006.
22. **Simo Särkkä**, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.
23. **Stepanov, O.A., Nosov, F.S., and Toropov, A.B**. About classification of map-aided navigation algorithms. *27th Saint Petersburg International Conference on Integrated Navigation Systems*, 2020.
24. **Stepanov, O.A.,** *Osnovy teorii otsenivaniya s prilozheniyami k zadacham obrabotki navigatsionnoi informatsii* (Fundamentals of the Estimation Theory with Applications to the Problems of Navigation Information Processing), Part 1, *Vvedenie v teoriyu otsenivaniya* (Introduction to the Estimation Theory), 3rd edn., St. Petersburg: Elektropribor, 2017.
25. **Stepanov, O.A.,** *Osnovy teorii otsenivaniya s prilozheniyami k zadacham obrabotki navigatsionnoi informatsii* (Fundamentals of the Estimation Theory with Applications to the Problems of Navigation Information Processing), Part 2, *Vvedenie v teoriyu fil'tratsii* (Introduction to the Filtering Theory), 3rd edn, St. Petersburg: Elektropribor, 2017.
26. **Stepanov, O.A.,** *Primenenie teorii nelineinoi fil'tratsii v zadachakh obrabotki navigatsionnoi informatsii* (Application of Nonlinear Filtering Theory for Navigation Data Processing), 3rd edn.,St. Petersburg: Elektropribor, 2003
27. **Stepanov, O.A.** Kalman filtering: Past and present. An outlook from Russia. (On the occasion of the 80th birthday of Rudolf Emil Kalman), *Gyroscopy and Navigation*, 2011, 2, (2), pp. 99–110.
28. **Stepanov, O.A.** Optimal and Sub-Optimal Filtering in Integrated Navigation Systems, in *Aerospace Navigation Systems*, Nebylov, A., Watson, J. (Eds.): (John Wiley & Sons, 2016, 1st edn.), pp. 244–298.
29. **Van Trees, H.L.,** *Detection, Estimation, and Modulation Theory*, Part 1, New York: Wiley, 1968**.**
30. **Van Trees, H.L. and Bell, K.L**., *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, San-Francisco: Wiley–IEEE Press, 2007.

Stepanov Oleg Andreevich

Estimation Algorithms for Navigation Data Processing

Tutorial

Original version
Editorial-Publishing Department of ITMO University
EP Department Head N. Gusarova
Signed to print
Order No
Printed circulation
Risograph printing

Editorial-Publishing Department of
ITMO University
197101, Saint Petersburg, Kronverkskiy pr., 49