MATHEMATICAL ASPECTS OF COMPUTER ENGINEERING

Advances in Science and Technology in the USSR

Design of Optimal Dynamic Analyzers: Mathematical Aspects of Wave Pattern Recognition

V P Belavkin and V P Maslov
Moscow Institute of Electronics and Mathematics
109028 Moscow USSR

Mir Publisher 1988

Contents

Preface

We give a review of the most important results on optimal tomography as mathematical wave-pattern recognition theory emerged in the 70's in connection with the problems of optimal estimation and hypothesis testing in quantum theory. The key problems in this theory is mathematical analysis and synthesis of the optimal dynamic analyzer discriminating between a given, discrete or continuous, family of pure or mixed *a priori* unknown wave patterns. Classical pattern recognition theory as a part of classical mathematical statistics cannot tackle such problems since it operates with given sample data and is not concerned how these date should be obtained from the physical wave states in an optimal way. In quantum theory this problem is sometimes referred as the problem of optimal measurement of an unknown quantum state, and is the main problem of the emerging mathematical theory of quantum statistics.

We develop the results of optimal quantum measurement theory, most of which belong to VPB [3]–[22], further into the direction of wave, rather than particle statistical estimation and hypothesis testing theory, with the aim to include not only quantum matter waves but also classical wave patterns like optical and acoustic waves. We apply the developed methods of this new pattern recognition theory to the problems of mathematical design of optimal wave analyzer discriminating the visual and sound patterns. We conclude that Hilbert space and operator methods developed in quantum theory are equally useful in the classical wave theory, as soon as the possible observations are restricted to only intensity distributions of waves, i.e. when the wave states are not the allowed observables, as they are not the observables of individual particles in the quantum theory. We will show that all characteristic attributes of quantum theory such as complementarity, entanglement or Heisenberg uncertainty relations are also attributes of the generalized wave pattern recognition theory.

Introduction

The problem of automatic recognition of coherent wave patterns is coming to the forefront in connection with one of the most important problems in building fifth generation computational systems, that are either of classical wave input and output or of quantum nature. The problem of efficiency for such wave pattern analyzers contains a number of mathematical subproblems, the most important of which have been studied in the seventeen's in connection with the quantum estimation and hypothesis testing theory. Without going into details of this and related quantum measurement theory these problems can be explained as the problems of (mathematical) design of an optimal analyzers of classical optical or acoustic waves, the devices occupying the center of the stage in the problem of automatic recognition of visual or sound patterns. An example of such a device is given in [45]: a receiver of acoustic waves $v(x - ct)$ whose idealized model is a point-like resonator (or cavity) capable of measuring the intensities of the vibrational modes excited by the waves. The modes are the natural vibrations of one or several standards placed at point $x = 0$ and are described by an orthonormal set of functions $\chi_k(t)$ on a given interval of observation $[0, T]$. A typical example of such a resonator is the spectrum analyzer, a device that measures the intensity distribution over the discrete frequencies $f_k = k/T$, $k \in N$, and can be represented by a selective filter of harmonic waves

$$v_k(x - ct) = 2\,\mathrm{Re}\,\phi^k \exp\{2\pi \mathrm{j} k(x - ct)/cT\}, \quad \mathrm{j} = \sqrt{-1},$$

in the output of which one can measure the positive numbers $\nu^k = |\phi^k|^2$ determined by the complex-valued amplitudes $\phi^k \in \mathbb{C}^1$. The spectrum selector described by the harmonic functions

$$\chi_k(t) = \exp\{-2\pi \mathrm{j} kt/T\},\ k = 0, 1, \ldots$$

which form an orthonormal set $\{\chi_k\}$ with respect to the scalar product

$$(\chi_i|\chi_k) = T^{-1} \int_0^T \chi_i(t)^* \chi_k(t)\,\mathrm{d}t,$$

is ideally suited for the discrimination of pure tones with multiple frequencies $\{f_i\}$, tones described by disjoint complex-valued amplitudes $\phi_i^k = 0$ at $i \neq k$ corresponding to the harmonic waves

$$v_i(x - ct) = 2\,\mathrm{Re}\,\varphi_i(t - x/c),$$

where $\varphi_i(t) = \sum_{k=0}^{\infty} \phi_i^k \chi_k(t)$. To establish which of the tones in $\{\varphi_i\}$ with different frequencies and nonzero intensities $\nu_i = |\phi_i^k|^2 \neq 0$ at $i = k$ is actually detected by such a receiver, it is sufficient to find the number i of the excited standard tuned to one of the harmonic modes in $\{\chi_i\}$ corresponding to the set $\{\varphi_i\} : \chi_i(t) = \varphi_i(t)/\|\varphi_i\|$. The vibrational energy of such a standard will coincide with the intensity $|(\chi_i \mid \varphi_i)|^2 = \|\varphi_i\|^2$ of the detected signal φ_i, while the other standards remain unexcited: $(\chi_k \mid \varphi_i) = 0$ at $k \neq i$. A segment of human speech of duration T containing a finished sentence consists, however, not of a single pure tone but, generally, of an infinitude of pure tones of different amplitudes and frequencies (the frequencies may be assumed to be multiples of $1/T$ at a fixed interval T of a single reception act). Unharmonic signals, described by spectral amplitudes $\varphi_i = [\phi_i^k]_{k=0}^{\infty}$ or, in the temporal representation, by the analytic signals

$$\varphi_i(t) = \sum_{k=0}^{\infty} \phi_i^k \exp\{-2\pi \mathrm{j} kt/T\},$$

may be indistinguishable in spectral measurements, even if they are orthogonal. For example, if the $\varphi_i(t)$, $i = 1, \ldots, m$ are disjoint pulses obtained through the shift by T of the pulsed signal $\varphi_0(t)$ of length $\Delta t = T/m$, these pulses have corresponding to them the orthogonal spectral amplitudes $\varphi_i^k = \phi_0^k \exp\{2\pi \mathrm{j} ik/m\}$ with the same intensity distributions $\nu_i^k = |\phi_i^k|^2 = |\phi_0^k|^2$, $i = 1, \ldots, m$, $k \in N$.

Orthogonal sound signals $\{\varphi_i\}$ on $[0, T]$ may be identified in a similar manner by selective filters matched with the signal modes $\chi_i(t) = \varphi_i(t)/\|\varphi_i\|$ that measure the intensity distribution $\nu^k = (\chi_k \mid \varphi_i)|^2$ in the modes $\{\chi_i\}$, a distribution that has a different form for different values of i, namely, $\nu_i^k = \|\varphi_i\|^2 \delta_i^k$. However, different sentences in human speech correspond ordinarily to nonorthogonal sound signals $\varphi_i(t)$, which from the viewpoint of their meaning are identified if they are collinear, that is, differ only in the total energy $\|\varphi_i\|^2 = \int |\varphi_i(t)|^2 \mathrm{d}t$. For the recognition of nonorthogonal signals $\{\varphi_i\}_{i=1}^m$ one cannot employ matched filtration since the filters described by the nonorthogonal modes $\chi_i = \varphi_i/\|\varphi_i\|$ are noncommutative and, hence, cannot be matched in a single selector; otherwise, the total measured intensity $\sum_{k=1}^m |(\chi_k \mid \chi_i)|^2$, would exceed the total energy $\|\varphi_i\|^2$ of the received signal φ_i if $(\varphi_i \mid \varphi_k) \neq 0$ at least for one $k \neq i$. Thus, we have an indefinite situation, formally similar to the incompatibility of noncommutative quantum mechanical observables, a situation arising from Bohr's complementarity principle [23] and Heisenberg's uncertainty relation [46]. Typical examples of noncommutative filters are the frequency and temporal filters, which are incompatible, just as position and momentum measurements are incompatible in a quantum mechanical system. So which of the disjoint selectors described by orthonormal sets of modes $\{\chi_k\}$ must we employ to discern the nonorthogonal sound signals from a given set $\{\varphi_i\}$? It is natural to look for the answer to this nontrivial question in the form of a solution to an optimization problem by selecting a criterion of discernment quality such that the optimal selector does not depend on the gauge transformation $\varphi_i \to \lambda\varphi_i$ for every complex-valued λ. The latter condition is satisfied by the criterion of the maximum of the total intensity $\sum_{i=1}^{m} |(\chi_i \mid \varphi_i)|^2$ of the true received amplitudes or by the criterion of the minimum of the lost intensity $\sum_{i \neq k} |(\chi_i \mid \varphi_i)|^2$. The corresponding extremal problem for arbitrary nonorthogonal amplitudes $\{\varphi_k\}$ describing quantum mechanical states normalized to prior probabilities was first studied in the general form in [14], [12], [16]. Particular solution of this problem for the case of two nonorthogonal amplitudes $\{\varphi_0, \varphi_1\}$ were obtained in [2], [34], while the case of several linearly independent amplitudes $\{\varphi_i\}$ was also considered in [40], [51].

The above-noted analogy between optimal recognition of sound signals and discernment of quantum mechanical states suggested the possibility of constructing a wave theory of noncommutative measurements within the scope of which one could solve more general problems of testing wave hypotheses for estimating the wave parameters. From the formal viewpoint this theory generalizes the quantum theory of optimal measurements, hypothesis testing, and estimation of parameters [30], [36], while actually it carries the mathematical methods and ideas developed in [3]-[22] into the new, practically more realistic, field of applications. A short author's review of optimal processing of quantum signals is given in [15]. (Additional literature on the quantum theory of detection, hypothesis testing, and parameter estimation can be found in the references cited in [30], [36].)

In the present text we give a systematic description of the wave theory of representation and measurement based on analogies with quantum mechanics. This theory is then employed to solve the problems of detection, discrimination, identification, and estimation of the parameters of sound signals and visual patterns within the framework of the noncommutative theory of testing wave hypotheses developed here. The idea of employing the methods of quantum mechanics for discerning wave patterns emerged at the beginning of the 1970s, when a seminar on quantum mechanics and pattern recognition was organized in the Physics Department of Moscow State University. The seminar was directed by V.P. Maslov, along with Yu.P. Pyt'ev, and the author, V.P. Belavkin, attended it. Interest in wave tomography as reconstruction of the complete wave field rather than only the energy illumination of the image in a certain plane was stimulated by the rapid development of holography, which was invented by Gabor [25], and then underwent a revival [24] when coherent sources of light, or lasers, were created. The emerging optimization problems of discerning wave fronts are similar to the problems of discerning quantum mechanical states and cannot be solved by classical methods [48] of pattern recognition since it is impossible to register directly and exactly the phase and amplitude of a wave field by measuring the energy parameters. A detailed study of these problems at the time showed [44] that the then existing quasiclassical methods of solving quantum mechanical problems were also inadequate, and the solution had to be postponed until a consistent noncommutative theory of measurements was developed in the then rapidly advancing field of quantum theory.

Some particular problems of optimal processing optical wave signals, such as those of optical localization [42], detection, and discrimination of two signals from closely positioned sources of coherent radiation [33], [32], have been already well studied by methods of quantum statistical and nonlinear optics [26], [41], [1].

In this book, in addition to discussing the noncommutative theory of measurements common for quantum states and classical sound signals and coherent optical waves, we provide solutions to a number of classical wave recognition problems from the quantum optimal measurement theory (obtained by the first author, V.P. Belavkin, in the 70's, when solving similar problems for quantum signals). Content and commentaries to the list of literature are presented in brief summaries at the beginning of each section. Similar summaries are given at the beginning of each subsection. The subsections are written as a series of articles of increasing complexity so that each can be read independently, although the best way to understand the material is to carefully read the articles in the order given.

Chapter 1

Representation and Measurement of Waves

In this section we discuss the mathematical apparatus of the wave theory of representation and measurement of sound and visual patterns. Along with pure wave patterns, which are described by coherent signals and fields, we also consider the representation of mixed patterns, which are described by partially coherent and incoherent signals and fields. In addition to the spatial-frequency (coordinate) and wave-temporal (momentum) representations we introduce the joint canonical representation, in which the pure and mixed patterns are described by entire functions and kernels in a phase complex space. We develop the mathematical theory of idea filters and quasifilters, disjoint selectors, and quasiselectors, which describe ordinary, successive, and indirect measurements of wave-pattern intensity distributions. Using this theory as a basis, we analyze coordinate and momenta measurements as well as quasimeasurements of joint coordinate-momentum distributions. The mathematical tools used here are in many respects similar to those used in the quantum theory of representations and measurements [46], which recently received a new impetus in connection with problems of quantum states recognition [14], [12], [16], [2], [34], [40], [51]; however, we give a wave rather than a statistical interpretation of the apparatus in accordance with the application considered here.

1.1 Mathematical Description of Wave Patterns

In this section we describe three basic types of representation of pure and mixed wave patterns; the coordinate, or spatial-frequency, the momentum, or wave-temporal, and the canonical, in which the wave patterns are represented by holomorphic amplitudes in the complex coordinate-momentum plane. The third representation, which emerged in quantum optics [26], proves useful in an analysis of the frequency-temporal structure of sound and visual patterns and in holography in an analysis of the spatial-temporal structure of such patterns.

1.1.1 Wave Sound and Visual Patterns

The sound and visual patterns considered here are commonly described by wave amplitudes $v(t, \mathbf{q})$ that are real-valued function of time t and coordinates $\mathbf{q}$ in a spatial-temporal region accessible for measurement. Although the simplest wave equations describing the behaviour of such physical fields are linear in amplitudes $v(t, \mathbf{q})$, for purposes of measurement of sound and visual patterns the most interesting are functions quadratic in v that describe the distribution of sound on the standards of the dynamic analyzer or of light on photodetectors, a distribution that in spatial-temporal measurements is described by the intensity function $v^2(t, \mathbf{q})/2$. More useful information is proved not by the intensities of a sound or visual pattern at points $(t, \mathbf{q})$ but by the distribution of the sound intensity at frequency f (a spatial-frequency distribution), as is common in an analysis of colour patterns. Such a distribution is determined by the intensity function

$$\iota(f, \mathbf{q}) = |\varphi(f, \mathbf{q})|^2, \quad f \geq 0, \tag{1.1}$$

with φ the complex-valued spectral amplitudes,

$$\varphi(f, \mathbf{q}) = \int_{-\infty}^{\infty} v(t, \mathbf{q}) \mathrm{e}^{2\pi \mathrm{j} f t} \, \mathrm{d}t,$$

used to represent the wave field $v(t, \mathbf{q})$ in the form of a linear combination of harmonic oscillations at each point $\mathbf{q}$:

$$v(t, \mathbf{q}) = 2 \operatorname{Re} \int_0^{\infty} \varphi(f, \mathbf{q}) \mathrm{e}^{-2\pi \mathrm{j} f t} \, \mathrm{d}f.$$

Bearing in mind the well-known advantage of employing complex-valued amplitudes, in what follows we consider complex-valued signals $\varphi(x) = \varphi(f, \mathbf{q}) \equiv \varphi(q)$, with $x = f(f, \mathbf{q}) \equiv q$, assuming that in a given spatial-frequency region of measurements Ω they possess a finite total intensity

$$I(\varphi) = \int_{\Omega} |\varphi(q)|^2 \, \mathrm{d}q. \tag{1.2}$$

When registering sound signals for which the spatial region of measurements is much smaller than the characteristic length of the sound wave, we can take for Ω a one-dimensional region, which is usually determined by a positive-frequency pass band $\Phi \in \mathbb{R}_+$ of the dynamic analyzer, assuming that $x = f$ at the point of its localization $\mathbf{q} = 0$; recognizing visual patterns usually requires only a three-dimensional region $\Omega = \Phi \times S$, where Φ is the optical frequency band and S the surface on which the pattern is localized; for static patterns $x = \mathbf{q} \quad (t = 0)$.

From the standpoint of physics the admissible amplitudes are smooth amplitudes $\varphi(q)$, $q \in \Omega$, with compact supports inside a $(d+1)$-dimensional region $\Omega \subseteq \mathbb{R}^{d+1}$ or, if Ω is noncompact, amplitudes that fall off rapidly at infinity together with all their derivatives. Such amplitudes generate a Hilbert space $\mathcal{H} = L^2(\Omega)$ of amplitudes $\chi(q)$ of finite intensity $\|\chi\|^2 < \infty$:

$$(\varphi \mid \chi) = \int_{\Omega} \varphi(q)^* \chi(q) \, \mathrm{d}q. \tag{1.3}$$

Generally the set $\mathcal{D}$ of basic amplitudes φ can form an arbitrary complex-values space with a positive Hermitian form $I(\varphi) = (\varphi \mid \varphi)$ that is invariant with respect to complex conjugation $\varphi^* (q) = \varphi(q)^*$. This form defines a finite intensity, $I(\varphi) \neq 0$, for any nonzero φ. We denote the completion of this space in norm $\|\varphi\| = I^{1/2}(\varphi)$ by $\mathcal{H}$ and consider it to be a Hilbert space equipped with an isometric involution $\chi \mapsto \chi^*$ with respect to the scalar produce (1.3), which is linear in the second argument $\varphi \in \mathcal{H}$.

The use of complex-valued amplitudes not only considerably simplifies the formulas for calculating the observed distributions of the fields but also makes it possible to employ analogies from quantum theory. Specifically, as in quantum theory, complex-valued amplitudes differing in a phase factor must be assumed "equal" since they lead to the same intensities defined by Hermitian forms of φ. Note that in the quantum description of optical and sound signals we usually take the mean number of the corresponding quanta (photons and phonons) as the intensity functions for light and sound, respectively. These quantities are determined by the same Hermitian forms of the complex-valued amplitudes φ as in the classical mode of description, provided that the quantum mechanical states are coherent [26], [41], that is, are described by Poisson probability amplitudes $|\varphi\rangle$. Thus, restricting ourselves to intensity measurements, we postulate that only distributions of quanta are observable, while the only characteristics of signals φ of interest to physics are those obtained as a result of measurement of such distributions.

1.1.2 Momentum Representation

In problems dealing with the recognition of moving patterns what may be of interest is not the spatial-frequency intensity distribution (1.1) but the momentum-temporal distribution described by the function

$$\widetilde{\iota}(t, \mathbf{p}) = |\widetilde{\varphi}(t, \mathbf{p})|^2, \quad (t, \mathbf{p}) \in \mathbb{R}^{d+1}, \tag{1.4}$$

where $\widetilde{\varphi}(t, \mathbf{p})$ is the involution Fourier transform,

$$\widetilde{\varphi}(t, \mathbf{p}) = \iint \varphi(f, \mathbf{q})^* \mathrm{e}^{2\pi \mathrm{j}(tf + \mathbf{p}\cdot\mathbf{q})}\, \mathrm{d}f\, \mathrm{d}\mathbf{q}. \tag{1.5}$$

We introduce the notation $x = (t, \mathbf{p}) \equiv p$. The representation of amplitudes φ in terms of the functions $\widetilde{\varphi}(x) = \widetilde{\varphi}(t, \mathbf{p}) \equiv \widetilde{\varphi}(p)$ is called the momentum representation.[1] Note that this representation differs from the common one by complex conjugation, $\widetilde{\varphi}^* = \widetilde{\varphi}^*$, but this difference is "unobservable" from the viewpoint of measuring intensities described by Hermitian forms (1.1) and (1.4), which are invariant under such conjugation.

Allowing for Plancherel's equality

$$\int |\widetilde{\varphi}(p)|^2 \mathrm{d}p = \int |\varphi(q)|^2 \mathrm{d}q, \tag{1.6}$$

[1] Thanks to the introduction of the involution $\varphi \mapsto \varphi^*$ in transformation (1.5), the inverse transformation to the coordinate representation is carried out by the same formula (1.5), $\varphi = \widetilde{\widetilde{\varphi}}$, which can be extended onto generalized amplitudes χ in the standard manner.

we find that the total intensity described by the distribution function (1.4) coincides with the total intensity (1.2) for any amplitude φ with support in Ω:

$$\int \widetilde{\iota}(p)\,\mathrm{d}p = I(\varphi) = \int \iota(q)\,\mathrm{d}q. \tag{1.7}$$

In what follows we consider the values of $\iota(q)$ and $\widetilde{\iota}(p)$ in (1.1) and (1.4) as being functions of φ, denoting by $\widetilde{\iota}(\varphi, q)$ and $\widetilde{\iota}(\varphi, p)$ the functionals connected by involutions $\widetilde{\iota}(\varphi, p) = \iota(\widetilde{\varphi}, p)$ and $\widetilde{\widetilde{\iota}}(\varphi, q) = \iota(\varphi, q)$, respectively. Note that the Fourier transformation $\iota(q) \mapsto \widetilde{\iota}(p)$ of the Hermitian functional $\iota(\varphi, q) = |\varphi(q)|^2$ cannot be reduced to the Fourier transformation of its value $\iota(\varphi)$ as a function of q. More than that, measuring the spatial distribution $\widetilde{\iota}(q)$ for a single value of φ does not generally make it possible in any way to calculate the corresponding distribution $\widetilde{\iota}(p)$, and vice versa. Nevertheless, these distributions satisfy certain relationships, the simplest of which are (1.7) and the inequality

$$\int (p-\overline{p})^2 \widetilde{\iota}(p)\,\mathrm{d}p \int (q-\overline{q})^2 \iota(q)\,\mathrm{d}q \geq \tfrac{1}{(4\pi)^2} I^2(\varphi) \tag{1.8}$$

(where $\overline{p} = \int p\widetilde{\iota}(p)\,\mathrm{d}p/I(\varphi)$ and $\overline{q} = \int q\iota(q)\,\mathrm{d}q/I(\varphi)$ for each of the components $p = p_k$ and $q = q_k$, $k = 0, \ldots, d$), which is known as the uncertainty relation. Using the commutation relations

$$\widehat{p}\widehat{q} - \widehat{q}\widehat{p} = (2\pi \mathrm{j})^{-1}\widehat{1} \tag{1.9}$$

For each pair of operators $\widehat{q}, \widehat{p}$ in the p-representation, with $\widehat{p} = p - \overline{p}$ and $\widehat{q} = \partial/\partial(2\pi \mathrm{j} p) = \overline{q}$, we can easily arrive at (1.8) as a corollary of Schwarz's inequality

$$\|\widehat{p}\widetilde{\varphi}\|\,\|\widehat{q}\widetilde{\varphi}\| \geq |(\widehat{p}\widetilde{\varphi} \mid \widehat{q}\widetilde{\varphi})| \geq |\mathrm{Im}(\widehat{p}\widetilde{\varphi} \mid \widehat{q}\widetilde{\varphi})| = \tfrac{1}{2\pi}\|\widetilde{\varphi}\|^2. \tag{1.10}$$

Indeed, according to definition (1.4) we have

$$\int (p-\overline{p})^2 \widetilde{\iota}(p)\,\mathrm{d}p = \int |(p-\overline{p})\widetilde{\varphi}(p)|^2\,\mathrm{d}p = \|\widehat{p}\widetilde{\varphi}\|^2$$

and, similarly, allowing for (1.6), for the Fourier transform $\widehat{p}\widetilde{\varphi}$ of the function $(q-\overline{q})\varphi(q)$ we obtain from (1.1)

$$\int (q-\overline{q})^2 \iota(q)\,\mathrm{d}q = \int |(q-\overline{q})\varphi(q)|^2\,\mathrm{d}q = \|\widehat{q}\widetilde{\varphi}\|^2.$$

Thus, inequality (1.8) is equivalent to (1.10), where $\|\widetilde{\varphi}\|^2 = I(\varphi)$. For nonzero amplitudes φ this inequality is usually written as

$$\sigma_p \sigma_q \geq 1/4\pi, \tag{1.11}$$

where σ_p and σ_q are the standard deviations,

$$\begin{aligned} \sigma_p^2 &= \int (p-\overline{p})^2 \widetilde{\iota}(p)\,\mathrm{d}p/I(\varphi), \\ \sigma_q^2 &= \int (q-\overline{q})^2 \iota(q)\,\mathrm{d}q/I(\varphi), \end{aligned} \tag{1.12}$$

of momentum p and coordinate q in the wave packet φ from their mean values $\overline{p}$ and $\overline{q}$. In this form (1.11) is similar to the quantum mechanical Heisenberg uncertainty relation; however, here the standard deviations (1.12) have no statistical meaning but characterize the extent to which the intensity distributions are localized in the coordinate and momentum spaces. The lower bound $1/(4\pi)$ in this relation is achieved only in the case of an unbounded region $\Omega = \mathbb{R}^{d+1}$ for the Gaussian amplitudes

$$\varphi(q) = C_q \exp\{2\pi \mathrm{j}(q - \frac{\overline{q}}{2})\overline{q} - \frac{|q-\overline{q}|^2}{4\sigma_q^2}\}. \tag{1.13}$$

These amplitudes, with the normalization constants $C_q = 1/(2\pi\sigma_p^2)^{(d+1)/4}$, have a similar form in the p-representation:

$$\widetilde{\varphi}(p) = C_p \exp\{2\pi \mathrm{j}(p - \frac{\overline{p}}{2})\overline{q} - \frac{|p-\overline{p}|^2}{4\sigma_p^2}\},$$

with $C_p = 1/(2\pi\sigma_p^2)^{(d+1)/4}$ and $\sigma_p\sigma_q = 1/(4\pi)$, are called standard canonical (Poisson) amplitudes and are denoted by $\psi_\alpha = |\alpha)$, with

$$\alpha = \frac{1}{2}(\frac{\overline{q}}{\sigma_q} + \mathrm{j}\frac{\overline{p}}{\sigma_p})$$

if σ_q and σ_p are fixed. Note that for $\alpha \neq \alpha'$ such amplitudes are nonorthogonal:

$$(\alpha \mid \alpha') = \exp\{-\frac{1}{2}|\alpha'|^2 + \alpha'\alpha^\dagger - \frac{1}{2}|\alpha|^2\}, \tag{1.14}$$

with $\alpha'\alpha^\dagger$ defined as the scalar product $\sum_{i=0}^{d} \alpha_i^* \alpha_i'$.

1.1.3 Mixed Signals

Due to limits in present-day technology, only a fraction of the information on the intensity distribution of amplitude φ in this or another region can usually be obtained when analyzing sound and visual patterns. For instance, in sound pattern recognition the common method is to use only the frequency or temporal distribution obtain through integration

$$\iota(f) = \int |\varphi(f,\mathbf{q})|^2 \,\mathrm{d}\mathbf{q}, \quad \widetilde{\iota}(t) = \int |\widetilde{\varphi}(t,\mathbf{p})|^2 \,\mathrm{d}\mathbf{p} \tag{1.15}$$

of distributions (1.1) and (1.4) over the spatial or wave region of measurement. In visual pattern recognition often only black and white patterns are considered. These are obtained as the result of mixing

$$\iota(\mathbf{q}) = \int |\varphi(f,\mathbf{q})|^2 \,\mathrm{d}f, \quad \widetilde{\iota}(\mathbf{p}) = \int |\widetilde{\varphi}(t,\mathbf{p})|^2 \,\mathrm{d}t \tag{1.16}$$

of the appropriate colour patterns in the spatial or wave region of measurement. To obtain such incomplete distributions there is no need to provide a total description of the signal by amplitude $\varphi(f, \mathbf{q})$. For instance, in describing sound it is sufficient to specify only the Hermitian kernel

$$S(f', f) = \int \varphi(f', \mathbf{q})\varphi^*(f, \mathbf{q})\, \mathrm{d}q,$$

for which $\iota(f) = S(f, f)$ and $\widetilde{\iota}(t) = \widetilde{S}(t, t)$, where

$$\widetilde{S}(t', t) = \int \mathrm{e}^{2\pi \mathrm{j}(ft' - f't)} S(f', f)\, \mathrm{d}f'\, \mathrm{d}f.$$

Monochrome patterns are defined by a similar kernel $S(\mathbf{q}', \mathbf{q})$, with $\iota(\mathbf{q}) = S(\mathbf{q}, \mathbf{q})$ and $\widetilde{\iota}(\mathbf{p}) = \widetilde{S}(\mathbf{p}, \mathbf{p})$. Having in mind the possibility of such mixing, we will describe signals in an abridged manner by nonnegative definite operators of intensity density, S, with kernels $S(q, q')$ that have a nonzero trace

$$\mathrm{Tr} S = \int_\Omega S(q, q)\, \mathrm{d}q \equiv \langle S, I \rangle\,, \tag{1.17}$$

which determines the total intensity, $\iota(S) = \langle S, I \rangle$, of the signal in Ω. To each amplitude $\psi(q)$ we assign a one-dimensional operator $S = |\psi)(\psi|$ with a kernel

$$S(q', q) = \psi(q')\psi^*(q), \tag{1.18}$$

which defines the amplitude $\psi(q)$ to within a nonessential phase factor $\mathrm{e}^{\mathrm{j}\theta}$. The diagonal values $\iota(q) = S(q, q)$ describe the coordinate distribution of the intensity of such a signal, while the momentum distribution is described by the diagonal values $\widetilde{\iota}(p) = \widetilde{S}(p, p)$ of the involution Fourier transform

$$\widetilde{S}(p', p) = \int \mathrm{e}^{2\pi \mathrm{j}(p' q^\intercal - q' p^\intercal)} S(q', q)\, \mathrm{d}q'\, \mathrm{d}q. \tag{1.19}$$

Each such kernel can be obtained as a result of mixing

$$S(q', q) = \int \psi_\alpha(q')\psi_\alpha^*(q)\nu(\mathrm{d}\alpha) \tag{1.20}$$

of the one-dimensional kernels corresponding to the normalized amplitudes $\{\psi_\alpha\}$, $\|\psi_\alpha\| = 1$, parameterized by a space A with a nonzero positive measure ν whose mass determines the total intensity $\langle S, I \rangle = \nu(A)$. For example, the kernel $S(\mathbf{q}', \mathbf{q})$ corresponding to a monochrome pattern generated by amplitude $\varphi(f, \mathbf{q})$ can be written in the form (1.20) for

$$\psi_f(\mathbf{q}) = \varphi(f, \mathbf{q})/\iota^{1/2}(f)$$

on the set A of frequencies Φ equipped with a nonzero measure $\nu(\mathrm{d}f) = \iota(f)\, \mathrm{d}f$.

In the case of an arbitrary Hilbert space $\mathcal{H}$, mixed signals are described by density operators S obtained as a result of weak integration

$$S = \int |\psi_\alpha)(\psi_a| \nu(d\alpha) \tag{1.21}$$

of one-dimensional density operators $S_\psi = |\psi)(\psi|$,

$$|\psi)(\psi| : \chi \in \mathcal{H} \mapsto \psi(\psi \mid \chi) = (\psi \mid \chi)\psi, \tag{1.22}$$

corresponding to the normalized values $\psi_\alpha \in \mathcal{H}, \alpha \in A$, of the vector function $\alpha \mapsto \psi_\alpha$. The operators (1.21) are kernel-positive and have a finite trace $\operatorname{Tr} S = \nu(A)$, with each operator $S : \mathcal{H} \mapsto \mathcal{H}$ being represented in the form (1.21) via, say, the spectral decomposition

$$S = \sum_i |\psi_i)(\psi_i|\nu_i. \tag{1.23}$$

Here $\{\psi_i\}$ is the maximal orthogonal set of normalized eigenvectors $\psi_i \in \mathcal{H}$ corresponding to zero eigenvalues $\nu_i : S\psi_i = \nu_i\psi_i$, which determine the trace $\operatorname{Tr} S = \sum_i \nu_i$.

1.1.4 Gaussian Signals

As an example let us consider the important class of mixed canonical signals defined by the integration

$$S = \int |\alpha)(\alpha|\nu(d\xi\, d\eta), \quad \alpha \in \mathbb{C}^{d+1}, \tag{1.24}$$

of canonical projectors corresponding to the amplitudes $\psi_{\xi\eta} = |\alpha)$, which in the coordinate representation have the following general Gaussian form (c.f. (1.13))

$$\psi_{\xi\eta}(q) = C \exp\{2\pi \mathrm{j}(q - \tfrac{1}{2}\xi)\eta^\intercal - \tfrac{1}{2}(q-\xi)\omega(q-\xi)^\intercal\}. \tag{1.25}$$

Here $\omega = \omega^\intercal$ is a symmetric complex-valued $(d+1)$-by-$(d+1)$ matrix with a positive definite real part

$$\omega + \omega^* = 2\pi(\upsilon^\dagger\upsilon)^{-1},$$

$\xi = \sqrt{2/\pi}\operatorname{Re}\alpha\upsilon$ and $\eta = \sqrt{2/\pi}\operatorname{Re}\mathrm{j}\alpha^\dagger\widetilde{\upsilon}$, with $\widetilde{\upsilon} = \upsilon^\dagger\omega/(2\pi)$, are $(d+1)$-dimensional rows, $\xi^\intercal$ and $\eta^\intercal$ are the corresponding columns,

$$|C|^2 = \det(\upsilon^\dagger\upsilon)^{-1} = |\upsilon|^{-2}$$

is the normalization constant, and $\upsilon^{*\intercal} = \upsilon^\dagger = \upsilon^{\intercal *}$ is the Hermitian conjugate of matrix υ.

Let $\zeta = (\xi, \eta)$ be a $2(d+1)$-dimensional row and $\nu(d\zeta) = \nu(d\xi\, d\eta)$ a Gaussian measure on $A = \mathbb{R}^{2(d+1)}$ normalized to certain number $J < \infty$ (the Gaussian intensity) and described (for J positive) by the following moments:

$$\overline{\xi} = J^{-1}\int \xi\nu(d\zeta) = \lambda, \quad \overline{\eta} = J^{-1}\int \eta\nu(d\zeta) = \varkappa, \tag{1.26}$$

$$\begin{bmatrix} \overline{\xi^\intercal\xi} & \overline{\xi^\intercal\eta} \\ \overline{\eta^\intercal\xi} & \overline{\eta^\intercal\eta} \end{bmatrix} = J^{-1}\int \zeta^\intercal\zeta\nu(d\zeta) = \begin{bmatrix} \lambda^\intercal\lambda & \lambda^\intercal\varkappa \\ \varkappa^\intercal\lambda & \varkappa^\intercal\varkappa \end{bmatrix} + \sigma_{\zeta\zeta}, \tag{1.27}$$

where $\sigma_{\xi\xi} = \begin{bmatrix} \sigma_{\xi\xi} & \sigma_{\xi\eta} \\ \sigma_{\eta\xi} & \sigma_{\eta\eta} \end{bmatrix}$ is a nonnegative define $2(d+1)$-by-$2(d+1)$ matrix. The signals that correspond to such a density operator (1.24) are characterized by the following first moments:

$$\overline{q} = J^{-1} \int Q(\psi_\zeta)\nu(d\zeta) = \lambda,\ \overline{p} = J^{-1} \int P(\psi_\zeta)\nu(d\zeta) = \varkappa, \tag{1.28}$$

$$\begin{aligned}
\overline{q^\intercal q} &= J^{-1} \int (Q^\intercal \psi_\zeta \mid Q\psi_\zeta)\nu(d\zeta) = \overline{\xi^\intercal \xi} + (\omega + \omega^*)^{-1}, \\
\overline{q^\intercal p} &= J^{-1} \int (Q^\intercal \psi_\zeta \mid P\psi_\zeta)\nu(d\zeta) = \overline{\xi^\intercal \eta} + j\upsilon^\intercal \widetilde{\upsilon}/(2\pi), \\
\overline{p^\intercal q} &= J^{-1} \int (P^\intercal \psi_\zeta \mid Q\psi_\zeta)\nu(d\zeta) = \overline{\eta^\intercal \xi} - j\widetilde{\upsilon}^\dagger \upsilon^*(2\pi), \\
\overline{p^\intercal p} &= J^{-1} \int (P^\intercal \psi_\zeta \mid P\psi_\zeta)\nu(d\zeta) = \overline{\eta^\intercal \eta} + (\widetilde{\omega} + \widetilde{\omega}^*)^{-1},
\end{aligned} \tag{1.29}$$

where $\widetilde{\omega}/(2\pi) = 2\pi/\omega^*$, Q and P are the rows of operators of position Q_k and momentum P_k defined in the q-representation via multiplication by q_k and differentiation with respect to q_k, or $(2\pi\mathrm{j})^{-1}\partial/\partial q_k$, and we have allowed for the fact that

$$Q(\psi_\zeta) = (\psi_\xi \mid Q\psi_\zeta) = \xi,\ P(\psi_\zeta) = (\psi_\zeta | P\psi_\zeta) = \eta,$$

$$\begin{bmatrix} (\widehat{q}^\intercal \psi_\zeta \mid \widehat{q}\psi_\zeta) & (\widehat{q}^\intercal \psi_\zeta \mid \widehat{p}\psi_\zeta) \\ (\widehat{p}^\intercal \psi_\zeta \mid \widehat{q}^\intercal \psi_\zeta) & (\widehat{p}^\intercal \psi_\zeta \mid \widehat{p}\psi_\zeta) \end{bmatrix} = (\upsilon^*, \mathrm{j}\widetilde{\upsilon}) + (\upsilon^*, \widetilde{\upsilon})/2\pi) \tag{1.30}$$

for the "shifted" operators $\widehat{q} = Q - \overline{q}I$ and $\widehat{p} = P - \overline{p}I$.

It can easily be demonstrated that for a nonsingular Gaussian measure described by density $n(\zeta) = \nu(d\zeta)/\mathrm{d}\zeta$ of the form

$$n(\zeta) = C \exp\left\{-\tfrac{1}{2}(\zeta - \theta)\sigma_{\zeta\zeta}^{-1}(\zeta - \theta)^\intercal\right\} \tag{1.31}$$

(with $\theta = (\varkappa, \lambda)$ and $C = J/\sqrt{\det 2\pi\sigma_{\zeta\zeta}}$) corresponding to a nonsingular correlation matrix $\sigma_{\zeta\zeta}$ we can select representation (1.24) of the density operator S by appropriate choice of matrix ω in such a manner that the representation will be described by density (1.31) with the matrix $\sigma_{\zeta\zeta}$ of the form

$$\sigma_{\zeta\zeta} = \frac{1}{\pi} \operatorname{Re}\left[(\upsilon^*, \mathrm{j}\widetilde{\upsilon})^\dagger s(\upsilon^*, \mathrm{j}\widetilde{\upsilon})\right] \tag{1.32}$$

where s is a complex-valued positive definite $(d+1)$-by-$(d+1)$ matrix. At this point it is expedient to introduce a complex-valued normal representation characterized by the transition made from $2(d+1)$ real variables $\zeta = (\xi, \eta)$ to a $(d+1)$-dimensional complex variables

$$\alpha = \frac{1}{\gamma}(\xi\omega + 2\pi\mathrm{j}\eta),$$

where $\gamma = (\omega+\omega^*)^{1/2}$ in terms of which the density (1.31) combined with (1.32) can be written in the following form:

$$n(\alpha, \alpha^*) = C \exp\{-(\alpha-\theta)^* s^{-1} (\alpha-\theta)^\intercal\}, \tag{1.33}$$

where $\theta = (\varkappa\omega + 2\pi \mathrm{j}\lambda)\gamma^{-1}$, and $C = J/\det s$ if density (1.33) is normalized to J with respect to $\mathrm{d}\alpha\,\mathrm{d}\alpha^* = \mathrm{d}\xi\,\mathrm{d}\eta$. Note that, as in the case with (1.13), amplitudes (1.25) must be written in the form

$$|\alpha)(q) = (\gamma/\sqrt{2\pi})^{1//2} \exp\left\{(q\gamma - \operatorname{Re}\alpha)\alpha^\dagger - \tfrac{1}{2} q\omega\, q^\intercal\right\}, \tag{1.34}$$

with the scalar product defined in(1.14). However, in contrast to (1.13), these amplitudes do not realize at $\operatorname{Im}\omega \neq 0$ the lower bound in the uncertainty relation (1.11) while they do realize a more exact lower bound defined by the matrix inequality

$$\det \begin{bmatrix} \sigma_{qq} & \sigma_{qp} \\ \sigma_{pq} & \sigma_{pp} \end{bmatrix} \geq 0, \text{ or } \sigma_{pp} \geq \sigma_{pq}\sigma_{qq}^{-2}\sigma_{qp}, \tag{1.35}$$

provided that matrix σ_{qq} is nonsingular. Here

$$\begin{aligned} \sigma_{qq} &= \sigma_{\xi\xi} + \gamma^{-1}, \qquad \sigma_{pp} = \sigma_{\eta\eta} + \omega^*\gamma^{-2}\omega/(2\pi)^2, \\ \sigma_{pq} &= \sigma_{\eta\xi} + \omega^*\gamma^{-1}/(2\pi\mathrm{j}),\ \sigma_{qp} = \sigma_{\xi\eta} - \omega\gamma^{-2}/(2\pi\mathrm{j}) \end{aligned}$$

are the elements of the correlation matrices,

$$\begin{aligned} \sigma_{qq} &= J^{-1} \int (\widehat{q}^\intercal \psi_\alpha \mid \widehat{q}\psi_\alpha)\nu\,(d\alpha)\,, \\ \sigma_{qp} &= J^{-1} \int (q^\intercal \psi_\alpha \mid \widehat{p}\psi_\alpha)\nu\,(d\alpha)\,, \\ \sigma_{pq} &= J^{-1} \int (\widehat{p}^\intercal \psi_\alpha \mid \widehat{q}\psi_\alpha)\nu\,(d\alpha)\,, \\ \sigma_{pp} &= J^{-1} \int (\widehat{p}^\intercal \psi_\alpha \mid \widehat{p}\psi_\alpha)\nu\,(d\alpha)\,, \end{aligned} \tag{1.36}$$

which are defined for nonmixed canonical signals in (1.30) and satisfy, obviously, the following relation:

$$\sigma_q(\sigma_p^2 - \rho^*\sigma_p^2\rho^\intercal)\sigma_q = 1/(4\pi)^2, \tag{1.37}$$

with $\sigma_p^2 = \sigma_{pp}, \sigma_q^2 = \sigma_{qq}$, and $\rho = \omega^{-1}\operatorname{Im}\omega$; at $\rho = 0$ this relation realizes the bound of (1.11). Otherwise, it realizes the bound to (1.35).

1.1.5 Canonical Representations

In problems dealing with the recognition of complicated sound and visual patterns, the most important information is usually contained in the momentum representation as well as in the coordinate representation. For example, in analyzing speech not only the frequency distribution of its intensity is important but so is its temporal distribution, in the same way as in colour pattern recognition it has proved important to know the wave structure in addition to the spatial structure. although there can be no joint coordinate-momentum representation that would enable calculating such distributions simultaneously (due to noncommutativity of position and momentum operators), the simultaneous estimate, say by the human ear, of the frequency and temporal structures of sound points to the possibility of building a mathematical model of such perception, which may prove extremely important for automatic speech recognition.

The simplest models of such joint coordinate-momentum representations are those whose densities are defined as the intensities

$$k(z) = |(\psi_z \mid \varphi)|^2, \quad z = (x, y) \in \mathbb{R}^{2(d+1)}, \tag{1.38}$$

of projections of amplitude φ on the canonical amplitudes (1.24) at $\zeta = z$, which are parametrized at a fixed ω by the estimate vectors $\xi = x$ and $\eta = y$ of the generalized coordinates $x = (x_0, \dots, x_d)$ and momenta $y = (y_0, \dots, y_d)$. We can directly verify that the density in x has the form

$$m(x) = \int k(x, y) dy = |v|^{-1/2} \int \mathrm{e}^{-\pi |(x-q)v^{-1}|^2} |\varphi(q)|^2 dq, \tag{1.39}$$

which in the limit of $|v| = \sqrt{\det v^\dagger v} \to \infty$ coincides with the coordinate distribution $\iota(q) = |\varphi(q)|^2$. Similarly, in the p-representation we find the density in y:

$$\begin{aligned} \widetilde{m}(y) &= \int k(x, y) dx \\ &= |\widetilde{v}|^{-1/2} \int \mathrm{e}^{-\pi |(y-p)\widetilde{v}^{-1}|^2} |\widetilde{\varphi}(p)|^2 dp. \end{aligned} \tag{1.40}$$

which in the limit of $|\widetilde{v}| \to \infty$ coincides with the momentum distribution $\iota(p) = |\widetilde{\varphi}(p)|^2$. Here for every amplitude φ we have

$$\begin{aligned} \|\varphi\|^2 &= \int m(x) dx = \iint k(x, y)\, dx\, dy \\ &= \int \widetilde{m}(y)\, dy = \|\widetilde{\varphi}\|^2, \end{aligned} \tag{1.41}$$

which means that the set of canonical amplitudes $\{\psi_z \mid z \in \mathbb{R}^{2d}\}$ is complete for every ω, with $\mathrm{Re}\,\omega$ positive. Thus, the $2(d+1)$-parametric set $\{\psi_z\}$ forms a nonorthogonal base that defines for each ω a canonical representation in which the diagonal elements of the kernel $(\psi_z \mid S\psi_2)$ of a signal described by density operator S yield the density of the distribution of the signal's intensity $\iota(S) = \langle S, I \rangle$. For mixed canonical signals (1.14) has the form

$$k(x, y) = \int \exp\{-|c - \alpha|^2\} n(\xi, \eta)\, \mathrm{d}\xi,\, \mathrm{d}\eta, \tag{1.42}$$

where $c = (x\omega + 2\pi \mathrm{j} y)\upsilon^\dagger / 2\pi$, specifically, for Gaussian signals (1.25) we arrive at the following Gaussian density:

$$k(z) = J \exp\left\{-\tfrac{1}{2}(z-\theta)\sigma_{zz}^{-1}(z-\theta)^\intercal\right\} / \sqrt{\det 2\pi\sigma_{zz}}, \tag{1.43}$$

where

$$\sigma_{zz} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} = \sigma_{\zeta\zeta} + 2(\upsilon^\dagger, \mathrm{j}\widetilde{\upsilon}) + (\upsilon^*, \mathrm{j}\widetilde{\upsilon}^\intercal)/\pi \tag{1.44}$$

is a $(d+1)$-by-$(d+1)$ correlation matrix (as usually $\mathrm{j} = \sqrt{-1}$).

A remarkable property of canonical representations is the possibility of calculating the intensity distribution in any representation knowing only one canonical distribution by analytically continuing the density $k(c, c^*) = k(x, y)$ to a kernel

$$k(c', c^*) = (c|S|c') \tag{1.45}$$

that is holomorphic in c' and $c^* \in \mathbb{C}^{d+1}$, where $|c) = \psi_z$ and similarly, $|c') = \psi_{z'}$ at

$$c' = (x'\omega + 2\pi\mathrm{j}y')\upsilon^\dagger/\sqrt{2\pi}$$

for $z' = (x', y')$. For one thing, for operator (1.24) with a density of the form (1.33) we obtain

$$k(c', c^*) = J(c \mid c')\mathrm{e}^{-(c^*-\theta^*)(s+1)^{-1}(c'-\theta')^\intercal} / |S.| \,. \tag{1.46}$$

The transition from kernels of the (1.45) type to, say, the coordinate representation of operator S is carried out by the following formula:

$$S = \int |c)(c' \mid k(c', c^*)(c \mid c')\, \mathrm{d}c\, \mathrm{d}c^*\, \mathrm{d}c'\, \mathrm{d}c^{*\prime}, \tag{1.47}$$

where $|c)(c'| : \varphi \mapsto (c' \,|\varphi)|\, c)$ are one-dimensional operators defined by the canonical amplitudes (1.24), respectively, at $\alpha = c$ and $c' \in \mathbb{C}^{d+1}$, $\mathrm{d}c\,\mathrm{d}c^* = \mathrm{d}x\,\mathrm{d}y$, $\mathrm{d}c'\,\mathrm{d}c^{*\prime} = \mathrm{d}x'\,\mathrm{d}y'$,

$$(c \mid c') = \exp\left\{-\tfrac{1}{2}|c|^2 + c^*c' - \tfrac{1}{2}|c'|^2\right\}. \tag{1.48}$$

Note that the canonical kernels (1.45), as operators S, act in the space of entire functions

$$h(c) = \mathrm{e}^{|c|^2/2}(\varphi \mid c), \tag{1.49}$$

which define the representation of amplitudes $\chi \in \mathcal{H}$ in the Bargmann space of all entire functions on $\mathbb{C}^{d+1}$, for which

$$\int |h(c)|^2 \exp\left\{-|c|^2/2\right\} \mathrm{d}c\,\mathrm{d}c^* < \infty. \tag{1.50}$$

Specifically, one-dimensional operators of density $S = |\varphi)(\varphi|$ have kernels

$$k(c', c^*) = (c \mid \varphi)(\varphi \mid c'). \tag{1.51}$$

1.2 Mathematical Models of Wave Pattern Analyzers

In this section we will consecutively introduce and describe mathematical models of an ideal filter, a quasifilter, a disjoint selector, and a quasiselector that make it possible to move to arbitrary representations necessary for solution of the problems of best wave-pattern recognition based on measurements of pattern intensities in a single representation. We will also discuss a dilation theory, based on the work done by Halmos [29] and Neumark [47], for designing ideal filters and selectors and their realization via indirect measurements, an idea that originated in quantum theory [46].

1.2.1 Ideal Filters

The simplest measurement of a signal is the determination of the intensity of the oscillations in the signal in a given mode described by a vector ψ normalized to unity, $\|\psi\| = 1$, belonging to the Hilbert space $\mathcal{H}$ of amplitudes φ admissible at the "in" terminals of the receiver and having a finite intensity $I(\varphi) = (\varphi \mid \varphi) < \infty$. The oscillation amplitude in mode ψ is determined by the projection $(\psi \mid \varphi)$ of the received amplitude φ on direction ψ, while the intensity is calculated according to the formula

$$E_\psi(\varphi) = (\varphi \mid \psi)(\psi \mid \varphi) = |(\varphi \mid \psi)^2, \tag{1.52}$$

similar to the transition amplitude of a quantum mechanical system from state φ to state ψ. The intensity given by formula (1.52) is a positive quantity, just as probability is; however, it can assume values greater than unity (but not greater than the total intensity $I(\varphi)$). The appropriate measuring device acts as an ideal filter if it receives a signal φ completely, provided that ψ and φ are collinear, and does not receive φ if φ and ψ are orthogonal. A mixed signal described by a density operator S excites in mode ψ oscillations of intensity

$$\varepsilon_\psi(S) = \langle S, E_\psi \rangle = (\psi \mid S\psi). \tag{1.53}$$

More general analyzers carry out the measurement of the intensity

$$E(\varphi) = (\varphi \mid E\varphi) = \|E\varphi\|^2 \tag{1.54}$$

of the projection $E\varphi$ of the received signal on an arbitrary subspace of $\mathcal{H}$ described by an orthoprojector $E = E^*E$. For example, an audio frequency filter with a pass band Δ is determined by an orthoprojector $E = I(\Delta)$ on the subspace of amplitudes $\psi(f)$ with support $\Delta = \Phi$ acting as the operator of multiplication by the indicator $1(f, \Delta)$ of set Δ. In a similar manner one can define spatial optical filters that cut out a visual field in a certain region of aperture Δ.

The reader will recall that an orthoprojector is any linear operator $E : \mathcal{H} \to \mathcal{H}$ satisfying the condition $E^* = E = E^2$, and to each Hilbert space $\mathcal{E} \subseteq \mathcal{H}$

there corresponds a unique orthoprojector for which $\mathcal{E} = E\mathcal{H}$. Bearing in mind this one-to-one relation, we will describe such analyzers by orthoprojectors and call them ideal filters, that is, as noted earlier, filters that pass a signal without distortion while measuring its intensity if $\varphi \in E\mathcal{H}$ and that do not pass a signal if it is orthogonal to space $E\mathcal{H}$. The set of all such filters is partially ordered with a smallest element and a greatest element for which we take the null operator 0 and the identity operator I; specifically, filter A is stronger than filter B, or $A \geq B$, if orthoprojector B is greater than A, or $AB = A$. Maximal filters, with the exception of the zero filter, are described by one-dimensional orthoprojectors $E_\chi = |\chi)(\chi|$ acting according to the formula

$$|\chi)(\chi|\varphi = \chi(\chi|\varphi) = (\chi|\varphi)\chi \tag{1.55}$$

and defining the normalized vector χ to within a phase factor $\mathrm{e}^{\mathrm{j}\theta}$. For every filter A there is a unique complementary filter $A^\perp$ such that $A + A^\perp = I$, with $A^{\perp\perp} = A$, and $A^\perp \geq B^\perp$ if $B \geq A$. The orthogonal complement $A \to A^\perp$ possesses all the properties of logical negation: if A "passes" $\varphi : A\varphi = \varphi$, then $A^\perp$ does not: $A^\perp\varphi = 0$, and vice versa, with $A \wedge A^\perp = 0$ and $A \vee A^\perp = I$ with respect to the operations of conjunction $\wedge$ and disjunction $\vee$, for which we take the upper and lower bounds

$$A \wedge B = \sup\{A, B\}, \quad A \wedge B = \inf\{A, b\}, \tag{1.56}$$

and with the duality formula being valid, or $(A \vee B)^\perp = A^\perp \vee B^\perp$.

Generally, filters A and B are said to be disjoint if $A \perp B$, that is, $A\varphi = \varphi$ implies $B\varphi = 0$ ($B\varphi = \varphi \Rightarrow A\varphi = 0$), or, which is equivalent, if $A^\perp \geq B$ ($B^\perp \geq A$); they are said to be incompatible if $A \wedge B = 0$, that is, if there is not a single signal that can pass completely through filter A and filter B in the sense that $A\varphi = \varphi$ and $B\varphi = \varphi$.

It can easily be shown that disjoint filters are incompatible, but not the other way round. For this reason the logic of filters is nondistributive, similar to the logic of quantum theory, which is also nondistributive. It satisfies a weaker condition of orthomodularity

$$(A \vee B^\perp) \wedge C = A \vee (B^\perp \wedge C) \text{ if } A \leq B \leq C. \tag{1.57}$$

It is in this respect that the logic of filters differs from Boolean logic, where incompatibility and disjointness mean the same. The intensity of a mixed signal S measured by an ideal filter E can be calculated via the formula

$$\varepsilon(S) = \langle S, E\rangle = \mathrm{Tr}\,(ES). \tag{1.58}$$

1.2.2 Disjoint Selectors

Complicated analyzers measure the intensities $E_i(\varphi) = |(\chi_i \mid \varphi)|^2$ of the received field φ simultaneously in several standard modes $\chi_i \in \mathcal{H}$, $i = 1, \ldots, m$, which, if the normalization conditions $\|\chi_i\| = 1$ for all i's, are necessarily orthogonal in view of the condition

$$\sum_{i=1}^{m} E_i(\varphi) \leq I(\varphi).$$

Otherwise, the total received intensity $\sum_{i=1}^{m} E_i(\varphi)$ could be greater than the total intensity $I(\varphi) = \|\varphi\|^2$ of the received signal φ. Such analyzers act as disjoint selectors, or ideal selective filters that split the received signal φ into orthogonal components $\varphi_i = (\chi_i \mid \varphi)\chi_i$, $i = 1, \ldots, m$. Signal φ is received completely by a selective filter if $E\varphi = \varphi$, where $E = \sum_{i=1}^{m} E_i$ is the appropriate nonselective filter defined by the one-dimensional orthoprojectors E_i on the subspaces generated by the standard modes χ_i.

More general selective filters are specified by arbitrary sets (or families) $\{E_i \mid i = 1, \ldots, m\}$ of projectors $E_i : \mathcal{H} \to \mathcal{H}$ that satisfy the condition of pairwise orthogonality $E_i E_k = 0$ for $i \neq k$. For example, a disjoint selector that measures the intensity of a signal in each region Δ_i of a Borel partition

$$\Omega = \sum_i \Delta_i : \Delta_i \subseteq \Omega$$

is described by an orthogonal set of $E_i = I(\Delta_i)$ of indicators $I(\Delta_i) = \{1(q, \Delta_i)\}$. Note that such a set $\{E_i\}$ may have an infinite number of members if space $\mathcal{H}$ is not finite-dimensional; in this case the received intensity is determined for each φ by an absolutely convergent series $\sum_{i=1}^{\infty} E_i(\varphi) \leq I(\varphi)$, where

$$E_i(\varphi) = (\varphi \mid E_i\varphi) = \|E_i\varphi\|^2 . \tag{1.59}$$

A selective measurement is said to be complete if the inequality $E(\varphi) \leq I(\varphi)$ is transformed into an equality for every $\varphi \in \mathcal{H}$, that is, if $\sum_{i=1}^{\infty} E_i = I$, in a strong operator topology (I is the identity operator in $\mathcal{H}$) and is said to be maximal if all the E_i are one-dimensional.

Complete filters are usually related to self-adjoint operators with a nondegenerate discrete spectrum $\{x_i\}$ through the spectral decomposition (or expansion)

$$A = \sum_{i=1}^{\infty} x_i E_i, \tag{1.60}$$

with each set $\{E_i\}$ being assigned a numbering self-adjoint operator $N = \sum_i iE_i$.

In addition to discrete filters there is another important class of filters, known as continuous filters, which are related to normal operators with a continuous spectrum $X \subseteq \mathbb{C}^1$. In accordance with von Neumann's theorem, to each such operator there is uniquely assigned a projector-valued measure E on X that specifies the orthogonal expansion (or decomposition) of unity $I = \int E(dx)$, so that

$$A = \int xE(dx),\ \mathcal{D}(A) = \left\{\chi \in \mathcal{H} : \int |x|^2 E(\chi, dx) < \infty\right\}. \tag{1.61}$$

Here a family $\{A_j \mid j = 1, \ldots, n\}$ of pairwise commutative normal operators $A_j : \mathcal{H} \to \mathcal{H}$ has corresponding to it a selective filter described by a projector-valued measure $E = \bigotimes_{j=1}^{n} E_j$ on $X \subseteq \mathbb{C}^m$ that defines a spectral representation $A = \int xE(dx)$ for the vector operator $A = (A_j)$. It is with these vector selective filters that the measurement of the intensity distribution (1.1) in the coordinate region is carried out. Such a distribution is described by the orthogonal decomposition of unity $I = \int I(dx)$ for the coordinate vector operator $Q = (Q_k, k = 0, 1, \ldots, d)$; the coordinates in the coordinate (or position) representation are given by the respective operator of multiplication by $q = (q_k)$, so that

$$I(\Delta)\varphi(q) = 1(q, \Delta)\varphi(q), \tag{1.62}$$

where $1(\Delta)$ is the indicator of the Borel subset $\Delta \subseteq \mathbb{R}^{d+1} : 1(\Delta, q) = 1$ for $q \in \Delta$ and $1(q, \Delta) = 0$ for $q \notin \Delta$. The result of such a measurement is the continuous measure $I(\varphi, dx)$ with a density

$$\iota(\varphi, x) = I(\varphi, dx)/\mathrm{d}x = |\varphi(x)|^2. \tag{1.63}$$

Note that the self-adjoint position operator

$$Q = \int qI(dq),\ \mathcal{D}(Q) = \left\{\chi \in \mathcal{H} : \int q^2|\chi(q)|^2 \mathrm{d}q < \infty\right\} \tag{1.64}$$

has a domain of definition $\mathcal{D}(Q)$ coinciding with $\mathcal{H} = L^2(Q)$ only in the case of a bounded region Ω. Generally speaking, operator Q is only a densely definite operator, such as the frequency operator $F = \int_0^\infty$ if $I(df)$ in the case of a semi-infinite band $\Phi = [0, \infty[$ of the spectrum.

In general, let X be an arbitrary set and $\mathcal{B}(X)$ the Borel algebra of its subsets. Every measure $E : \Delta \in \mathcal{B} \mapsto E(\Delta)$ with values in the orthoprojectors of the Hilbert space $\mathcal{H}$ is said to be a disjoint selector, it is called a complete selector if $E(X) = I$. Disjoint selectors measure the intensity distribution in the received signal φ on X according to the formula

$$E(\varphi, \Delta) = (\varphi \mid E(\Delta)\varphi) = \|E(\Delta)\varphi\|^2, \tag{1.65}$$

and define for each φ a positive measure on X of finite mass $E(\varphi, X) \leq I(\varphi)$, coinciding with $I(\varphi) = \|\varphi\|^2$ in the case of a complete selector.

We will say that selector E' on X' majorizes selector E on X (denoted $E' \gtrsim E$) if there exists a measurable mapping $f : X' \to X$ with respect to which

$$E(\Delta) = E'(f^{-1}(\Delta)) \text{ for every } \Delta \in \mathcal{B}(X) \tag{1.66}$$

where $f^{-1}(\Delta) = \{x' \in X' \mid f(x') \in \Delta\}$ is the inverse image of set $\Delta \subseteq X$, and the selectors E' and E are equivalent, $E' \simeq E$, if $E' \lesssim E$, too.

1.2.3 Successive Filters and Quasifilters

The common practice in processing sound and visual patterns is to use analyzers that act not in the initial Hilbert space $\mathcal{H}$ generated by the amplitudes φ on the "in" terminals but in an extension of this space. For example, temporal measurements of sound signals $\varphi(f)$ with a restricted frequency band Φ are reduced to determining the intensity of these signals in this or that temporal interval $\Delta \in \mathbb{R}^1_+$ via the orthoprojector $\widetilde{I}(\Delta)$ of multiplication of the signal in the temporal representation by the indicator function $1(t, \Delta)$:

$$\widetilde{I}(\varphi, \Delta) = \int 1(t, \Delta)|\widetilde{\varphi}(t)|^2 \mathrm{d}t = \left\| \widetilde{I}(\Delta)\varphi \right\|^2 \tag{1.67}$$

where $\widetilde{I}(\Delta)$ acts in the space of signals of unlimited bandwidth, $L^2(\mathbb{R})$. In a similar manner space $\mathcal{H}$ is extended to $L^2(\mathbb{R}^d)$ in the wave processing of optical fields observed on a limited aperture $S \subset \mathbb{R}^d$, which results in determining the intensities of the fields in this or another momentum interval $\Delta \in \mathbb{R}^d$.

In general, such an extension is described by an isometric embedding $F : \mathcal{H} \to \mathcal{H}'$ of Hilbert space $\mathcal{H}$ into another space $\mathcal{H}'$, for which for examples considered here we can take the Hilbert space $\mathcal{H}' = L^2(\mathbb{R}^{d+1})$ into which the space $\mathcal{H} = L^2(\Omega)$ is isometrically embedded via the Fourier transform $F : \varphi \mapsto \widetilde{\varphi}^2$. An ideal filter described in $\mathcal{H}'$ by the orthoprojector E measures the intensity of amplitude $\varphi \in \mathcal{H}$ defined by the Hermitian form

$$D(\varphi) = \|EF\varphi\|^2 = (F\varphi \mid EF\varphi) = (\varphi \mid D\varphi), \tag{1.68}$$

where $D = F^*EF$ is a positive contraction operator in $\mathcal{H}$. Formula (1.68) shows that this intensity can be considered the result of successive action of two ideal filters, F and E, with $\mathcal{H}$ being identified with a subspace $F\mathcal{H} \subset \mathcal{H}'$, where filter F is described by the orthoprojector F that cuts subspace $\mathcal{H}$ out of $\mathcal{H}'$. For example, temporal measurement of narrow-band signals is the result of noncommutative action of a frequency filter $F = I(\Delta f)$ and a temporal filter $E = \widetilde{I}(\Delta t)$, the result is effectively described by the Hermitian form (1.68) defined by the operator

$$D = I(\Delta f)\widetilde{I}(\Delta t)I(\Delta f).$$

It is, therefore, advisable to generalize the concept of a filter by describing it in space $\mathcal{H}$ by any operator $D : \mathcal{H} \to \mathcal{H}$ that satisfies the condition

$$I \geq D^* = D \geq 0, \tag{1.69}$$

and calling it a quasifilter if $D \neq D^2$.

The basis for this extension is the Halmos theorem [29], according to which every quasifilter described by operator (1.69) can be considered as a reduction (projection) $D = F^*EF$ on $\mathcal{H}$ of an ideal filter E acting in an extension $\mathcal{H}'$. For the Hilbert space $\mathcal{H}'$ we can always take the doubling $\mathcal{H}' = \mathcal{H} \oplus \mathcal{H} = \mathbb{C}^2 \otimes \mathcal{H}$ of space $\mathcal{H}$ with embedding $F : \varphi \mapsto (\varphi, 0)$, selecting the operators

$$E_{11} = D,\ E_{12} = \sqrt{D(I - D)} = E_{21},\ E_{22} = I - D \tag{1.70}$$

for the blocks of orthoprojector E. Note that allowance for consecutive action of several noncommutative ideal filters $E_1, \ldots, E_n$ in the initial Hilbert space also leads to the notion of a quasifilter. These ideal filters then measure the intensity

$$D(\varphi) = \|E_n \dots E_1\varphi\|^2 = (\varphi \mid D\varphi), \tag{1.71}$$
$$D = E_1 \dots E_{n-1}E_nE_{n-1}\dots E_1, \tag{1.72}$$

and the result can be considered the effect of linear nonideal filters not necessarily described by Hermitian contraction operators $A : \mathcal{H} \to \mathcal{H}$, $\|A\| \leq 1$, with the nonideal filters damping and distorting the amplitudes and with $D = A^*A$ in the formula for the appropriate intensity:

$$D(\varphi) = \|A\varphi\|^2 = (\varphi \mid D\varphi). \tag{1.73}$$

1.2.4 Quasiselectors and Indirect Measurements

In a similar manner we can introduce generalized selectors, which are defined on a Borel space X by a positive operator-valued measure $M : \Delta \in \mathcal{B}(X) \to M(\Delta)$ specifying in the Hilbert space $\mathcal{H}$ a weak decomposition $D = \int M(dx)$ of an operator D satisfying condition (1.69) in the following sense:

$$D(\varphi) = \int M(\varphi, dx) \text{ for every } \varphi. \tag{1.74}$$

Here, as usual, $D(\varphi) = (\varphi \mid D\varphi)$ is a Hermitian form defined by the operator of total effect D, while

$$M(\varphi, dx) = (\varphi \mid M(dx)\varphi) \tag{1.75}$$

is the distribution of intensity of X corresponding to amplitude φ and measured by such a selector. Note that the expansion of operator D defined by measure $M(dx)$ may not necessarily be orthogonal even if the operator is a projector, that is, if the total filter is ideal, such nondisjoint selective filters will be called quasiselective filters, or simply quasiselectors. A quasiselector is said to be complete if $M(X) = I$ and maximal if $M \lesssim M' \Rightarrow M' \simeq M$ in the same sense as in (1.66).

An example of a complete maximal quasiselector for sound signals and optical fields observed in a restricted region Ω of coordinates $\mathbb{R}^{d+1} \ni q = (f, \mathbf{q})$ is the analyzer of the momentum distribution (1.4), which in the q-representation is defined by formula (1.75) via the operator-valued measure

$$M(dp) = F^*I(dp)F \equiv \widetilde{I}(dp), \tag{1.76}$$

where F is the Fourier transform (1.5), and $I(D) = \{1(p, D)\}$ is the projector-valued measure on $\mathbb{R}^{d+1}$ described in the space $\mathcal{H}' = L^2(\mathbb{R}^{d+1})$ of the proper representation of the generalized-momentum operator $p = (t, \mathbf{p})$ by the indicator measure $1(p, dx)$.

Note that the momentum operator defined in $\mathcal{H} = L^2(\Omega)$ by the nonorthogonal integral expansion

$$P = \int p\widetilde{I}(dp),\ \mathcal{D}(P) = \left\{\chi \in \mathcal{H} : \int p^2|\widetilde{\chi}(p)|^2 \mathrm{d}p < \infty\right\}$$

is always unbounded with a spectrum $\mathbb{R}^{d+1}$ and, for $\Omega \neq \mathbb{R}^{d+1}$, non-selfadjoint, notwithstanding the fact that the form of the total momentum $P(\varphi) = \int p\widetilde{I}(\varphi, dp)$ is always Hermitian. Nevertheless, this operator always uniquely defines a nonorthogonal expansion $I = \int \widetilde{I}(dp)$ via the condition

$$\int p^2(\chi \mid \widetilde{I}(dp)\chi) = (P\chi \mid P\chi) \quad \forall \chi \in \mathcal{D}(P),$$

and is a restriction to functions $\varphi(q) = 0$ for $q \notin \Omega$ of the operator $(2\pi \mathrm{j})^{-1}\partial/\partial q$ that is self-adjoint in $L^2(\mathbb{R}^{d+1})$ with a domain of definition

$$\mathcal{D}(P) = \{\varphi \in L^2(\mathbb{R}^{d+1}) : \left\|\partial^2\varphi/\partial q^2\right\|^2 < \infty\}.$$

For instance, time operator $T = \int t\widetilde{I}(dt)$ in the space $\mathcal{H} = L^2(\Phi)$ with a semi-infinite band $\Phi = [0, \infty[$ is a symmetric but not a self-adjoint operator $(2\pi \mathrm{j})^{-1}\partial/\partial f$ with a domain of definition

$$\mathcal{D}(T) = \left\{\chi \in L^2(0, \infty) : \chi(0) = 0, \int_0^\infty |\partial\chi(f)/\partial f|^2 \mathrm{d}f < \infty\right\}.$$

Similarly, the validity of the representation

$$M(dx) = F^* E(dx) F,\ F^* F = I, \tag{1.77}$$

for an arbitrary quasiselector M in the form of the projection of the disjoint selector E described in the extended space $\mathcal{H}'$ by an orthogonal projector-valued measure $E(dx)$ is ensured by the Neumark theorem [47]. For mixed signals the intensity, as a function of the density operator S, is described by a measure $\mu(S, dx)$ defined by the following linear form:

$$\mu(S, dx) = \langle S, M(dx)\rangle\,. \tag{1.78}$$

Quasiselective filters also emerge as a result of reducing the description of indirect measurement of the received signal via the disjoint selection $E_0(dx)$ of the initially uncorrelated reference signal interacting with the received signal; this reference signal generates a Hilbert space $\mathcal{H}_0$. Specifically, if S_0 is the density operator of the normalized reference signal ($\mathrm{Tr}\, S_0 = 1$) and U is a unitary operator describing in the tensor product $\mathcal{H} \otimes \mathcal{H}_0$ the result of the interaction $S' = U(S \otimes S_0)U^*$ with the received signal S, then the intensity distribution corresponding to such indirect measurement may be effectively calculated via formula (1.78) as a result of the quasimeasurement

$$\langle S, M(dx)\rangle = \langle S', I \otimes S_0(dx)\rangle = \langle S \otimes S_0, E'(dx)\rangle$$

described by the operator-valued measure

$$M(dx) = \mathrm{Tr}\ \left[(I \otimes S_0)E'(dx)|\mathcal{H}\right], \tag{1.79}$$

where $E'(dx) = U^*(I \otimes E_0(dx))U$, and $\mathrm{Tr}\,\{\cdot|\mathcal{H}\}$ is the partial trace in $\mathcal{H} \otimes \mathcal{H}_0$ defined for factorable density operators $S \otimes S_0$ via the formula

$$\mathrm{Tr}\,\{S \otimes S_0|\mathcal{H}\} = S\mathrm{Tr}\,S_0.$$

The indirect calculation of the intensity distribution over the frequency (colour) $f \in \Phi$ of static monochrome patterns is an example of the above-mentioned type of measurement. It can be carried out as a result of the wave processing of the patterns in which the intensity distribution over the momenta p in such patterns is calculated.

Using the Neumark theorem as a basis, let us give an explicit description of a construction that makes it possible to reduce any quasimeasurement to an indirect measurement. To this end we take for $\mathcal{H}_0$ the space $\mathcal{H}'$ of the Neumark construction and for the reference signal a normalized amplitude $\psi' = F\psi$, $\|\psi\| = 1$, and introduce the linear operator U in $\mathcal{H} \otimes \mathcal{H}'$ (which is defined by the Neumark isometry $F : \mathcal{H} \to \mathcal{H}'$, $F^*F = I$) in the following manner:

$$U : \varphi \otimes \varphi' \mapsto F^*\varphi' \otimes F\varphi + \varphi \otimes (1 - FF^*)\varphi' \tag{1.80}$$

with the generating elements being $\varphi \otimes \varphi'$, $\varphi \in \mathcal{H}, \varphi' \in \mathcal{H}'$. It can be directly verified that $U = U^*$ and $U^2 = I$ and, hence, $U^*U = I = UU^*$. Taking for $E_0(dx)$ the Neumark expansion $E(dx)$ in $\mathcal{H}'$, assuming that $S_0 = |\psi)(\psi|$ and allowing for (1.77), we arrive at an indirect measurement whose reduction (1.79) yields the initial measure $M(dx)$:

$$\mathrm{Tr}\,\{(I \otimes |\psi')(\psi'|)E'(dx)\} = (\psi \mid \psi)F^*E(dx)F = M(dx).$$

1.2.5 Canonical Operators and Measurements

Bearing in mind the invariance of the domains of definitions of operators Q and P with respect to the self-adjoint operators of multiplication by q and differentiation with respect to q, or $(2\pi \mathrm{j})^{-1}\,\partial/\partial q$, which on $\mathcal{H} = L^2(\Omega)$ coincide with Q and P, respectively, in what follows we will take for Q and P in $L^2(\mathbb{R}^{d+1})$ their extensions, while always assuming that the region where these operators act is $\Omega \subset \mathbb{R}^{d+1}$. Such operators are known as canonical and satisfy the commutation relations (1.9) in the common domain $\mathcal{D}(P) \cap \mathcal{D}(Q)$.

Let us now discuss simultaneous measurement of the coordinate (or position) and momentum distributions. In view of the noncommutativity of Q and P, there can be no joint orthogonal decomposition of unity for these operators; there can even be no joint nonorthogonal decomposition of $M(dx\,dy)$ for which the following would be true

$$I(dq) = \int M(dq\,dy),\ \widetilde{I}(dp) = \int M(dx\,dp). \tag{1.81}$$

Otherwise, in view of the Neumark theorem, there would be commutative self-adjoint operators in $\mathcal{H}' \supset \mathcal{H}$ coinciding on $\mathcal{H}$ with the noncommutative operators Q and P, which is impossible.

Another interesting question is the relation to these operators of the measurements of the canonical distributions (1.38). Such canonical measurements are described, obviously, by continuous with respect to $\mathrm{d}z = \mathrm{d}x\,\mathrm{d}y$ nonorthogonal measures $K(dz) = k(z)\mathrm{d}z$ with projector-valued densities

$$k(z) = |\psi_z)(\psi_z| = |c)(c| \equiv k(c, c^*), \tag{1.82}$$

which are defined by canonical amplitudes (1.25) at $\zeta = z$ and a certain ω or, in complex variables, by (1.34) at $\alpha = c$. The respective quasiselective filters, which are parameterized by symmetric ω matrices with a nonsingular real part $\omega + \omega^*$ and now will be called canonical filters, are, obviously, maximal and, because of (1.41), complete:

$$\int |\psi_z)(\psi_z|\mathrm{d}z = I = \iint |c)(c)\,\mathrm{d}c\,\mathrm{d}c^*, \tag{1.83}$$

where $\mathrm{d}c\,\mathrm{d}c^* = \mathrm{d}x\,\mathrm{d}y$.

By fixing ω and directly integrating we find that the quasimeasurement of an intensity distribution in $x \in \mathbb{R}^{d+1}$ is described by a continuous measure

$$M(dx) = \int K(dx\,dy) = m(x)\mathrm{d}x$$

diagonal in the q-representation,

$$m(x) = \int k(x, y)\mathrm{d}y = |\upsilon|^{-1} \int \mathrm{e}^{-\pi|(x-q)\upsilon^{-1}|^2} I(dq). \tag{1.84}$$

with a Gaussian density and $\upsilon^*\upsilon = 2\pi(\omega + \omega^*)^{-1}$, while the quasimeasurement of an intensity distribution in $y \in \mathbb{R}^{d+1}$ is described by an operator measure $\widetilde{M}(dy) = \int K(dx\,dy) = \widetilde{m}(y)\,dy$, where

$$\widetilde{m}(y) = \int k(x, y)dx = |\widetilde{\upsilon}|^{-1} \int \mathrm{e}^{-\pi|(y-p)\widetilde{\upsilon}^{-1}|^2} \widetilde{I}(dp), \tag{1.85}$$

with a Gaussian density and

$$\widetilde{\upsilon}^\dagger\widetilde{\upsilon} = 2\pi(\widetilde{\omega} + \widetilde{\omega}^*)^{-1},$$

where $\widetilde{\omega}/2\pi = 2\pi/\omega^*$. The operator measures M and $\widetilde{M}$ on $\mathbb{R}^{d+1}$, which define nonorthogonal expansions of operators Q and P, that is,

$$Q = \int xm(x)\mathrm{d}x \text{ and } P = \int y\widetilde{m}(y)\,\mathrm{d}y, \tag{1.86}$$

describe, in contrast to the spectral measures I and $\widetilde{I}$, inaccurate measurements of position and momentum distributions, which are obtained by smoothing out (1.39) and (1.40) with Gaussian weighting functions m and $\widetilde{m}$. Nevertheless, the canonical operator measure K that generates the two spectral measures possesses certain spectral properties with respect to the complex-valued combinations of the two respective operators:

$$A = \frac{1}{\sqrt{2\pi}}(Q\omega + 2\pi \mathrm{j} P)\upsilon^{\dagger}. \qquad (1.87)$$

Namely, applying A directly to the canonical amplitudes (1.34), we can easily verify that it is well-defined on these amplitudes:

$$A|\alpha) = \alpha|\alpha),\ \alpha \in \mathbb{C}^{d+1}, \qquad (1.88)$$

which, therefore, form a proper base for A in $\mathcal{H} = L^2(\mathbb{R}^{d+1})$.

Hermitian conjugate operators $C = A^*$ are diagonal in the Bargmann representation,

$$(\widehat{c}h)(c) = \mathrm{e}^{|c|^2/2}(C\chi|c) = \mathrm{e}^{|c|^2/2}c(\chi|c) = ch(c),$$

with a domain of definition $\mathcal{D}(\widehat{c}) = \{\chi \in \mathcal{H} : \|\widehat{c}h\| < \infty\}$, where

$$\|\widehat{c}h\|^2 = \int |c|^2 h(c)|^2 \mathrm{e}^{-|c|^2}\,\mathrm{d}c\,\mathrm{d}c^*, \qquad (1.89)$$

on which domain there is also defined the operator A by differentiation $(\widehat{a}h)(c) = \partial h(c)/\partial c$. Thus, in the initial representation we obtain the nonorthogonal "spectral" decompositions

$$\begin{aligned} A &= \int cK(dz),\ \ A^* = \int c^* K(dz) \\ \mathcal{D}(A^*) &= \left\{\chi \in \mathcal{H} : \int |c|^2 |(\chi \mid c)^2\,\mathrm{d}c\,\mathrm{d}c^* < \infty\right\}. \end{aligned}$$

Now let us describe a simple realization of a canonical measurement by an indirect measurement defined in the tensor product $\mathcal{H}\otimes\mathcal{H}_0$, where $\mathcal{H}_0$ is a copy of $\mathcal{H}$. To this end we take the commutative self-adjoint operators

$$X = Q \otimes I_0 + I \otimes Q_0,\ Y = P \otimes I_0 - I \otimes P_0, \qquad (1.90)$$

where $Q_0 = q_0$, $P_0 = (2\pi \mathrm{j})^{-1}\partial/\partial q_0$, and I_0 is the identity operator in $\mathcal{H}_0 = L^2(\mathbb{R}^{d+1})$. Suppose that $E(dz)$ is the orthogonal spectral measure of the set $Z = (X, Y)$ and that

$$\psi_0(q_0) = |\frac{\omega + \omega^*}{2\pi}|^{1/4} \exp\left\{-\tfrac{1}{2} q_0 \omega q_0^{\intercal}\right\} = |0)_0 \qquad (1.91)$$

is the basic canonical amplitude in $\mathcal{H}_0$. We take an arbitrary amplitude $\chi \in \mathcal{H}$ and the corresponding tensor product

$$(\chi \otimes \psi_0^*)(q, q_0) = \chi(q)\psi_0^*(q_0)$$

in $\mathcal{H}\oplus\mathcal{H}_0$ and define the characteristic function of the corresponding distribution thus:

$$\Upsilon(u, u^*) = \int \mathrm{e}^{\mathrm{j}(u^* c^{\intercal} + c^* u^{\intercal})}(\chi \otimes \psi_0^* \mid E(dz)\chi \otimes \psi_0^*), \qquad (1.92)$$

$$z = (x, y) \in \mathbb{R}^{2(d+1)}$$

where $u, u^* \in \mathbb{C}^{2(d+1)}$ and as usual,

$$c = \frac{1}{\sqrt{2\pi}}(x\omega + 2\pi \mathrm{j} y)\upsilon^\dagger.$$

We write this function in terms of normal operators

$$B = \int cE(dz) = \frac{1}{\sqrt{2\pi}}(X\omega + 2\pi \mathrm{j} Y)\upsilon^\dagger = A \otimes I_0 + I \otimes C_0, \tag{1.93}$$

with A the operators (1.87) and

$$C_0 = \frac{1}{\sqrt{2\pi}}(Q_0\omega - 2\pi \mathrm{j} P_0)\upsilon^\dagger,$$

in the form

$$\Upsilon(u, u^*) = (\mathrm{e}^{\mathrm{j}B^* u^\intercal} \chi \otimes \psi_0^* | \mathrm{e}^{\mathrm{j}B^* u^\intercal} \chi \otimes \psi_0^*) = (\mathrm{e}^{\mathrm{j}A^* u^\intercal} \chi | \mathrm{e}^{\mathrm{j}A^* u^\intercal} \chi),$$

where we have allowed for the property $C_0\psi_0^* = 0$ for the basic amplitude (1.91). Employing now the completeness property (1.83) of canonical amplitudes, we obtain

$$\begin{aligned} \Upsilon(u, u^*) &= \int (\mathrm{e}^{\mathrm{j}A^* u'} \chi | c)(c | \mathrm{e}^{\mathrm{j}A^* u'} \chi) \, \mathrm{d}c \, \mathrm{d}c^* \\ &= \int \mathrm{e}^{\mathrm{j}u^* c' + \mathrm{j}c^* u'} |(\chi | c)|^2 \, \mathrm{d}c \, \mathrm{d}c^* = \widetilde{k}(u, u^*). \end{aligned}$$

Thus, the characteristic function (1.92) of the indirect measurement of intensity of amplitude $\chi \in \mathcal{H}$ coincides for the ground state ψ_0^*, when calculated in the z-representation of operators (1.90), with the characteristic function $\widetilde{k}$ of the canonical distribution $k(c, c^*) = |(\chi \mid c)|^2$ for this amplitude.

Chapter 2

Optimal Wave Detection and Discrimination

In this chapter we develop the wave theory of hypothesis testing for solving problems of optimal recognition of sound and visual patterns. We formulate the necessary and sufficient conditions for the optimality of two-alternative and multialternative detection of wave patterns according to the maximum criterion for the measured intensity of acoustic signals and optical fields. We consider problems involving the discrimination of a wave pattern against an acoustic or optical background, problems involving the discrimination of pure nonorthogonal signals and fields, and problems involving the recognition of mixed patterns described by noncommutative density operators. Complete solution of the last type of problem is then obtained for the case of mixing two pure patterns. The discussed results of solution of the corresponding extremal problems follow from the methods of linear programming in Banach partially ordered operator spaces [43]. The results generalize the corresponding results of the quantum detection and estimation theory, which have been obtained for the two-alternative case by Helstrom [30] and for the multialternative case by Belavkin [14], [12]. The necessary and sufficient optimality conditions for the quantum theory of hypothesis testing have been discussed by Kennedy [40], Yuen and Lax [53], Kholevo [39], Belavkin [14], and Belavkin and Vancjan [10].

2.1 Optimal Detection of Sound and Visual Patterns

In this section we will discuss the problem of detecting wave patterns that are in a partially coherent superposition with an acoustic or optical background. The problem is complicated by the presence of interference. We start by considering the superposition principle for generalized mixed amplitudes. We then formulate the necessary and sufficient conditions for the optimality of detection and give solutions to a number of problems considered in the quantum case in the review [15].

2.1.1 The Superposition Principle

The problem of detecting a sound or visual pattern described by a wave amplitude $\varphi(q)$ taken from the Hilbert space $\mathcal{H} = L^2(\Omega)$ can be solved in a trivial manner by measuring the total intensity $I(\varphi) = \|\varphi\|^2$ only in the absence of an acoustic or optical background consisting of other signals and fields in the frequency-spatial region Ω considered. If in the region of measurement there is another signal or field described by amplitude $\varphi_0 \in \mathcal{H}$ the question of whether a wave pattern φ is present cannot generally be unambiguously solved by simply measuring the total intensity of the resulting amplitude ψ, which may be higher or lower than the background intensity. Such a phenomenon is called interference and is the result of the wave superposition principle $\psi + \varphi + \varphi_0$, according to which the complex-valued amplitudes of the coherent signals φ and φ_0 rather than the intensities of these signals, are added. The intensity of the resulting signal has the form

$$\|\psi\|^2 = \|\varphi\|^2 + 2\,\mathrm{Re}(\varphi \mid \varphi_0) + \|\varphi_0\|^2 . \tag{2.1}$$

To describe the result of the superposition of a mixed pattern and a partially coherent background caused, say, by thermal fluctuations that have an infinite total intensity of the acoustic or optical field, we can employ the correlation theory by considering generalized random amplitudes within the second-order statistical theory.

Partially coherent signals and fields determined in a similar manner in the framework of the classical or the quantum theory are commonly described by bounded operators F from $\mathcal{H}$ to another Hilbert space $\mathcal{K}$; for ordinary nonrandom amplitudes $\psi \in \mathcal{H}$ these operators are usually represented by the functional $F_\psi\chi = (\psi \mid \chi)$, denoted by $F_\psi = (\psi|$ and acting from $\mathcal{H}$ to $\mathcal{K} = \mathbb{C}$. The mean intensity of random oscillations excited in mode $\chi \in \mathcal{H}$, $\|\chi\| = 1$, is determined by a Hermitian form in F:

$$E_\chi(F) = (F\chi \mid F\chi) = (\chi \mid F^*F\chi), \tag{2.2}$$

and is calculated for common mixed signals via formula (1.53) with the aid of a (generally infinite trace) density operator $P = F^*F$ of the intensity $\iota(P) \in [0, \infty]$, where F^* is the Hermitian conjugate operator $\mathcal{K} \to \mathcal{H}$ acting for $F = (\psi|$ as an operator of multiplication $c \mapsto \psi c$ from $K = \mathbb{C}$ to $\mathcal{H}$. The intensity (2.2) is a measurable quantity bounded by the norm of the positive operator P and equal, via the duality theorem, to

$$\varepsilon_\chi(P) = (\chi \mid P\chi) \leq \|P\| = \inf\{\varepsilon \mid \varepsilon I \geq P\}, \tag{2.3}$$

which for the case of white noise $P = \varepsilon I$, described by the isometric operator $T = F/\sqrt{\varepsilon}$, $T^*T = I$, determines the local intensity $\varepsilon = \varepsilon_\chi(P)$, the same for all modes $\chi \in \mathcal{H}$. Note that every partially coherent signal F can be considered as the result of action of a contraction filter $D = P/\|P\|$ on white noise of local intensity $\varepsilon = \|P\|$ if we employ the polar decomposition $F = TP^{1/2}$, which determines uniquely the isometry operator T on the range of values $F^*\mathcal{K}$ of operator F^*.

For generalized signals with infinite trace density operators P it proves expedient, however, to consider only such quasimeasurements for which the operators D of the total effect lead to finite intensities:

$$D(F) \equiv \mathrm{Tr}\,(FDF^*) = \mathrm{Tr}\,(PD). \tag{2.4}$$

In addition to one-dimensional projector $D = E_\chi = |\chi)(\chi|$, for which the intensity (2.2) is determined by the bounded form (2.3), we can always consider finite-dimensional operators $D = \sum_i |\chi_i)(\chi_i|$ as well as any trace class operator $0 \le D < I$, since $\mathrm{Tr}\,(SD) \le \varepsilon \mathrm{Tr}\, D$ if $S < \varepsilon I$.

Extending the superposition principle to generalized amplitudes $F, F_0 : \mathcal{H} \to \mathcal{K}$, we find that the result $G = F + F_0$ of addition of the generalized signal F and the background F_0 is described by a density operator $R = G^*G$, that is the sum of operators $P = F^*F$ and $P_0 = F_0^*F_0$ only if $\mathrm{Re}\, F^*F_0 = 0$. The latter condition, which defines the incoherence relation between F and F_0, cannot be met for nonrandom amplitudes $F = (\varphi|$ and $F_0 = (\varphi_0|$ since $F^*F_0 = |\varphi)(\varphi_0| \ne 0$ even in the event of orthogonality $(\varphi \mid \varphi_0) = 0$ if $\varphi \ne 0$ or $\varphi_0 \ne 0$, although the total intensity (2.1) is equal to the sum $\|\varphi\|^2 + \|\varphi_0\|^2$.

Generally, the resulting density operator R can be represented in the form

$$R = P + P^{1/2}CP_0^{1/2} + P_0^{1/2}C^*P^{1/2} + P_0, \tag{2.5}$$

where $C = T^*T_0$ is the operator of mutual coherence of signal $F = TF^{1/2}$ and noise $F_0 = T_0P^{1/2}$ determined by the partial isometries $T : F^*\mathcal{K} \to \mathcal{H}$ and $T_0 : F_0^*\mathcal{K} \to \mathcal{H}$.

Note that C is a contracting operator:

$$\|C\| \le \|T^*\|\,\|T_0\| = 1,$$

and a partially isometric operator if $F_0\mathcal{H} \subseteq F\mathcal{H}$. The latter condition determines the coherence relation between the generalized amplitude F_0 and amplitude F, which is always met for nonrandom amplitudes $F = (\varphi|$ and $F_0 = (\varphi_0|$ for which $F_0\mathcal{H} = \mathbb{C} = F\mathcal{H}$. Representing the partially coherent amplitude F_0 in the form of a sum of the component $H_0 = TCP_0^{1/2}$ coherent with F and the component $W = F_0 - H_0$ that is incoherent and doing the same with the resulting amplitude G, or

$$G = H_1 + W,$$

where $H_1 = F + H_0$, we can isolate from operators P_0 and R a common density operator of the incoherent background $N = W^*W$ by writing the two operators, with allowance made for the fact that $W^*H_i = 0$, in the form $P_0 = S_0 + N$ and $R = S_1 + N$, where $S_i = H_i^*H_i$ are the operators $S_0 = P_0^{1/2}C^*CP_0^{1/2}$ and

$$S_1 = P + P^{1/2}US_0^{1/2} + S_0^{1/2}U^*D^{1/2} + S_0, \tag{2.6}$$

with U the partially isometric operator of polar expansion, and $CP_0^{1/2} = US_0^{1/2}$. In contrast to P_0 and R, for a trace class operator P the operators S_i are usually also trace class operators of rank $r(S_i) \le r(P)$ and one-dimensional operators if $r(P) = 1$.

Infinite trace operators P may also be replaced with trace class operators if we consider finite total intensities (2.4) with respect to a fixed D, assuming that $S = D^{1/2}PD^{1/2}$. The effective operator D is then replaced with the orthoprojector E on the subspace $\mathcal{E} = D\mathcal{H}$ that determines the total intensity $\varepsilon(S) = \mathrm{Tr}\ S$ by taking the trace $\varepsilon(S) = \mathrm{Tr}\,(ES)$ on $\mathcal{E}$.

2.1.2 Classical Detection

The simplest detection problem, that of isolating a pattern described by a kernel operator $P > 0$ from an incoherent mixture $R = P + N$ of this pattern with the background N, is solved by measuring the intensity of one of the possible signals, $R_0 = N$ or $R_1 = R$, by comparing this signal with the background level $\iota(N) = \mathrm{Tr}\, N$. To this end it has proved sufficient to limit oneself to measuring the total degree of contrast $\iota(C) = \mathrm{Tr}\, C$ of the received signal by calculating the trace $\langle C, E\rangle = \mathrm{Tr}\,(CE)$ of the appropriate operator $C_i = R_i - N$, $i = 0, 1$, on any subspace $\mathcal{E} = E\mathcal{H}$, $CE = C$, with the trace assuming finite values $\langle C_0, E\rangle = 0$ in the absence of a pattern, $i = 0$, and $\langle C_1, E\rangle = \mathrm{Tr}\, P$, $i = 1$, in the presence of a pattern even for an infinitely high level of the background $\iota(N) = \infty$.

In the case of a partially coherent superposition R of pattern P and background P_0, the difference $C = R - P_0$ may be a nonpositive trace class operator with a zero or even negative trace, with the result that the detection criterion, which is based on the condition that the total degree of contrast $\iota(C)$ is positive, may lead to incorrect results. Even if $\iota(C)$ is positive, which is the case when the superposition $\psi = \varphi_0 + \varphi$ of orthogonal amplitudes, $(\varphi \mid \varphi_0) = 0$, is coherent, that is,

$$\iota(C) = \|\psi\|^2 - \|\varphi_0\|^2 = \|\varphi\|^2\,,$$

we can considerably increase the degree of contrast of amplitudes φ_0 and ψ if we sum, say, the coordinate distribution of the degree of contrast,

$$c(x) = |\psi(x)|^2 - |\varphi_0(x)|^2 = |\varphi(x)|^2 + 2\,\mathrm{Re}\,\varphi^*(x)\varphi_0(x), \qquad (2.7)$$

not over the entire region Ω but only that part of the region where $c(x)$ is positive. As a result we arrive at the following classical problem of optimal detection of a pattern in a coordinate (frequency-spatial) region Ω: we must find a measurable subregion $\Delta^{\mathrm{o}} \subseteq \Omega$ in which the upper bound

$$\varkappa_I^{\mathrm{o}}(C) = \sup_{\Delta\subseteq\Omega} \langle C, I(\Delta)\rangle = \int_{\Delta^*} c(x)\,\mathrm{d}x \qquad (2.8)$$

of the integral of the contrast function $c(x) = C(x, x)$ is attained.

This function is determined by the diagonal values of the kernel $C(x', x)$, which is the difference between the generalized matrix elements $R(x', x)$ and $N(x', x)$ of operators R and N in the coordinate representation.

It is sufficient to consider the supremum (2.8) in the class of measurable subsets $\Delta \subseteq \Omega$ of the coordinate region $\Omega = \{x \in \mathbb{R}^{d+1} | c(x) \neq 0\}$, the support of the integrable function $c(x)$, in which the supremum is attained only on the set

$$\Delta^{\mathrm{o}} = \{x \in \Omega \mid c(x) > 0\} \equiv \Omega_+. \tag{2.9}$$

Its value, $\varkappa_I^{\mathrm{o}}(C) = \int_{\Omega+} c(x)\,\mathrm{d}x$, coincides, obviously, with the integral over Ω of the positive part

$$c_+(x) = \max\{0, c(x)\} = \tfrac{1}{2}(c(x) + |c(x)|), \tag{2.10}$$

where the functions c determine the solution to the duality problem

$$\langle c \rangle_+ = \inf_{b \geq 0} \left\{ \int_\Omega b(x)\,\mathrm{d}x \mid b \geq c \right\} = \int_\Omega c_+(x)\,\mathrm{d}x. \tag{2.11}$$

The lower bound (2.11) over all positive integrable functions $b(x) \in L^1_+(\Omega)$, majorizing almost everywhere the function c, is attained at $b^{\mathrm{o}} = 0 \vee c = c_+$ and determines on the space of integrable functions c a positive gauge $\langle c \rangle_+$, which is zero only when $c \leq 0$. The set (2.9) specifies the optimal band of the frequency-spatial filter in which the best quality of detection, (2.8), is achieved.

Reasoning along similar lines, we can solve the problem of optimal detection in the momentum (or temporal-wave) space $X = \mathbb{R}^{d+1}$,

$$\varkappa_I^{\mathrm{o}}(C) = \sup_{\Delta \subseteq X} \left\langle C, \widetilde{I}(\Delta) \right\rangle = \int_{\Delta^{\mathrm{o}}} \widetilde{c}(x)\,\mathrm{d}x, \tag{2.12}$$

where $\widetilde{c}(x) = \widetilde{C}(x, x)$ are the diagonal elements of the difference $\widetilde{R}(x', x) - \widetilde{N}(x', x)$ of the operators R and N in the momentum representation; in coherent superposition these diagonal elements are

$$\widetilde{c}(x) = |\widetilde{\psi}(x)|^2 - |\widetilde{\varphi}_0(x)|^2 = |\varphi(x)|^2 + 2\,\mathrm{Re}\,\widetilde{\varphi}^*(x)\widetilde{\varphi}_0(x). \tag{2.13}$$

The quality of such detection, $\varkappa_I^{\mathrm{o}}(C) = \left\langle \widetilde{C} \right\rangle_+$, based on a momentum quasi-measurement may differ considerably from (2.11). For example, the canonical amplitudes (1.25) $\varphi_0 = \psi_{00}$ and $\psi = \psi_{0\eta}$, which are similarly localized in the coordinate representation, differ by their momenta, $\eta \neq 0$, and can be thought of as two hypotheses, corresponding to the absence and presence of a complex-valued amplitude $\varphi = \psi_{0\eta} - \psi_{00}$ in the coherent superposition $\psi = \varphi + \varphi_0$, that cannot be distinguished by the measurement of

$$|\varphi_0(x)|^2 = |\psi(x)|^2$$

($c_+ = 0$ since $c(x) = 0$ for all $x \in \Omega$). At the same time, such wave packets are easily distinguished in the momentum representation:

$$\left\langle \widetilde{C} \right\rangle_+ = |\widetilde{\upsilon}|^{-1} \int (\mathrm{e}^{-\pi|(x-\eta)\upsilon^{-1}|^2} - \mathrm{e}^{-\pi|x\widetilde{\upsilon}^{-1}|^2})\,\mathrm{d}x \simeq 1$$

if $|\eta v^{-1}| \gg 1$ since in this case $\widetilde{c}_+(x) \simeq |\widetilde{\psi}_{0\eta}(x)|^2$.

In general, for every quasiselective measurement of intensity on a Borel space X with a positive operator measure $M(\Delta) \leq I, \Delta \subseteq X$, optimal detection is determined by the solution to the problem

$$\varkappa_M^{\mathrm{o}}(C) = \sup_{\Delta \subseteq X} \langle C, M(\Delta)\rangle = \varkappa(\Delta^{\mathrm{o}}) \tag{2.14}$$

of finding the upper bound of the degree-of-contrast measure $\varkappa(\Delta) = \langle C, M(\Delta)\rangle$. The supremum (2.14) is attained on the $|\varkappa|$-measurable set Δ^{o}, the support of the positive part

$$\varkappa_+ = 0 \vee \varkappa = (\varkappa + |\varkappa|)/2$$

of measure $\varkappa$:

$$\Delta^{\mathrm{o}} = \cap\{\overline{\Delta} : \varkappa_+(\Delta) = 0\} \equiv \Delta_+, \tag{2.15}$$

which realizes the lower bound in the positive measures $\lambda \geq \varkappa$ of finite variation:

$$\langle \varkappa \rangle_+ = \inf_{\lambda \geq 0}\{\lambda(X) \mid \lambda \geq \varkappa\} = \varkappa_+(X), \tag{2.16}$$

which determines the gauge $\langle \varkappa \rangle_+ = 0 \Leftrightarrow \varkappa \leq 0$ of measure $\varkappa$.

2.1.3 Optimal Detection

As the example in Section 2.1.2 shows, the quality of detection, which is determined for a given intensity distribution on X by the degree-of-contrast measure $\mu(C, \Delta) = \langle C, M(\Delta)\rangle$, must be optimized not only with respect to measurement regions $\Delta \subseteq X$ but also with respect to the methods of measurement of this quantity. These methods are determined by the ways in which the positive operator-valued measure $M(\Delta) \leq E$ is specified, where E is any orthoprojector in $\mathcal{H}$ satisfying the condition $CE = C$. Here it is sufficient to find at least one resolving operator $D = M(\Delta)$ that realizes the upper bound of the maximal degree of contrast (2.14):

$$\varkappa^{\mathrm{o}}(C) = \sup_{D \geq 0}\{\langle D, D\rangle \mid D \leq E\}. \tag{2.17}$$

Employing the methods of linear programming in partially ordered Banach spaces [43], we can formulate the necessary and sufficient conditions for the optimality of the detection operator D employing criterion (2.17), which is determined by the trace class degree-of-contrast operator $C = R - P_0$.

Theorem 1 *The upper bound* (2.17) *is attained on operator* $0 \leq D^{\mathrm{o}} \leq E$ *if and only if*

$$B^{\mathrm{o}}(E - D^{\mathrm{o}}) = 0, \quad (B^{\mathrm{o}} - C)\, D^{\mathrm{o}} = 0, \tag{2.18}$$

where $B^{\mathrm{o}} \geq 0, C$. *The operator* B^{o} *here is the solution to the duality*

$$\langle C \rangle_+ = \inf_{B \geq 0}\{\langle B, E\rangle \mid B \geq C\} \tag{2.19}$$

for which the conditions (2.18) *for admissible* D^{o} *are also necessary and sufficient, with* $\varkappa^{\mathrm{o}}(C) = \langle C \rangle_+$.

Proof. The sufficiency of the optimality conditions (2.18) for solving problems (2.17) and (2.19) can be verified directly by employing the property of the monotonicity of the trace,

$$B \geq C \Rightarrow \mathrm{Tr}\,(BD)\mathrm{Tr}\,(CD),$$

for every positive operator D. Allowing for the fact that $B^{\mathrm{o}}E = B^{\mathrm{o}}D^{\mathrm{o}} = CD^{\mathrm{o}}$ for every $0 \leq D \leq E$, we obtain

$$\langle C, D\rangle = \mathrm{Tr}\,(CD) \leq \mathrm{Tr}\,(B^{\mathrm{o}}D) \leq \mathrm{Tr}\,(B^{\mathrm{o}}E) = \langle C, D^{\mathrm{o}}\rangle\,.$$

Similarly, for every $B \geq 0$ and every C we obtain

$$\langle B, E\rangle = \mathrm{Tr}\,(BE) \geq \mathrm{Tr}\,(BD^{\mathrm{o}}) \geq \mathrm{Tr}\,(CD)^{\mathrm{o}} = \langle B^{\mathrm{o}}, E\rangle\,.$$

The necessity of the optimality conditions (2.18) follows from the fact that the inequality

$$\langle C, D\rangle = \mathrm{Tr}\,(CD) \leq \mathrm{Tr}\,(BD) \leq \mathrm{Tr}\,(BE) = \langle B, E\rangle\,, \tag{2.20}$$

which is valid for all operators D and B admissible in problems (2.17) and (2.19), must transform into the equality

$$\langle C, D^{\mathrm{o}}\rangle = \langle B^{\mathrm{o}}, E\rangle$$

on the extremal operators D^{o} and B^{o}, in accordance with Lagrange's principle of duality:

$$\begin{aligned} \sup_{D\geq 0}\{\langle C, D\rangle \mid D \leq E\} &= \sup_{D\geq 0}\inf_{R\geq 0}\{\langle C, D\rangle + \langle B, E - D\rangle\} \\ &= \inf_{E\geq 0}\sup_{D\geq 0}\{\langle C - B, D\rangle + \langle B, E\rangle\} \\ &= \inf_{B\geq 0}\{\langle B, E\rangle \mid B \geq C\}\,. \end{aligned}$$

Whereby, allowing for the fact that the trace of the product of positive operators is zero if and only if the product itself is zero, we arrive at conditions (2.18) via the following relation:

$$\mathrm{Tr}\,[B^{\mathrm{o}}(E - D^{\mathrm{o}})] + \mathrm{Tr}\,[(B^{\mathrm{o}} - c)D^{\mathrm{o}}] = \mathrm{Tr}\,(B^{\mathrm{o}}E) - \mathrm{Tr}\,(CD^{\mathrm{o}}) = 0.$$

The proof of the theorem is complete. ■

Note that the solutions to problems (2.17) and (2.19) exist for every Hermitian trace class operator C and every bounded positive orthoprojector E; the solution to problem (2.17) is unique only if E is the minimal of the orthoprojectors for which $CE = C$, while the solution to problem (2.19) is unique only if E is the maximal $E = I$ of the orthoprojector E. Indeed, employing the spectral representation of operator C, we write this operator in the form of the orthogonal sum

$$C = \sum \varkappa_n|\chi_n)(\chi_n| = C_+ + C_- \tag{2.21}$$

of the positive and negative operators

$$C_+ = \sum_{\varkappa_n>0} \varkappa_n|\chi_n)(\chi_n|, C_- = \sum_{\varkappa_n<0} \varkappa_n|\chi_n)(\chi_n|, \tag{2.22}$$

where we have allowed for the fact that a Hermitian trace class operator has a discrete spectrum of finite multiplicity, $\varkappa_n \in \mathbb{R}$, which can be found by solving the eigenvalue problem $C\chi = \varkappa\chi$. The orthoprojector E satisfying condition $CE = C$ can be written in the form of the orthogonal sum

$$E = E_+ + E_0 + E_-, \tag{2.23}$$

where $E_0 = E - E_+ - E_-$ with

$$E_+ = \sum_{\varkappa_n>0} |\chi_n)(\chi_n|, E_- = \sum_{\varkappa_n<0} |\chi_n)(\chi_n|.$$

The operators $D^o = E_+$, $B^o = C_+$ are, obviously, admissible: $0 \leq E_+ \leq E$, $C_+ \geq 0$, C and optimal:

$$\begin{aligned} C_+(E - E_+) &= C_+(E_0 + E) = 0, \\ (C_+ - C)E_+ &= -C_-E_+ = 0. \end{aligned} \tag{2.24}$$

Every other solution D^o to problem (2.17) satisfies conditions (2.18) for $B^o = C_+$;

$$\begin{aligned} C_+(E - D^o) &= C_+ - C_+D^o = 0, \\ (C_+ - C)D^o &= -C_-D^o = 0, \end{aligned}$$

in view of which $E_+ = E_+D^o$ and $E_-D^o = 0$, that is,

$$E_+ \leq D^o \leq E - E_- = E_+ + E_0. \tag{2.25}$$

Similarly, every solution B^o to problem (2.19) satisfies conditions (2.18) for $D^o = E_+$:

$$B^o(E - E_+) = 0, \quad (B^o - C)E_+ = B^oE_+ - C_+ = 0,$$

which imply that B is commutative with E_+ and, hence can be represented in the form of the orthogonal sum $B^o = B_+ + B_0$, with

$$B_+ = B^oE_+ = C_+, \quad B_0(E - E_+) = 0,$$

that is,

$$B^o = C_+ + B_0, \quad B_0 \geq 0, \quad B_0E = 0. \tag{2.26}$$

Thus, the general solution to the problem of optimal detection is determined by the quasifilter (2.25) of the form $D^o = E_+ + D_0$, where D_0 is an arbitrary operator, $0 \leq D_0 \leq E_0$, and an ideal filter $D^o = E_+$ if $E = E_+ + E_-$. The general solution to the duality problem (2.19) is determined by the operator of the form (2.26), with $B_0 = 0$ at $E = I$. The maximal possible degree of contrast realized by the optimal detector D^o is given by the expression

$$\varkappa_+(R - P_0) = \mathrm{Tr}\,(R - P_0)_+ = \sum_{\varkappa_n>0} \varkappa_n. \tag{2.27}$$

2.1.4 Coherent and Quasioptimal Detection

Let us consider the particular problem of optimal detection of a wave pattern described by a common amplitude $\varphi \in \mathcal{H}$ in a partially coherent mixture with a generalized random amplitude $H_0 : \mathcal{H} \to \mathcal{K}$. The resulting amplitude

$$G = |\xi)(\varphi| + H_0,$$

with $\xi \in \mathcal{K}$ a normalized vector $\|\xi\| = 1$, defines a density operator $R = G^*G$ of the form

$$R = |\varphi)(\varphi| + |\varphi)(\varphi_0| + |\varphi_0)(\varphi| + P_0, \tag{2.28}$$

with $\varphi_0 = F_0^*\xi \in \mathcal{H}$ and $P_0 = F_0^*F_0$ the background-density operator. Thus, we are required to solve the extremal problem (2.17) for the two-dimensional degree-of-contrast operator $C = R - P_0$ of the form

$$\begin{aligned} C &= |\varphi)(\varphi| + |\varphi)(\varphi_0| + |\varphi_0)(\varphi| \\ &= |\psi)(\psi| - |\varphi_0)(\varphi_0|, \end{aligned}$$

which corresponds to the coherent superposition $\psi = \varphi + \varphi_0$ of the common amplitudes φ and φ_0. We will consider this problem in the minimal subspace $\mathcal{E} \subset \mathcal{H}$ generated by the amplitudes $\psi_0 = \varphi_0$ and $\psi_1 = \varphi_0 + \varphi$. For its solution we find the eigenvectors and eigenvalues of operator C by constructing the secular equation $C\chi = \varkappa\chi$ for the coefficients of the expansion

$$\chi = \alpha_0\psi_0 + \alpha_1\psi_1$$

in the base $\{\psi_0, \psi_1\}$ of space $\mathcal{E}$:

$$\psi_1(\psi_1 \mid \alpha_0\psi_0 + \alpha_1\psi_1) - \psi_0(\psi_0 \mid \alpha_0\psi_0 + \alpha_1\psi_1) = \varkappa(\alpha_0\psi_0 + \alpha_1\psi_1). \tag{2.29}$$

Introducing the notation $\nu_i = \|\psi_i\|^2,\ i = 0, 1$, $\beta = (\psi_0 \mid \psi_1)$, and equating the coefficients of $\psi_i,\ i = 0, 1$, in (2.19), we arrive at a system of two homogeneous equations,

$$(\nu_0 + \varkappa)\alpha_0 + \beta\alpha_1 = 0, \quad \bar{\beta}\alpha_0 + (\nu_1 - \varkappa)\alpha_1 = 0. \tag{2.30}$$

This system has nonzero solutions only if the system determinant is zero, or

$$(\nu_0 + \varkappa)(\nu_1 - \varkappa) - |\beta|^2 = 0. \tag{2.31}$$

Solving this quadratic equation for $\varkappa$, we obtain the eigenvalues:

$$\varkappa_\pm = \frac{\nu_1 - \nu_2}{2} \pm \sqrt{\left(\frac{\nu_1 + \nu_0}{2}\right)^2 - |\beta|^2}, \tag{2.32}$$

which are real in view of the Schwarz inequality

$$|\beta|^2 = |(\psi_0 \mid \psi_1)|^2 \leq |\psi_0|^2|\psi_1|^2 = \nu_0\nu_1,$$

and, obviously, have opposite signs: $\pm\varkappa_\pm \geq 0$. At $\beta = 0$ the amplitudes ψ_1 and ψ_0 by measuring the degree of contrast of oscillations in the resulting mode $\chi_+ = \psi_1/\sqrt{\nu_1}$, which is equal to the intensity of oscillations at $\varkappa = \nu_1$ in this mode if the received signal is ψ_1 and to zero if the received signal is ψ_0. In the opposite case $|\beta|^2 = \nu_0\nu_1$ of the colinearity of ψ_1 and ψ_0, the values $\varkappa_\pm$ are equal respectively, to the positive and negative parts of the difference $\nu_1 - \nu_0$:

$$\varkappa_\pm = \tfrac{1}{2}(\nu_1 - \nu_0 \pm |\nu_1 - \nu_0|) = (\nu_1 - \nu_0)_\pm.$$

The corresponding optimal detection is reduced to the measurement of the positive degree of contrast $\varkappa_+ = \nu_1 - \nu_0$ in the mode $\chi = \psi_1/\sqrt{\nu_1} = \psi_0/\sqrt{\nu_0}$ if $\nu_1 > \nu_0$, in the opposite case, $\nu_0 \geq \nu_1$, the degree of contrast $\varkappa_+$ is zero and no measurement is carried out, or $\chi_+ = 0$. The optimal detection of a wave pattern φ of intensity $\mu = \|\varphi\|^2 \neq 0$ in the coherent superposition $\psi = \varphi + \varphi_0$ is therefore reduced to the measurement of the maximal degree of contrast

$$\varkappa_+ = \sqrt{\mu\nu_0}\left(\operatorname{Re}\gamma + \sqrt{\gamma} + \sqrt{(\operatorname{Re}\gamma + \sqrt{\gamma})^2 + 1 - |\gamma|^2}\right), \tag{2.33}$$

where $\gamma = (\varphi_0 \mid \varphi)/\sqrt{\mu\nu_0}$ is the coefficient of colinearity of amplitudes φ and φ_0, and λ is the signal-to-noise ratio. The corresponding ideal filter $E_+ = |\chi_+)(\chi_+|$ is defined at $\varkappa_+ \neq 0$ by the eigenvector $\chi_+ = \alpha_+\varphi + \alpha_0\varphi_0$ with coefficient

$$\alpha_+ = \sqrt{\nu_0/\mu\alpha_0}\left(\mathrm{j}\operatorname{Im}\gamma + \sqrt{\gamma} + \sqrt{(\operatorname{Re}\gamma + \sqrt{\lambda})^2 + 1 - |\gamma|^2}\right), \tag{2.34}$$

$\alpha_0 > 0$, found from the normalization condition $\|\chi_+\| = 1$. The case where $\varkappa_- = 0$, and therefore $\chi_+ = 0$, is possible in the minimal subspace $\mathcal{E}$ only if φ and φ_0 are colinear, when $|\gamma| = 1$, and

$$\varkappa_+ = \sqrt{\mu\nu_0}(\cos\theta + \sqrt{\lambda})_+, \tag{2.35}$$

where $\cos\theta = \operatorname{Re}\gamma$. The optimal filter in this case is matched with the signal mode $\chi_+ = \varphi/\sqrt{\mu}$ if $\cos\theta > -\sqrt{\lambda}$ and $\chi = 0$ in the opposite case if $\cos\theta \leq -\sqrt{\lambda}$ which is possible only if $\lambda \leq 1$.

The same filter $\chi_0 = \varphi/\sqrt{\mu}$ matched with ψ is used to describe the asymptotically optimal detection at large signal-to-noise ratios

$$\lambda = 1/\varepsilon \gg (\operatorname{Re}\gamma)^2.$$

The degree of contrast is then

$$\varkappa_0 = \mu(1 + \sqrt{\varepsilon}\operatorname{Re}\gamma), \tag{2.36}$$

which coincides with (2.33) to within ε. In the next order we obtain a filter matched with the resulting mode $\chi_1 = \psi/\sqrt{\nu_1}$ and realizing the degree of contrast

$$\varkappa_1 = \mu\left(1 + \frac{\sqrt{\varepsilon}}{2}\operatorname{Re}\gamma + \frac{\varepsilon}{4} - \frac{\varepsilon|\sqrt{\varepsilon}/2 + \gamma|^2}{4[1 + (\sqrt{\varepsilon}/2)\operatorname{Re}\gamma + \varepsilon/4]}\right). \tag{2.37}$$

For an orthogonal background, $\varphi \perp \varphi_0$, we have $\gamma = 0$, and the normalized eigenvector χ_+ corresponding to the eigenvalue

$$\varkappa_+ = \sqrt{\mu\nu_0}\left(\sqrt{1+\lambda}+\sqrt{\lambda}\right) \tag{2.38}$$

can be written in the form

$$\chi_+ = \frac{\left(\varphi_0/\sqrt{\nu_0}+\left(\sqrt{1+\lambda}+\sqrt{\lambda}\right)\varphi/\sqrt{\mu}\right)}{2(1+\lambda)+\sqrt{\lambda(1+\lambda)}}.$$

For $\operatorname{Im}\lambda \neq 1$ and a low signal-to-noise ratio $\lambda \ll 1-(\operatorname{Im}\gamma)^2$, the maximal degree of contrast is realized at

$$\chi_+ = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{\mu}}e^{j\theta}\varphi + \frac{1}{\sqrt{\nu_0}}\varphi_0\right), \tag{2.39}$$

with $\sin\theta = \operatorname{Im}\gamma$, and is determined asyptotically by the expression

$$\varkappa_+ \cong \sqrt{\mu\nu_0}\left(\operatorname{Re}\lambda + \cos\theta + \varepsilon\lambda(\operatorname{Re}\gamma+1)\right), \tag{2.40}$$

with $\cos\theta = \sqrt{1-(\operatorname{Im}\gamma)^2}$. For one, at $\gamma = 0$ we get $\varkappa_+ \cong \sqrt{\mu\nu_0}(1+\sqrt{\lambda})$.

In the general case of the partially coherent superposition $G = F + F_0$, the solution of the eigenvalue problem for the degree-of-contrast operator $C = G^*G - F_0^*F_0$ constitutes a complicated mathematical problem. If we isolate the coherent component from the generalized amplitude F_0, we can represent the latter in the form

$$F_0 = \tfrac{1}{2}\sqrt{\varepsilon}FA^* + W,$$

with W the incoherent component, $F^*W = 0$, and A is an operator in $\mathcal{H}$, which we assume to be bounded: $\|A\| \leq 1$. Next we select the positive constant ε in an appropriate manner. Operator C then assumes the form

$$C = P + \tfrac{1}{2}\sqrt{\varepsilon}(PA^* + AP) \tag{2.41}$$

and is a trace class operator if $P = F^*F$ is an operator with a finite trace.

For high signal-to-noise ratios $\lambda = 1/\varepsilon \gg 1$, the eigenvectors χ_n and the corresponding eigenvalues $\varkappa_n$ of operator C can be found via perturbation theory methods. In the first order in $\sqrt{\varepsilon}$ the eigenvectors coincide with the eigenvectors φ_n of the signal density operator P, that is, $P\varphi_{0n} = \mu_n\varphi_{0n}$, and realize the following degrees of contrast:

$$\varkappa_{0n} = (\varphi_{0n} \mid C\varphi_{0n}) = \mu_n(1+\sqrt{\varepsilon}\operatorname{Re}\gamma_n), \tag{2.42}$$

with $\gamma_n = (A\varphi_n \mid \varphi_n)$. The corresponding quasioptimal detection is reduced, therefore, to measuring the total degree of contrast

$$\varkappa_0 = \operatorname{Tr}(CE_0) = \sum_{n\in N_+}\mu_n(1+\sqrt{\varepsilon}\operatorname{Re}\gamma_n) \tag{2.43}$$

matched with the signal orthogonal modes φ_n of the ideal filter

$$E_0 = \sum_{n\in N_+} |\varphi_n)(\varphi_n|, \quad N_+ = \{n : \mathrm{Re}\,\gamma_n > -1/\sqrt{\varepsilon}\}. \tag{2.44}$$

When the intensities of the signal and the noise are comparable, $\varepsilon \approx 1$, the quality of such detection may be considerably lower than that of optimal detection. In particular, for the above example of orthogonal φ and φ_0 we have $\gamma = 0$ and $\varkappa_0 = \mu$, while the quality of optimal detection (2.28) equal to

$$\varkappa_+ = \mu(1 + \sqrt{1+\varepsilon}) > \mu$$

is more than twice as great as $\varkappa_0$ if the signal intensity is less than half of the intensity of the noise, and we have

$$\frac{\varkappa_+}{\varkappa_0} = \frac{1}{2}(1 + \sqrt{1+\varepsilon}) \to \infty \text{ as } \varepsilon = \frac{\nu_0}{4\mu} \to \infty. \tag{2.45}$$

2.2 Multialternative Detection of Wave Patterns

In this section we will consider the problem of detecting one of several simple or mixed wave patterns that is in a partially coherent super-position with the background. We will introduce the necessary and sufficient conditions for the optimality of such detection, using the criterion of the maximum degree of contrast, and give these conditions a concrete meaning for the problem of separating such patterns from an incoherent background. We will also give the complete solution to the problem of identifying nonorthogonal waves of the same intensity. This solution formally coincides with that of the problem of optimal discrimination between pure quantum states obtained in [14], [12]. Finally, we will discuss the quasioptimal method of multialternative detection based on the perturbation theory for degree-of-contrast operators.

2.2.1 Statement of the Pattern Identification Problem

The problem of m-alternative detection of sound and visual patterns described in a given spatial-frequency region Ω by the wave amplitudes $\varphi_i(q)$, $i = 1, \dots, m$ belonging to the Hilbert space $\mathcal{H} = L^2(\Omega)$ can be solved in a trivial manner in the absence of an audio or optical background, $\varphi_0 = 0$, only on the assumption that these amplitudes are pairwise orthogonal, $(\varphi_i \mid \varphi_k) = 0$ for $i \neq k$. It is sufficient to measure the intensity distribution $\varepsilon_i = |(\psi \mid \chi)|^2$ in the received signal $\psi \in \{\varphi_i\}_{i=1}^m$ over the orthogonal modes $\chi_k = \varphi_k/\|\varphi_k\|$, $i = 1, \dots, m$, to determine correctly the pattern $\varphi = \varphi_i$ with a nonzero intensity $\mu_i = \|\varphi_i\|^2$ by specifying the number of the excited mode $i : \varepsilon_i = \mu_i \neq 0$. The other modes χ_k, $k \neq i$, remain unexcited in the process, and the case $\varepsilon_i = 0$ for all $i = 1, \dots, m$ means that these patterns are absent from the measurement region Ω.

The simplest problem of multialternative detection in noise, the problem of isolating one of a set of orthogonal amplitudes $\{\varphi_i\}$ from an incoherent mixture

$$R_i = |\varphi_i)(\varphi_i| + N$$

with an optical or acoustic background not necessarily described by a trace class density operator N, has the same solution if we compare the intensities

$$\varepsilon_i = (\chi_i \mid R\chi_i) = \mu_i + \nu_i$$

of the received signal $R \in \{R_i\}_{i=0}^m$ not with zero but with the background level $\nu_i = (\chi_i \mid N\chi_i)$ in the orthogonal modes $\chi_i = \varphi_i / \|\varphi_i\|$, $i = 1, \ldots, m$.

In the case of nonorthogonal amplitudes $\{\varphi_i\}_{i=1}^m$ there is no way of measuring the intensity distribution in the received signal directly over the modes $\varphi_i / \|\varphi_i\|$. We must therefore find a set $\{\chi_i\}_{i=1}^m \subset \mathcal{H}$ that satisfies the condition

$$\sum_{i=1}^m |\chi_i)(\chi_i| \leq I$$

and for which a pattern φ_i can be confidently reconstructed from the distribution $\varkappa_i = |(\psi \mid \chi_i)|^2$ corresponding to the received signal $\psi \in \{\varphi_i\}_{i=1}^m$.

The same problem emerges in the multialternative detection of patterns $\{\varphi_i\}_{i=1}^m$ in the coherent superposition $\psi_i = \varphi_i + \varphi_0$ with a nonzero background amplitude φ_0 even when all the amplitudes $\{\varphi_i\}_{i=1}^m$ are mutually orthogonal and orthogonal to φ_0. Although in the latter case we can still measure the intensities in the orthogonal signal modes $\chi_i = \varphi_i / \|\varphi_i\|$ and this yields a total degree of contrast $\varkappa_0 = \sum_{i=1}^m \mu_i$, we can achieve a higher quality of detection if we minimize the expression

$$\varkappa = \sum_{i=1}^m (\chi_i \mid C_i \chi_i) = \sum_{i=1}^m |(\chi_i \mid \psi_i)|^2 - |(\chi_i \mid \varphi_0)|^2, \tag{2.46}$$

where

$$\begin{aligned} C_i &= |\varphi_i)(\varphi_i| + |\varphi_0)(\varphi_i| + |\varphi_i)(\varphi_0| \\ &= |\psi_i)(\psi_i| - |\varphi_0)(\varphi_0|, \end{aligned} \tag{2.47}$$

as we did in the case with $m = 1$ in the example at the end of Section 2.1.

Note that in the coordinate representation the wave patterns $\{\varphi_i\}_{i=1}^m$ may be indistinguishable even if they are orthogonal, as is the case, say, for harmonic amplitudes

$$\varphi_i(f) = \exp\{2\pi \mathrm{j} f i / \Phi\}$$

which are orthogonal in the frequency interval $[0, \Phi]$ and have the same homogeneous distributions, $|\varphi_i(f)|^2 = 1$. The maximal quality of m-alternative detection achievable through measurements of the coordinate distribution of the contrast degree

$$c_i(x) = |\varphi_i(x)|^2 + 2\,\mathrm{Re}\,\varphi_i^*(x)\varphi_0 = |\psi_i(x)|^2 - |\varphi_0(x)|^2 \tag{2.48}$$

is determined by the solution to the extremal problem

$$\varkappa_I^{\mathrm{o}}(C) = \sup_{\sum_{i=1}^m \Delta_i \subseteq \Omega} \sum_{i=1}^m \langle C_i, I(\Delta_i) \rangle = \sum_{i=1}^m \int_{\Delta_i^{\mathrm{o}}} c_i(x)\, \mathrm{d}x, \tag{2.49}$$

where the supremum is taken over the measurable nonintersecting subsets $\Delta_i \subset \Omega$ of a coordinate region Ω that can be bound by the union of the supports of the integrable functions $c_i(x)$, $i = 1, \ldots, m$. This limit is attained, obviously, on the partitions

$$\Omega_+ = \sum_{i=1}^m \Delta_i^{\mathrm{o}}$$

of the measurable set Ω_+ in every point of which at least one of the functions $c_i(x)$ is positive and coincides on Δ_c^{o} with the upper envelope

$$c_\vee(x) = \max_{i=1,\ldots,m} c_i(x).$$

Thus, the total degree of contrast (2.49) coincides with the integral over Ω of the positive part

$$c_+(x) = \max(0, c_\vee(x))$$

of $c_\vee$, which determines the solution of the duality problem

$$\langle c \rangle_+ = \inf_{b \geq 0} \left\{ \int_\Omega b(x)\, \mathrm{d}x | b \geq c_i,\ i = 1, \ldots, m \right\} = \int_\Omega c_+(x)\, \mathrm{d}x. \tag{2.50}$$

The lower bound (2.50) over all positive integrable functions $b(x) \geq 0$, which almost everywhere majorize every function $c_i(x)$, defines the positive gauge of the vector function $c = \{c_i\}_{i=1}^m$, $\langle c \rangle_+ = 0 \Leftrightarrow c_i(x) \leq 0$. The best m-alternative detection in the coordinate region Ω is reduced, therefore, to a search among the Δ_i for the regions Δ_i^{o} on which the measured degree of contrast $c(x)$ is positive and reaches $c_+(x)$; in the opposite case of $c(x) < c_+(x)$ the pattern may not be detected for all $x \in \Omega_+$ and it can be assumed to be undetected if $c(x) \leq 0$ for all $x \in \Omega$.

2.2.2 The Optimality Conditions

Let us consider the problem of maximizing the quality of m-alternative detection of mixed patterns $F_i : \mathcal{H} \to \mathcal{K}$ with trace class density operators $P_i = F_i^* F_i$, $i = 1, \ldots, m$, in a partially coherent superposition $G_i = F_i + F_0$ for which the mutual densities

$$F_i^* F_0 = \sqrt{\varepsilon} P_i A_i^* / 2$$

are determined by the contraction operators $A_i : \mathcal{H} \to \mathcal{H}$ ($\varepsilon > 0$ is a parameter). The respective extremal problem is formulated for trace class operators $\mathbf{C} = (C_i)_{i=1}^m$ of the degree of contrast,

$$C_i = R_i - P_0 = P_i + \frac{1}{2}\sqrt{\varepsilon}(P_i A_i^* + A_i P_i),\ i = 1, \ldots, m. \tag{2.51}$$

$R_i = G_i^* G_i$, in the class of quasiselective measurements described by any resolving operators $\{D_i\}_{i=1}^m$:

$$\varkappa^{\rm o}(\mathbf{C}) = \sup_{D_i \geq 0} \left\{ \sum_{i=1}^m \langle C_i,\ D_i \rangle \,\Big|\, \sum_{i=1}^m\ D_i \leq E \right\}, \tag{2.52}$$

where E is an operator for which $C_i E = C_i$ for all $i = 1, \ldots, m$.

Theorem 2 *The upper bound* (2.52) *is attained on the admissible operators* $D_i,\ i = 1, \ldots, m$, *if and only if there is a trace class operator* $B^{\rm o} \geq 0$, $C_i,\ i = 1, \ldots, m$, *such that*

$$B^{\rm o}(E - D^{\rm o}) = 0,\ (B^{\rm o} - C_i)D_i^{\rm o} = 0,\ i = 1, \ldots, m, \tag{2.53}$$

with $D^{\rm o} = \sum_{i=1}^m D_i^{\rm o}$. *The operator* $B^{\rm o}$ *then is the solution to the duality problem*

$$\langle \mathbf{C} \rangle_+ = \inf_{B \geq 0} \left\{ \langle B, E \rangle \,| B \geq C,\ i = 1, \ldots, m \right\}, \tag{2.54}$$

with $\varkappa^{\rm o}(\mathbf{C}) = \langle \mathbf{C} \rangle_+$, *for which conditions* (2.53) *are also necessary and sufficient when* $D_i^{\rm o} \geq 0$, $\sum_{i=1}^m D_i^{\rm o} \leq E$.

Proof. The proof, which is similar to the proof of a particular case of this theorem, Theorem 1, will be found as a corollary of a more general theorem, Theorem 5.

It can also be easily proved that the solution to problem (2.52) exists for all trace class operators C_i and determines on subspace $\mathcal{E} = E\mathcal{H}$ for which $C_i E = C_i$ for all $i = 1, \ldots, m$, a unique solution $B^{\rm o} = B^{\rm o} E$ to problem (2.54).

Indeed, the lower bound (2.54) determines for the vector operators $\mathbf{C} = (C_i)_{i=1}^m$ be the gauge $\langle \mathbf{C} \rangle_+$, a positive homogeneous sublinear functional on the space of families $(C_i)_{i=1}^m$ of the trace class operators $C_i,\ i = 1, \ldots, m$, that possesses the property $\langle \mathbf{C} \rangle_+ = 0 \Leftrightarrow C_i \leq 0$ for all i's. Bearing in mind that every linear functional $\mathbf{C} \to \langle \mathbf{C}, \mathbf{D} \rangle$ satisfying the condition $\langle \mathbf{C}, \mathbf{D} \rangle \leq \langle \mathbf{C} \rangle_+$ is positive and has the form

$$\langle \mathbf{C}, \mathbf{D} \rangle = \sum_{i=1}^m \mathrm{Tr}\,(C_i, D_i),$$

where $\sum_{i=1}^m D_i = E$, we find that the set that is conjugate to $\left\{ \mathbf{C} | \langle \mathbf{C} \rangle_+ \leq 1 \right\}$ consists of the resolving families $\mathbf{D} = \{D_i\}_{i=1}^m$ of bound operators that are admissible in problem (2.52). The existence of a solution to problem (2.52) follows, therefore, from the Hahn-Banach theorem, according to which for every vector $\mathbf{C}^{\rm o}$ of a calibrated space there exists a supporting functional $\mathbf{D}^{\rm o}$ defined by the conditions $\langle \mathbf{C}, \mathbf{D}^{\rm o} \rangle \leq \langle \mathbf{C} \rangle_+$ and $\langle \mathbf{C}^{\rm o}, \mathbf{D}^{\rm o} \rangle = \langle \mathbf{C}^{\rm o} \rangle_+$. For every solution $\mathbf{D}^{\rm o}$ to problem (2.52) the solution of the conjugate problem (2.54) on the subspace $\mathcal{E} = E\mathcal{H}$ is determined uniquely by the formula

$$B^{\rm o} E = B^{\rm o} D^{\rm o} = \sum_{i=1}^m C_i D_i^{\rm o} \tag{2.55}$$

which is obtained by adding (2.53) over $i = 1, \dots, m$. Note that the above proof of the existence of a solution to problem (2.52) and of the uniqueness of the solution to problem (2.54) remains valid for the case of an infinite number of patterns $m = \infty$ if we require that $\langle C_i, E \rangle = \operatorname{Tr} C_i \to 0$ as $i \to \infty$.

Conditions (2.53) can easily be met for $m > 1$ by analogy with the case of $m = 1$ only for commutative C_i, when these operators have a joint spectral representation

$$C_i = \sum_{i=1}^{m} \varkappa_{in} |\chi_n)(\chi_n|, \ (\chi_n \mid \chi_m) = \delta_{mn}. \tag{2.56}$$

The orthoprojector E can be resolved into an orthogonal sum $E = E_0 + \sum_{i=1}^{m} E_i$, where

$$E_i = \sum_{n \in \mathbb{N}_i} |\chi_n)(\chi_n|, \ \ \mathbb{N}_i \subseteq \{n \in \mathbb{N} | \varkappa_{in} \geq \{0, \ \varkappa_{kn} : \ k \neq i\}\}$$

(points n at which $\varkappa_{in} = \max_{j=1,\dots,m} \varkappa_{jn} = \varkappa_{kn}$ refer to any one of the nonintersecting sets $\mathbb{N}_i, \ \mathbb{N}_k$). The operators

$$D_i^{\mathrm{o}} = E_i, \ B^{\mathrm{o}} = \sum_{i=1}^{m} \sum_{n \in \mathbb{N}_i} \varkappa_{in} |\chi_n)(\chi_n| = \sum_{i=1}^{m} C_i E_i \tag{2.57}$$

are, therefore, admissible and optimal:

$$B^{\mathrm{o}}(E - D^{\mathrm{o}}) = B^{\mathrm{o}} E_0 = 0, \ (B^{\mathrm{o}} - C_i) E_i = C_i E_i - C_i E_i = 0.$$

Thus, optimal m-alternative detection in the commutative case is reduced to measuring the discrete distribution of the degree of contrast $\varkappa$ in the proper representation of operators C_i. The total maximal degree of contrast in this case is determined from the formula

$$\varkappa^{\mathrm{o}}(\mathbf{C}) = \sum_{i=1}^{m} \sum_{n \in \mathbb{N}_i} \varkappa_{in} = \sum_{n \in \mathbb{N}} \max_i \{\varkappa_{in} \vee 0\}. \tag{2.58}$$

■

2.2.3 Solution of Optimal Identification

Let us consider the important case of positive operators $C_i = H_i^+ H_i = S_i$ which occur, say, in the case of an incoherent superposition of wave patterns F_i with a background F_0, when the degrees of contrast (2.51) are the density operators $P_i = F_i^+ F_i$. The corresponding extremal problem (2.52) of pattern recognition, which is known as the optimal identification problem, is not trivial for noncommutative $S_i, \ i = 1, \dots, m$, for $m > 1$ even if these patterns are pure, that is, are described by nonorthogonal amplitudes $\psi_i, \ i = 1, \dots, m$. For the case of $m = 2$, however, the optimal identification problem can easily be solved by reducing it to the problem of optimal detection with one degree-of-contrast operator $C = S_1 - S_2$. Indeed, allowing for the fact that the admissible operators $B = L$ in the duality problem (2.54) are determined by the conditions $L \geq B > 0,\ i = 1, 2$, we can proceed from (2.54) to (2.19) by carrying out the substitution

$$\inf \langle L, E \rangle = \langle S_2, E \rangle + \inf \langle B, E \rangle$$

where $B = L - S_2$ is the admissible operator of problem (2.19):

$$B \geq \{S_1 - S_2, S_2 - S_2\} = \{C, 0\}.$$

Thus, the solution D^{o} to problem (2.17) makes it possible to represent the solution to problem (2.52) in the form

$$\varkappa^{\text{o}}(\mathbf{S}) = \langle C, D^{\text{o}} \rangle + \langle S_2, E \rangle = \langle S_1, D^{\text{o}} \rangle + \langle S_2, E - D^{\text{o}} \rangle\,,$$

which yields the optimal decision operators $D_1^{\text{o}} = D^{\text{o}}$ and $D_2^{\text{o}} = E - D^{\text{o}}$.

To investigate the problem of identifying wave patterns in the multialternative case with $m > 2$, we restrict the space $\mathcal{H}$ by the minimal space $\mathcal{E}^{\text{o}} \subseteq \mathcal{H}$ containing all the ranges of values $\mathcal{H}_i = S_i\mathcal{H}$. Since the $S_i \geq 0$, every operator B admissible to problem (2.54) is determined by the conditions $B \geq S_i$, $i = 1, \ldots, m$, in view of which it is nonsingular on the subspace $\mathcal{E}^{\text{o}}$ in the sense that $BD = 0 \Rightarrow D = 0$ for every operator D in $\mathcal{E}^{\text{o}}$. Otherwise, operator $D^+(B - S_i)D$ could be negative for at least one $i \in 1, \ldots, m$. This last fact means that the first condition in (2.53) is met only if $E = D^{\text{o}}$, that is, the optimal decision operators D_i^{o} determine the decomposition of unity $E^{\text{o}} = \sum_{i=1}^m D_i^{\text{o}}$, the orthoprojector on subspace $\mathcal{E}^{\text{o}}$, and operator B^{o} can be found uniquely by summation of the remaining optimality conditions in (2.53).

When the subspaces $\mathcal{K}_1 = H_i\mathcal{H}$ have a low dimensionality, say, ordinary amplitudes $H_i = (\psi_i|$ for which $\mathcal{K}_i = \mathbb{C}$, it has proved expedient to represent the solution to problem (2.52) via the following.

Theorem 3 *The optimal decision operators D^{o} determined by conditions* (2.53) *for $C_i = H_i^* H_i,\ i = 1, \ldots, m$, have the following form in space $\mathcal{E}^{\text{o}}$*

$$D_i^{\text{o}} = (L^{\text{o}})^{-1} H_i^* \mu_i^{\text{o}} H_i (L^{\text{o}})^{-1},\ i = 1, \ldots, m, \tag{2.59}$$

where $L^{\text{o}} = \left(\sum_{i=1}^m H_i^* \mu_i H_i\right)^{1/2} = B^{\text{o}}$ is the solution to problem (2.54), *and the μ_i are trace class positive operators in $\mathcal{K}_i$ defined by the conditions*

$$(1_i - H_i(L^{\text{o}})^{-1} H_i^*)\mu_i^{\text{o}} = 0,\ 1_i \geq H_i(L^{\text{o}})^{-1} H_i^* \tag{2.60}$$

(1_i *are the identity operators in $\mathcal{K}_i$). If these conditions are met, maximal intensity of graded signals $\varkappa^{\text{o}} = \sum_{i=1}^m \operatorname{Tr} \mu_i^{\text{o}}$ is achieved.*

Proof. Multiplying the remaining equations in (2.53) from the right by $(L^{\text{o}})^{1/2}$ and from the left by $(L^{\text{o}})^{-1/2}$, where $L^{\text{o}} = B^{\text{o}}$, we can rewrite the optimality conditions in the form

$$(E^{\text{o}} - F_i^* F_i) M_i^{\text{o}} = 0,\ E^{\text{o}} \geq F_i^* F_i,\ i = 1, \ldots, m, \tag{2.61}$$

where $F_i = H_i(L^{\circ})^{1/2}$ and

$$M_i^{\circ} = (L^{\circ})^{-1/2} D_i^{\circ} (L^{\circ})^{1/2}.$$

Thus,

$$M_i^{\circ} = F_i^* F_i M_i^{\circ} = M_i^{\circ} F_i^* F_i = F_i^* \mu_i^{\circ} F_i, \tag{2.62}$$

where $\mu_i^{\circ} = F_i M_i^{\circ} F_i^*$, which leads to (2.59) if we carry out the inverse transformation. If we substitute (2.62) into (2.61) and multiply the result from the right by F_i^* and from the left by $(F_i^*)^{-1}$, we arrive at (2.60) if we allow for the reversibility of the operators $F^* : \mathcal{K}_i \to \mathcal{E}$. The inequalities (2.60) are simply the inequality (2.61) in the form $F_i F_i^* \leq 1_i$. The operator L° is determined by summation $\sum_{i=1}^{m} D_i^{\circ} = E^{\circ}$ of the optimal decision operators (2.59), which yields $L^{\circ} = \sum_{i=1}^{m} H_i^{\circ} \mu_i^{\circ} H_i$, and this determines uniquely the positive operator $B^{\circ} = L^{\circ}$.

The proved theorem reduces the solution of the optimal identification problem to finding the operators μ_i° that satisfy conditions (2.60), which in the case of finitely mixed patterns H_i constitute finite-dimensional algebraic equations and inequalities. For one, for pure patterns $H_i = (\psi_i|$, conditions (2.60) have the scalar form

$$\mu_i^{\circ} = (\psi_i \mid (L^{\circ})^{-1} \psi_i) \mu_i^{\circ}, \; 1 \geq (\psi_i \mid (L^{\circ})^{-1} \psi_i), \; i = 1, \ldots, m. \tag{2.63}$$

where $L^{\circ} = (|\psi_i) \mu_i^{\circ} (\psi_i|)^{1/2}$. The numerical positive solutions of the system of algebraic equations (2.63) determine the one-dimensional decision operators

$$D_i^{\circ} = |\chi_i)(\chi_i|, \; \chi_i = (L^{\circ})^{-1} \psi_i \sqrt{\mu_i^{\circ}} \tag{2.64}$$

(which are equal to zero for those i's for which $(\psi_i \mid (L^{\circ})^{-1} \psi_i < 1)$ and the quality of the optimal solution, $\varkappa^{\circ} = \sum_{i=1}^{m} \mu_i^{\circ}$).

Solution of the pattern identification problem makes it possible to establish the quasioptimal multialternative detection scheme using the maximum criterion of the total degree of contrast (2.51) as the first approximation in $\sqrt{\varepsilon}$ for decision operators of the form

$$D_i = (F_{0i} + \sqrt{\varepsilon} F_{1i})^* (F_{0i} + \sqrt{\varepsilon} F_{1i}) = F_{\varepsilon i}^* F_{\varepsilon i}. \tag{2.65}$$

Assuming that $F_{0i} = \sqrt{\mu_i^{\circ}} H_i (L^{\circ})^{-1}$ and $D_{0i} = F_{0i}^* F_{0i}$, in the first order in the signal-to-noise ratio $\varepsilon \ll 1$ we obtain the following formula for the degree of contrast of quasioptimal detection $\varkappa_0 = \sum_{i=1}^{m} S_i D_{0i}$

$$\varkappa_0 = \sum_{i=1}^{m} \mathrm{Tr}_{\mathcal{K}_i} \mu_i^{\circ} (1 + \sqrt{\varepsilon}(\gamma_i + \gamma_i^*)/2), \tag{2.66}$$

where $\gamma_i = H_i A_i^* (L^{\circ})^{-1} H_i^*$, or $\gamma_i = (A_i \psi_i | (L^{\circ})^{-1} \psi_i)$ when $\mathcal{K}_i = \mathbb{C}$. ■

2.2.4 The Signal Representation

It has proved expedient to represent solution (2.59) to the problem of optimal identification of wave patterns in the so-called signal space, $\mathcal{K}^m = \bigoplus_{i=1}^m \mathcal{K}_i$, which is the direct sum of Hilbert spaces $\mathcal{K}_i = H_i\mathcal{H}$ and which, in the case of ordinary amplitudes $H_i = (\psi_i|$, is equal to $\mathbb{C}^m$. Such decomposition is carried out via the partially isometric operator $V : \mathcal{H} \to \mathcal{K}^m$ of the polar expansion $H = \sigma^{1/2}V$, $\sigma = HH^*$, for the operator $H : \varphi \in \mathcal{H} \mapsto [H_i\varphi]_{i=1}^m$ from $\mathcal{H}$ into $\mathcal{K}^m$, which is defined uniquely on the subspace $\mathcal{E}^{\rm o}$ by the conditions $V^*V = E^{\rm o}$ and $VV^* = \varepsilon^{\rm o}$, where $\varepsilon^{\rm o}$ is the support of the correlation matrix $\sigma = \sigma\varepsilon^{\rm o}$. Note that the m-by-m matrix $\sigma = [\sigma_{ik}]$ consisting of operator components $\sigma_{ik} = H_iH_k^*$, $i, k = 1, \dots, m$ ($\sigma_{ik} = (\psi_i|\psi_k)$ if $H_i = (\psi_i|$), is positive and, in the case of the linear independence of the signals H_i, nonsingular with support $\varepsilon^{\rm o} = \bigoplus_{i=1}^m 1_i \equiv 1^m$. The components $V_i = 1_iV$, $i = 1, \dots, m$, of the isometric operator $V : \mathcal{E}^{\rm o} \to \mathcal{K}^m$ determined by the diagonal projectors 1_i from $\mathcal{K}^m$ onto $\mathcal{K}_i$ bring about, obviously, the decomposition of the unit element

$$E^{\rm o} = V^*V = \sum_{i=1}^m V^*1_iV = \sum_{i=1}^m V_i^*V_i \tag{2.67}$$

of space $\mathcal{E}^{\rm o}$ and are orthogonal if $\varepsilon^{\rm o} = 1^m$:

$$V_iV_k^* = \varepsilon_{ik}^{\rm o} = 1_i\varepsilon^{\rm o}1_k = \delta_{ik}1_k.$$

Representing the operators H_i in the form $H_i = h_iV$, with $h_i = 1_i\sigma^{1/2}$, we can write the necessary and sufficient conditions for the optimality of the separating operators D_i in the following form:

$$(\lambda^{\rm o} - \sigma_i)\delta_i^{\rm o} = 0,\ \lambda^{\rm o} \geq \sigma_i := h_i^*h_i,\ i = 1, \dots, m, \tag{2.68}$$

which are simply conditions for the decomposition of the m-by-m projection matrix $\varepsilon^{\rm o} = \sum_{i=1}^m \delta_1^{\rm o}$, $\delta_i^{\rm o} = VD_i^{\rm o}V^*$. Theorem 3 in this case assumes the form of

Theorem 4 *The optimal decomposition of the support $\varepsilon^{\rm o}$ of the correlation matrix σ defined by conditions* (2.68) *has the form*

$$\delta_i^{\rm o} = \lambda^{{\rm o}-1}\, h\mu_i^{\rm o}h\lambda^{{\rm o}-1},\ i = 1, \dots, m, \tag{2.69}$$

where $h = \sigma^{1/2}$, $\lambda^{\rm o} = (h\mu^{\rm o}h)^{1/2}$, and $\mu^{\rm o} = \bigoplus_{i=1}^m \mu_i$ is a diagonal matrix $\mu^{\rm o} = [\mu_i^{\rm o}\delta_{ik}]$ consisting of the positive operators $\mu_i^{\rm o} : \mathcal{K}_i \to \mathcal{K}_i$ and defined by the conditions

$$\mu^{\rm o} = \epsilon(h\lambda^{{\rm o}-1}h)\mu^{\rm o},\ 1^m \geq \epsilon(h\lambda^{{\rm o}-1}h), \tag{2.70}$$

or $\mu^{\rm o} = \epsilon(\sqrt{\sigma\mu^{\rm o}})$ if σ is nonsingular ($\varepsilon^{\rm o} = 1$), where $\epsilon : a \mapsto \sum_{i=1}^m 1_ia1_i$ is the partial diagonalization operation $[a_{ik}] \mapsto [a_{ik}\delta_{ik}]$ of the m-by-m block-matrices $a = [a_{ik}]$ consisting of operators $a_{ik} : \mathcal{K}_k \to \mathcal{K}_i$. The quality of optimal identification is determined by the trace in $\mathcal{K}^m$, or $\varkappa^{\rm o} = \mathrm{Tr}\ \mu^{\rm o}$.

Proof. Representation (2.69) can be obtained directly via the isomorphism V of spaces $E^{\mathrm{o}}\mathcal{H}$ and $\varepsilon^{\mathrm{o}}\mathcal{K}^m$. Here $(\lambda^{\mathrm{o}})^{-1} = V(L^{\mathrm{o}})^{-1}V^*$, an m-by-m matrix with elements $(\lambda^{ki})^{-1} : \mathcal{K}_1 \to \mathcal{K}_k$, is the inverse of matrix $\lambda^{\mathrm{o}} = VL^{\mathrm{o}}V^*$ with respect to ε^{o}: $\lambda^{\mathrm{o}-1}\lambda^{\mathrm{o}} = \varepsilon^{\mathrm{o}} = \lambda^{\mathrm{o}}\lambda^{\mathrm{o}-1}$. Matrix $\lambda^{\mathrm{o}} = VL^{\mathrm{o}}V^*$ consisting of operator elements $\lambda^{\mathrm{o}}_{ik} : \mathcal{K}_k \to \mathcal{K}_i$ is directly expressible in terms of the square root $h = \sqrt{\sigma}$ of the correlation matrix

$$\lambda^{\mathrm{o}} = \left(\sum_{i=1}^{m} h_i^* \mu_i^{\mathrm{o}} h_i\right)^{1/2} = \sqrt{h\mu^{\mathrm{o}}h}, \tag{2.71}$$

while the conditions (2.60) for determining the operators μ_i^{o}, which in the signal representation have form

$$(1_i h_i \lambda^{\mathrm{o}-1} h_i^*)\mu_i^{\mathrm{o}} = 0, \quad 1_i \geq h_i \lambda^{\mathrm{o}-1} h_i^*, \tag{2.72}$$

represent the element-by-element notation for the conditions (2.70) imposed on the diagonal elements in $\mathcal{K}^m$. If σ is nonsingular (which means that h is nonsingular, too), we can rewrite (2.70) in the following simple form:

$$\mu^{\mathrm{o}} = \epsilon(h\lambda^{\mathrm{o}-1}h\mu^{\mathrm{o}}) = \epsilon(h\lambda^{\mathrm{o}}h^{-1}) = \epsilon(\sqrt{\delta\mu^{\mathrm{o}}}), \tag{2.73}$$

where we have allowed for the fact that $\epsilon(a)\mu^{\mathrm{o}} = \epsilon(a\mu^{\mathrm{o}})$ (because μ^{o} is diagonal) and that $\sigma\mu^{\mathrm{o}} = (h\lambda^{\mathrm{o}}h^{-1})^2$, in accordance with (2.70) can be resolved explicitly. Let us assume that the diagonal part $\epsilon(h)$ of matrix $h = \sigma^{1/2}$ is commutative with h. Then conditions (2.70) are met at $\mu^{\mathrm{o}} = \epsilon(\sqrt{\sigma})^2$, that is, at $\mu_i^{\mathrm{o}} = (h_{ii})^2$, $i = 1, \ldots, m$. Indeed, the diagonal matrix μ^{o} in this case is commutative with h and

$$\lambda^{\mathrm{o}} = \sqrt{h\mu^{\mathrm{o}}h} = \sqrt{h^2\mu^{\mathrm{o}}} = h\sqrt{\mu^{\mathrm{o}}} = h\epsilon(h) = \sqrt{\sigma}\epsilon(\sqrt{\sigma}).$$

Moreover, $h\lambda^{\mathrm{o}-1}h = h\epsilon(h)^{-1}$, where $\epsilon(h)^{-1}$ is the diagonal that is the inverse of $\epsilon(h)$, which always exists because the diagonal elements $\sigma_{ii} = H_i H_i^*$ of the correlation matrix σ are nonsingular and, hence, so are the diagonal elements h_{ii} of the matrix $h = \sqrt{\sigma}$ on the spaces $\mathcal{K}_i = H_i\mathcal{H}$. Thus,

$$\epsilon(h\lambda^{\mathrm{o}-1}h) = \epsilon(h\epsilon(h)^{-1}) = \epsilon(h)\epsilon(h)^{-1} = 1^m,$$

and conditions (2.70) are satisfied. The optimal decision operators then assume the form

$$\delta_i^{\mathrm{o}} = 1_i,\ D_i^{\mathrm{o}} = V^*1, = V_i^*V_i,\ i = 1, \ldots, m, \tag{2.74}$$

where $V = h^{-1}H$, while the quality of optimal separation is determined by the total intensity:

$$\varkappa^{\mathrm{o}} = \sum_{i=1}^{m} \mathrm{Tr}\,(h_{ii})^2 = \sum_{i=1}^{m} \mathrm{Tr}\,(\sigma_{ii}^{1/2})^2. \tag{2.75}$$

The above-noted property of commutativity manifests itself, for one thing, in the case where all diagonal operators σ_{ii} coincide and are multiples of the identity element $1_i = 1$ of space $\mathcal{K}_i = \mathcal{K}$, which is the same for all $i = 1, \ldots, m$. In view of the assumption that σ_{ii} is a trace class operator and, hence, $\mu_i^{\mathrm{o}} = (\sigma_{ii})^2$, this is possible only for a finite-dimensional $\mathcal{K}$. In Section 2.2.5 we consider concrete equidiagonal families of ordinary amplitudes $H_i = (\psi_i|$, for which $\mathcal{K} = \mathbb{C}$. ■

2.2.5 Separation of Cyclic Systems

Let $\{\psi_i\}_{i=1}^m$ be a family (or set) of nonorthogonal wave amplitudes $\psi_i \in \mathcal{H}$ that describe sound or visual patterns with a correlation matrix $\sigma = [(\psi_i|\psi_k)]$ whose square root, $h = \sqrt{\sigma}$, has the same diagonal elements $h_{ii} = a = h_{kk}$ for all $i, k = 1, \ldots, m$. The optimal identification of wave patterns $\{\psi_i\}$ is described by the one-dimensional separating operators $\{D_k^{\mathrm{o}}\}_{k=1}^m$ of the form (2.57), where $\chi_k^{\mathrm{o}} = C_k^*,\ k = 1, \ldots, m$, is generally an overcomplete system of polar decomposition,

$$\psi_k = \sum_{i=1}^m V_k^* \sigma_{ki}^{1/2} = \sum_{i=1}^m \chi_k^{\mathrm{o}} h_{ki},$$

$$\sum_{k=1}^m |\chi_k^{\mathrm{o}})(\chi_k^{\mathrm{o}}| = \sum_{k=1}^m V_k^* V_k = V^* V = E^{\mathrm{o}},$$

in the space $\mathcal{E}^{\mathrm{o}}$ induced by the set $\{\psi_k\}$. Bearing in mind that $\mu = a^2 = (\mathrm{Tr}\, h/m)^2$, we can represent the maximal intensity $\varkappa^{\mathrm{o}} = ma^2$ of optimally separated amplitudes $\{\chi_i^{\mathrm{o}}\}$ in the following invariant form:

$$\varkappa^{\mathrm{o}} = \frac{1}{m}(\mathrm{Tr}\, h)^2 = \frac{1}{m}(\mathrm{Tr}\, \sigma^{1/2})^2.$$

Let us consider the following example when the above-mentioned condition of the equidiagonality of matrix $h = \sqrt{\sigma}$ is met. We will call the system $\{\psi_i\}$ of amplitudes of equal intensity $\|\psi_i\|^2 = \nu$ equiangular if $(\psi_i|\psi_k) = \nu\gamma$ for every $i \neq k$, that is, if the cosines of all mutual angles are equal to γ. This is possible in the case when $\gamma \geq 1/(1-m)$, say, when $\psi_i = \varphi_0 + \varphi_i$, where $\{\varphi_i\}_{i=0}^m$ is an orthogonal system of amplitudes with intensities $\|\varphi_0\|^2 = \nu\gamma$ and $\|\varphi_i\|^2 = \nu(1-\gamma)$ at $i \neq 0$. Representing the respective correlation matrix σ in the form

$$\sigma = \nu((1-\gamma)1^m + \gamma x^\intercal x),\ x = (1, \ldots, 1) \in \mathbb{C}^m, \tag{2.76}$$

and using the formula

$$f(1^m + \tau x^\intercal x) = f(1)1^m + \frac{1}{xx^\intercal}[f(1+\tau xx^\intercal) - f(1))x^\intercal x]$$

to invert it and extract a square root, we can write out the optimal system $\{\chi_i^{\mathrm{o}}\}$ for $\gamma \in](1-m)^{-1}, 1[$ explicitly:

$$\chi_k^{\mathrm{o}} = \frac{1}{\sqrt{1-\gamma}}\left(\frac{1}{\sqrt{\mu}}\psi_k - (1-(1+\frac{m\gamma}{1-\gamma})^{-1/2})\frac{1}{m}\sum_{i=1}^m \frac{1}{\sqrt{\mu}}\psi_i\right).$$

The intensity of the signals separated by this orthogonal system is

$$\varkappa^{\mathrm{o}} = \nu(m - (1-\frac{1}{m})(\sqrt{1-\gamma+m\gamma} - \sqrt{1-\gamma}))^2,$$

and admits the maximal value $\varkappa^{\rm o} = m\nu$ in the event of orthogonality $\gamma = 0$ of the family $\{\psi_i\}$ and the value $\varkappa^{\rm o} = \mu$ in the case of colinearity of $\{\psi_k\}$.

For one, when the $\psi_i = |\alpha_i)$ are canonical equiangular amplitudes α_i defined by a $(d+1)$-by-$(d+1)$ matrix of the scalar products of vectors $\alpha_i \in \mathbb{C}^{d+1}$ of the form $\alpha_i^* \alpha_k^{\intercal} = \lambda\delta$ for $i \neq k$, $|\alpha_i|^2 = \lambda$ for all $i = 1, \ldots, m$, the quantity $\gamma = \exp\{\lambda(\delta - 1)\}$ does not vanish and the maximal intensity of optimal separation is always lower than $m\mu$ even if the vectors $\{\alpha_i\}$ are orthogonal ($\delta = 0$) and tends to $m\gamma$ only as $\lambda \to \infty$. Note that the maximal intensity of separation of canonical amplitudes is reached on simplex vectors $\alpha_i \in \mathbb{C}^{d+1}$ defined by the condition $\delta = (1-m)^{-1}$; for one thing, at $m = 2$ the intensity of separation of a pair of canonical amplitudes,

$$\varkappa^{\rm o} = \nu(1 - \sqrt{1-\gamma^2}) = 1 - \sqrt{1 - {\rm e}^{2\lambda(\delta-1)}},$$

can be attained at $\delta = -1$ by employing orthogonal vectors α_i, while at $\delta = 0$ this can be done only by doubling $\lambda = |\alpha_i|^2$.

Equiangular systems constitute a particular case of cyclic systems, which are defined by the condition that the correlation matrix σ_{ik} remain invariant under translations $s \in \mathbb{Z} : (i,k) \mapsto (i+s,\ k+s)$, that is, at $\sigma_{ik} = \sigma(i-k)$. Such translation invariant systems as containing only a finite number m of distinct amplitudes must satisfy also the cyclicity condition $\sigma(l) = \sigma(l+s)$ for $l = i - k < 0$. Since the matrix $h = \sqrt{\sigma}$, as any other matrix function of σ, also depends solely on the difference in the indices, or $h_{i,k} = h(i-k)$, the equidiagonality condition $h_{i,k} = a = h(0)$ is certain to be met and the solution to the problem of separating any cyclic system can be written explicitly.

Let us take the case of cyclic canonical amplitudes $\psi_k = |\alpha_k)$ defined by complex numbers $\alpha_k \in \mathbb{C}$ whose real and imaginary parts can be interpreted as the mean frequency and duration of the wave packet $|\alpha_i)$. There can be only two cases of the cyclicity of amplitudes $|\alpha_i)$ corresponding to the equidistant distribution of points α_i along a circle or a straight line with the center at $\alpha = 0$.

(1) *Optimal estimation of phase* Let $\alpha_i = \sqrt{\lambda}{\rm e}^{2\pi i {\rm j} k/m}, {\rm j} = \sqrt{-1}$. In this case we have a cyclic system

$$\sigma_{ik} = \exp\left\{\lambda({\rm e}^{-2\pi{\rm j}(i-k)/m} - 1\right\} = \sigma(i-k).$$

To extract the square root of matrix σ one should diagonalize it by a discrete Fourier transformation applying the unitary matrix

$$U_{in} = \exp\{2\pi i {\rm j} n/m\}/\sqrt{m},\ n = 0, \ldots, m-1.$$

A continuous analog of this problem, to which one can pass if m is sent to infinity, is the estimation of phase θ of the vector $\alpha_0 = \sqrt{\lambda}\exp\{2\pi{\rm j}\theta\}$ of the canonical amplitude $\psi_\theta = |\alpha_\theta)$ on the interval $[0,1]$. Diagonalizing matrix

$$\sigma_{x\theta} = (\alpha_x|\alpha_\theta) = \exp\left\{\lambda({\rm e}^{-2\pi{\rm j}(x-\theta)} - 1\right\}$$

via a discrete-continuous Fourier transformation $\mu_{xn} = \exp\{2\pi \mathrm{j} x n\}$, $n \in \mathbb{Z}$, we obtain its eigenvalues

$$\lambda_n = \lambda^n \mathrm{e}^{-\lambda}/n!,\ n = 0, 1, \ldots; \qquad \lambda_n = 0,\ n < 0.$$

The optimal system of decision vectors $\chi_x^{\mathrm{o}},\ x \in [0, 1]$, has the form

$$\chi_x^{\mathrm{o}} = \sum_{n=0}^{\infty} \mathrm{e}^{2\pi \mathrm{j} x n} |n),\ \text{where } |n) = \frac{1}{\sqrt{n!}} (A^*)^n |0),$$

with A^* the creation operator in $\mathcal{H} = L^2(\mathbb{R})$. It can easily be verified that the system χ_x^{o} defines the decomposition of unity,

$$I = \int_0^1 |\chi_x^{\mathrm{o}})(\chi_x^{\mathrm{o}}|\, \mathrm{d}x = \sum_{n,m=0}^{\infty} |n)(m| \int_0^1 \mathrm{e}^{2\pi \mathrm{j} x (n-m)}\, \mathrm{d}x = \sum_{n=0}^{\infty} |n)(n|,$$

but is not orthogonal.

(2) *Optimal estimation of amplitude.* Let us take $\alpha_i = i\Delta \mathrm{e}^{\mathrm{j}\theta}$, where $i \in \mathbb{Z}$, $\Delta > 0$, and $\mathrm{j} = \sqrt{-1}$. In this case the cyclicity condition is satisfied:

$$\sigma_{ik} = \exp\{-\Delta^2 (i-k)^2/2\} = \sigma(i-k).$$

The matrix $\sigma = [\sigma_{ik}]$ is diagonalized by the discrete-continuous Fourier transformation $U_{i\lambda} = \exp\{2\pi \mathrm{j} i \lambda\}$, $\lambda \in [0, 1]$. For $\Delta \ll 1$ the problem of optimal separation of the respective coherent amplitudes is reduced to the problem of optimal estimation of the real parameter $x \in \mathbb{R}$ of the coherent amplitude $\psi_x = |x \mathrm{e}^{\mathrm{j}\theta})$. This estimation is realized by measuring the intensity in the proper representation of the self-adjoint operator $\mathrm{Re}\, A \mathrm{e}^{-\mathrm{j}\theta}$ in space $\mathcal{H} = L^2(\mathbb{R})$. At $\theta = 0$ this is the frequency representation, while at $\theta = \pi/2$ it is the temporal representation.

2.3 Optimal Discrimination of Mixed Waves

In this section we will take up the problem of testing wave hypotheses based on measuring the appropriate intensity distributions. We will derive the necessary and sufficient conditions for optimal testing of such hypotheses by the minimum criterion of parasitic contrast at a fixed level of the received signal by employing a method of linear programming in partially ordered Banach operator spaces. In a specific case these conditions formally coincide with conditions obtained earlier in [53] on the optimality of quantum measurements by the minimum criterion for the error probability. A general geometric solution will be given for the case of a two-dimensional space, which is sufficient for describing the recognition of the polarization of a plane wave. This solution is similar to the solution of the problem of measuring quantum mechanical spin [14].

2.3.1 Wave Pattern Hypotheses

The problem of recognizing sound and visual patterns based on measurements of the intensity of the received audio or optical wave can be formulated within the framework of the wave theory of hypothesis testing discussed below.

Let $H_i,\ i = 1, \ldots, m$, be bounded operators from the Hilbert space $\mathcal{H}$ to another Hilbert space $\mathcal{K}$ describing the possible generalized random amplitudes at the "in" terminals of the receiver with density operators $S_i = H_i^* H_i$ with a trace $\operatorname{Tr} S_i < \infty$. The reader will recall that at $\mathcal{K} = \mathbb{C}$ the operators H_i correspond to ordinary amplitudes $\psi_i \in \mathcal{H},\ i = 0, \ldots, m$, which define the bounded functionals $H_i = (\psi_i| : \chi \in \mathcal{H} \mapsto (\psi_i|\chi)$. Each operator $H_i,\ i = 0, \ldots, m$, can be thought of as a hypothesis, according to which at the "in" terminals of the receiver there is one of the possible simple or mixed patterns $G_i,\ i = 1, \ldots, m$, in a partially coherent superposition $H_i = G_i + H_0$ with the absence of wave patterns G_i. The problem of m-alternative detection of wave patterns $G_i,\ i = 1, \ldots, m$, may, therefore, be considered as a problem of testing $m + 1$ hypotheses $H_i,\ i = 0, \ldots, m$, and vice versa.

The optimal testing of the hypotheses $H_i,\ i = 0, \ldots, m$ is determined by the solution to the problem of finding a quasiselective measurement $D = \{D_i\}_{i=0}^m$ that maximizes the quality functional

$$\varkappa(R, D) = \sum_{i=0}^{m} \langle R_i, D_i \rangle \,,\ D_i \geq 0,\ \sum_{i=0}^{m} D_i = E,$$

where $R = (R_i)_{i=0}^m$ are trace class operators with a common support $E : R_i E = R_i$ for all $i = 0, \ldots, m$, operators that are usually represented by linear combinations $R_i = \sum_{k=0}^m c_i^k S_k$ of density operators $S_i = H_i^* H_i$. For instance, in the problem of m-alternative optimal detection by maximum of the total contrast (2.52) criterion, the operators R_i are in effect the degrees of contrast $R_0 = 0$, $R_i = S_i - S_0 = C_i,\ i = 1, \ldots, m$, and the admissible operators $\{D_i\}_{i=0}^m$ are determined by the decision operators $D_i,\ i = 1, \ldots, m$ and $D_0 = E - D$, where $D = \sum_{i=1}^m D_i$. For the problem of discriminating between the hypotheses H_i we can consider more general criteria defined, say, by the operators

$$R_0 = \sum_{i=1}^{m} C_i,\ R_i = (1 + \lambda_i) C_i,\ \lambda_i \geq 0,\ i = 1, \ldots, m, \tag{2.77}$$

that appear in the problem of suppressing parasitic degrees of contrast $\langle C_i, D_k \rangle$ for $i \neq k$:

$$\tau^{\mathrm{o}}(\mathbb{C}) = \inf_{D_i \geq 0} \left\{ \sum_{i=1}^{m} \left\langle C_i, \sum_{k \neq i} D_k \right\rangle \Big| \langle C_i, D_i \rangle \geq \varepsilon_i,\ \sum_{i=1}^{m} D_i \leq E \right\} \tag{2.78}$$

under the condition that the useful degrees of contrast $\langle C_i, D_i \rangle,\ i = 1, \ldots, m$, are not lower than given levels ε_i. Indeed, if we solve the extremal problem

$$\begin{aligned} \varkappa^{\mathrm{o}}(R) &= \left\{ \sum_{i=0}^{m} \langle R_i, D_i \rangle \Big| \sum_{i=1}^{m} D_i = E \right\} \\ &= \sup_{D_i \geq 0} \Big\{ \sum_{i=1}^{m} \Big(\langle C_i, D_i \rangle (1 + \lambda_i) \\ &\qquad + \langle C_i, E - D \rangle \Big) \Big| \sum_{i=1}^{m} D_i \leq E \Big\}, \end{aligned} \tag{2.79}$$

for the operators $R_i = R_i(\lambda)$ defined in (2.77), we can write the solution to the problem (2.78) in the form

$$\begin{aligned} \tau^{\mathrm{o}}(\mathbb{C}) &= \sum_{i=1}^{m} \langle C_i, E \rangle + \sup_{\lambda_i \geq 0} \left\{ \sum_{i=1}^{m} \lambda_i \varepsilon_i - \varkappa^{\mathrm{o}}(R) \right\} \\ &= \sup_{\lambda_i \geq 0} \inf_{D_i \geq 0} \Big\{ \sum_{i=1}^{m} \Big(\Big\langle C_i, \sum_{k \neq i} D_k \Big\rangle \\ &\qquad + \lambda_i (\varepsilon_i - \langle C_i, D_i \rangle) \Big) \Big| \sum_{i=1}^{m} D_i \leq E \Big\}, \end{aligned} \tag{2.80}$$

provided that we employ Lagrange's method of multipliers $\lambda_i,\ i = 1, \ldots, m$.

Let us start with the classical variant

$$\begin{aligned} \varkappa_M^{\mathrm{o}}(R) &= \sup_{\Delta_i > 0} \left\{ \sum_{i=0}^{m} \langle R_i, M(\Delta_i) \rangle \Big| \sum_{i=0}^{m} \Delta_i = X \right\} \\ &= \sum_{i=0}^{m} \mu_i(\Delta_i^{\mathrm{o}}) \end{aligned} \tag{2.81}$$

of the problem (2.79) of optimal testing of hypotheses H in a fixed measurement described by the decomposition $E = \int M(dx)$ of an orthoprojector $E,\ R_i E = R_i$, on a Borel space X. This may be the coordinate selective measurement $M(dx) = I(dx),\ X = \Omega$, or x the momentum quasiselective measurement $M(dx) = \widetilde{I}(dx),\ X = \mathbb{R}^{d+1}$, or the canonical quasimeasurement $M(dx) = |x)(x|\mathrm{d}x,\ X = \mathbb{R}^{d+1}$, described in Section 1.2. The upper bound (2.81) in measurable partitions $X = \sum_{i=0}^{m} \Delta_i$ reaches the gauge

$$\langle \mu \rangle = \inf \{ \lambda(X) | \lambda \geq \mu_i,\ i = 0, \ldots, m \} = \mu_\vee(X) \tag{2.82}$$

of the family $(\mu_i)_{i=0}^m$ of measures $\mu_i(\Delta) = \langle C_i, M(\Delta) \rangle$, where the infimum is taken over all the measures of finite variation $|\lambda|(X) < \infty$ that majorize all μ_i.

Indeed, $\varkappa_M^{\mathrm{o}}(C) \leq \varkappa_\vee(\mu)$, since for every measurable partition $\Omega = \sum_{i=0}^{m} \Delta_i \subseteq X$, obviously,

$$\sum_{i=0}^{m} \mu_i(\Delta_i) \leq \sum_{i=0}^{m} \lambda(\Delta_i) = \lambda(\Omega) \leq \lambda(X).$$

The lower bound (2.82) is attained at the upper bound $\mu_\vee = \vee_{i=0}^m \mu_i$ of the family of measures $\{\mu_i\}_{i=0}^m$, defined as $\mu_\vee \geq \{\mu_i\}_{i=0}^m$, $\lambda \geq \mu_i \Rightarrow \lambda \geq \mu_\vee$, and is equal to the supremum (2.50) reached on the partitions $\Omega = \sum_{i=0}^m \Delta_i^\circ$ of the support $\Omega \subseteq X$ of measure $\mu_\vee$ into regions Δ_i°, on which it coincides with the respective measure μ_i :

$$\mu_\vee(\Delta_i^\circ) = \max_{k=0,\ldots,m} \mu_k(\Delta_i^\circ) = \mu_i(\Delta_i^\circ). \tag{2.83}$$

In view of the last relationship, determining a hypothesis H_i for a given measurement M is reduced to searching for the number of the nonempty region Δ_i° on which the measured degree of contrast reaches the envelope $\mu_\vee(\Delta_i^\circ)$ of the family $\{\mu_i\}$. However, this method does not enable us to find the wave patterns for which $\mu(\Delta_i^\circ) < \mu_\vee(\Delta_i^\circ)$ for all $i = 0, \ldots, m$.

2.3.2 Optimal Multialternative Testing

To obtain a satisfactory solution to the problem of wave pattern recognition one must look for the supremum (2.81) not only over the measurement regions Δ_i but also over the various methods of such a measurement, which are described by the resolving operators $D_i = M(\Delta_i)$. Thus, there emerges a nonclassical extremal problem (2.79), which may be considered as part of the conditional problem (2.78) of testing the hypotheses H_i in the degrees of contrast $\langle C_i, D_i \rangle$, which are compared with given level $\varepsilon_i, i = 1, \ldots, m$. The necessary and sufficient conditions for solving this problem are formulated in the following.

Theorem 5 *Theorem The upper bound* (2.79) *is attained on operators* $D_i^\circ, i = 0, \ldots, m$ *if and only if there is a trace class operator* $L^\circ \geq R_i,\ i = 0, \ldots, m,$ *such that*

$$(L^\circ - R_i) D_i^\circ = 0,\ i = 0, \ldots, m. \tag{2.84}$$

The operator L° *is then the solution to the duality problem*

$$\langle R \rangle_+ = \inf_L \{\langle L, E \rangle \,|L \geq R_i,\ i = 0, \ldots, m\} \tag{2.85}$$

for which conditions (2.84) *are also necessary and sufficient for* $D_i^\circ \geq 0$, $\sum_{i=0}^m D_i^\circ = E$, *and* $\varkappa^\circ(R) = \langle R \rangle_+$. *The solution* L° *to this problem for operators* $R_i = R_i^\lambda,\ i = 0, \ldots, m,$ *of the form* (2.77) *represents the solution to the constraint extremal problem* (2.78) *in the Lagrange form*

$$\tau^\circ(\mathbb{C}) = \sum_{i=1}^m (\langle C_i, E \rangle + \lambda_i^\circ \varepsilon_i) - \langle L^\circ, E \rangle \,, \tag{2.86}$$

where the parameters $\lambda_i^\circ \geq 0$ *are to be found from*

$$\lambda_i^\circ \left(\varepsilon_i - \langle C_i, D_i^\circ \rangle\right) = 0,\ \ \varepsilon_i \leq \langle C_i, D_i^\circ \rangle \,, i = 1, \ldots, m, \tag{2.87}$$

Proof. The sufficiency of the optimality conditions (2.84) for (2.83) and (2.85) can be verified directly by employing the property of monotonicity of the trace,

$$L \geq R_i \Rightarrow \mathrm{Tr}\,(LD_i) \geq \mathrm{Tr}\;(R_i D_i),$$

for $D_i \geq 0$. Allowing for the equality $L^{\mathrm{o}}E = \sum_{i=0}^{m} R_i D_i^{\mathrm{o}}$, which is obtained via summation of (2.84) over $i = 0, \ldots, m$, for every family $(D_i)_{i=0}^{m}$ admissible in (2.83) we have

$$\begin{aligned}\sum_{i=0}^{m} \langle R_i, D_i\rangle &= \sum_{i=0}^{m} \mathrm{Tr}\,(R_i D_i) \leq \sum_{i=0}^{m} \mathrm{Tr}\,(L^{\mathrm{o}} D_i) \\ &= \mathrm{Tr}\,(L^{\mathrm{o}}E) = \sum_{i=0}^{m} \langle R_i, D_i^{\mathrm{o}}\rangle\,.\end{aligned}$$

In a similar manner for every operator L admissible in (2.85) we have

$$\langle L, E\rangle = \mathrm{Tr}\,(LE) = \sum_{i=0}^{m} \mathrm{Tr}\,(LD_i^{\mathrm{o}}) \geq \sum_{i=0}^{m} R_i D_i^{\mathrm{o}} = \langle L^{\mathrm{o}}, E\rangle\,.$$

The necessity of the optimality conditions (2.84) follows from Lagrange's duality principle

$$\begin{aligned}&\sup_{D_i \geq 0} \left\{ \sum_{i=0}^{m} \langle R_i, D_i\rangle \,\Big|\, \sum_{i=0}^{m} D_i = E \right\} \\ = \;& \sup_{D_i \geq 0} \inf_{L} \left\{ \sum_{i=0}^{m} \langle R_i, D_i\rangle + \left\langle L, E - \sum_{i=0}^{m} D_i \right\rangle \right\} \\ = \;& \inf_{L} \sup_{D_i \geq 0} \left\{ \sum_{i=0}^{m} \langle R_i - L, D_i\rangle + \langle L, E\rangle \right\} \\ = \;& \inf_{L} \left\{ \langle L, E\rangle \,| L \geq R_i, i = 0, \ldots, m \right\},\end{aligned}$$

according to which $\sum_{i=0}^{m} \langle R_i, D_i\rangle = \varkappa^{\mathrm{o}}(R) = \langle R\rangle = \langle L^{\mathrm{o}}, E\rangle$ and

$$\sum_{i=0}^{m} \mathrm{Tr}\;[(L^{\mathrm{o}} - R_i)D_i^{\mathrm{o}}] = \mathrm{Tr}\,(L^{\mathrm{o}}E) - \sum_{i=0}^{m} \langle R_i, D_i^{\mathrm{o}}\rangle\,.$$

The necessary and sufficient condition for this sum of traces of products of positive operators to vanish is, obviously, Equation (2.84).

Employment of the duality principle in the conditional problem (2.78) reduces this problem by the elementary Lagrange method to problem (2.80), for which the necessity and sufficiency of conditions (2.87) can be verified directly. The proof of the theorem is complete. ■

Note that the above proof remains unchanged in the case of an infinite number of hypotheses, $m = \infty$. From this theory follows, for one thing, Theorem

2 if we put $R_0 = 0,\ R_i = C_i,\ i = 1, \dots, m$, and $L = B$. The existence of a solution to problem (2.83) and the uniqueness on the subspace $\mathcal{E} = E\mathcal{H}$ of the solution to problem (2.85) can be obtained from the proof in Section 2.2.2 of these assertions for problems (2.52) and (2.54) to which (2.83) and (2.85) are reduced by the substitutions $C_i = R_i - R_0$ and $B = L - R_0$.

In the case of positive R_i's the problem of testing the hypotheses $H_i, i = 0, \dots, m$, can be solved as a problem of separating $m-1$ signals $H_i = R_i^{1/2}$ by applying Theorems 3 and 4. For nonpositive R_i's it has also proved expedient to go over to the signal space $\mathcal{K}^{m+1} = \bigoplus_{i=0}^{m} \mathcal{K}_i,\ \mathcal{K}_i = H_i\mathcal{H},\ i = 0, \dots, m$, via a partially isometric operator $V : \mathcal{H} \to \mathcal{K}^{m+1}$ of polar decomposition $H = \sigma^{1/2}V, \sigma = HH^*$ for the operator $H : \varphi \in \mathcal{H} \mapsto [H_i\varphi]_{i=0}^{m}$. As a result, the optimality conditions for the decision operators D_i° can be written in the form of conditions imposed on the decomposition $\varepsilon = \sum_{i=0}^{m} \delta_i^{\circ}$, where $\delta_i^{\circ} = VD_i^{\circ}V^*$, of the support $\varepsilon^{\circ} = VV^*$ of the signal correlation matrix $\sigma_{ik} = H_iH_k^*$ with $i, k = 0, \dots, m$:

$$(\lambda_i^{\circ} - \rho_i)\,\delta_i^{\circ} = 0,\ \lambda_i^{\circ} \geq \rho_i = \sum_{k=0}^{m} h_k^* c_i^k h_k,\ i = 0, \dots, m.$$

Here $\lambda_i^{\circ} = VL^{\circ}V^*,\ h_i = 1_i h,\ h = \sqrt{\sigma}$ and c_i^k is the quality matrix, which defines the operators $R_i = \sum_{k=0}^{m} H_k^* c_i^k H_k$ and which, for a fixed m, it has proved expedient to consider as being a diagonal operator $c_i = \bigoplus_{k=0}^{m} c_i^k 1_k$ in space K^{m+1} because then the signal matrices $\rho_i = VR_iV^*$ can be represented in the form $\rho_i = hc_ih$.

Even if the amplitudes H_i are ordinary, that is, $H_i = (\psi_i|$ and hence the correlation matrix is a number matrix $\sigma_{ik} = (\psi_i|\psi_k)$, it is difficult to write conditions of optimality explicitly for $m > 1$ for a nonsingular matrix σ. Below we will study this problem for the case where the rank of matrix is equal to 2 and, hence, all the operators R_i, L, and D_i can be represented by 2-by-2 matrices in space $\mathcal{E}^{\circ} = \mathbb{C}^2$.

2.3.3 2-d Wave Pattern Recognition

To the operators $\{R_i\}$ in the optimization problem (2.84) we assign Hermitian matrices that can be considered nonnegative without loss of generality. Any 2-by-2 matrix can be decomposed in Pauli matrices, which are

$$\begin{aligned} 1 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\ \sigma_y = \begin{bmatrix} 0 & -\mathrm{j} \\ \mathrm{j} & 0 \end{bmatrix},\ \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \\ R &= \begin{bmatrix} \nu + z & x - \mathrm{j}y \\ x + \mathrm{j}y & \nu - z \end{bmatrix} = \nu + x\sigma_x + y\sigma_y + z\sigma_z = \nu + \mathbf{r}\cdot\boldsymbol{\sigma},\ \mathrm{j} = \sqrt{-1}, \end{aligned}$$

where x, y, z and ν are real if matrix R is Hermitian, and $\hat{\mathbf{r}} = x\sigma_x + y\sigma_y + z\sigma_z \equiv \mathbf{r}\cdot\boldsymbol{\sigma}$ is a vector operator represented by vector $\mathbf{r} = (x, y, z)$ of three-dimensional real space $\mathbb{R}^3$. The product of $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$, with $\mathbf{r} \in \mathbb{R}^3$ and $\mathbf{s} \in \mathbb{R}^3$, is equal to $\hat{\mathbf{r}}\hat{\mathbf{s}} = \mathbf{r}\cdot\mathbf{s}+\mathrm{j}(\mathbf{r}\times\mathbf{s})$, where $\mathbf{r}\cdot\mathbf{s}$ and $\mathbf{r}\times\mathbf{s}$ are the scalar and vector products of

$\mathbf{r}$ and $\mathbf{s}$. Note that Tr $R = 2\nu$ and Det $R = \nu^2 - |\mathbf{r}|^2$ (with $|\mathbf{r}| = \sqrt{\mathbf{r}\cdot\mathbf{r}}$), and that the nonnegativity condition $R \geq 0$ assumes the form $\nu = |\mathbf{r}| \neq 0$, and rank $R = 0$ at $\nu = 0$.

The operators $R_i = (\nu_i + \hat{\mathbf{r}}_i)/2,\ i = 0, \ldots, m$, where $|\nu_i| \geq |\mathbf{r}_i|$ and $\sum_{i=0}^m \nu_i = 1$, can be interpreted as density operators related to the tested wave hypotheses with prior intensities $\nu_i =$ Tr R_i and represented by vectors $\mathbf{r}_i \in \mathbb{R}^3$, which are known as polarization vectors. A similar problem arises when we must identify the photon polarization or the electron spin [14]. Let us assume that polarizations $\{\mathbf{r}_i\}$ satisfy the inequalities

$$|\mathbf{r}_k - \mathbf{r}_i| > |\nu_k - \nu_i| \tag{2.88}$$

for all $k \neq i$; in the opposite case, that is, $|\nu_k - \nu_i| \geq |\mathbf{r}_k - \mathbf{r}_i|$, the k-th hypothesis dominates the i-th hypothesis or vice versa: $R_k > R_i$ or $R_k = R_i$ or $R_k < R_i$, and one of the hypotheses (with the smaller ν) can be ignored.

The decision operators D_i in the Pauli representation $D_i = \delta_i + \hat{\mathbf{d}}_i$ are described by nonnegative numbers $\delta_i \geq 0$ and vectors $\mathbf{d}_i \in \mathbb{R}^3$ ($|\mathbf{d}_i| \leq \delta_i$), with the decomposition of unity $\sum_{i=0}^m D_i = 1$ assuming the form

$$\sum_{i=0}^m \delta_i = 1,\ \sum_{i=0}^m \mathbf{d}_i = 0.$$

The solution of the problem of optimal recognition of polarizations $\mathbf{r}_i$ with intensities ν_i can be reduced to finding a real number λ_i° and a vector $\mathbf{l}^{\circ} \in \mathbb{R}^3$ that defines the operator $L^{\circ} = (\lambda_i^{\circ} + \hat{\mathbf{l}}^{\circ})/2$ that satisfies conditions (2.84) for a collection of $D_i^{\circ} = \delta_i^{\circ} + \hat{\mathbf{l}}_i^{\circ}, i = 0, \ldots, m$.

Theorem 6 *The solution to the problem of optimal recognition of polarizations* $\{\mathbf{r}_i\}$ *satisfying together with* $\{\nu_i\}$ *condition* (2.88) *can be found if and only if there is a collection of numbers* $\mu_i^{\circ} \geq 0,\ i = 0, \ldots, m$ *such that*

$$\left|\sum_{k=0}^m (\mathbf{r}_i - \mathbf{r}_k)\mu_k^{\circ}\right| + \sum_{k=0}^m (\nu_i - \nu_k)\mu_k^{\circ} \geq 1,\ i = 0, \ldots, m, \tag{2.89}$$

where the equality takes place at least for those i*'s for which* $\mu_i^{\circ} \neq 0$. *The optimal decision operators have the form* $\delta_i^{\circ} = |\mathbf{d}_i^{\circ}|$, $\mathbf{d}_i^{\circ} = \mu_i^{\circ}(\mathbf{r}_i - \mathbf{l}_i^{\circ})$, *where*

$$\mathbf{l}^{\circ} = \sum_{i=0}^m \mu_i^{\circ}\mathbf{r}_i \Big/ \sum_{i=0}^m \mu_i^{\circ}, \tag{2.90}$$

and the maximal received intensity is

$$\varkappa^{\circ} = \left(1 + \sum_{i=0}^m \mu_i^{\circ}\nu_i\right) \Big/ \sum_{i=0}^m \mu_i^{\circ} = \lambda_i^{\circ}. \tag{2.91}$$

Proof. In terms of $\lambda_i^{\text{o}}, \mathbf{l}^{\text{o}}, \nu_i$, and $\mathbf{r}_i$, the first optimality condition (2.84), $L^{\text{o}} - R_i \geq 0$, has the form

$$\lambda_i^{\text{o}} \geq |\mathbf{r}_i - \mathbf{l}^{\text{o}}| + \nu. \tag{2.92}$$

The equation $(L^{\text{o}} - R_i)D_i^{\text{o}} = 0$ can be written in another form if we nullify the scalar and real vector parts of the product $(\lambda_i^{\text{o}} + \nu_i - \hat{\mathbf{l}}^{\text{o}} - \hat{\mathbf{r}}_i)(\delta_i^{\text{o}} + \hat{\mathbf{d}}_i)$:

$$(\lambda_i^{\text{o}} - \nu_i)d_i^{\text{o}} + (\mathbf{l}^{\text{o}} - \mathbf{r}_i)\delta_i^{\text{o}} = 0,\ (\lambda_i^{\text{o}} - \nu_i)\delta_i^{\text{o}} + (\mathbf{l}^{\text{o}} - \mathbf{r}_i)\cdot \mathbf{d}_i^{\text{o}} = 0. \tag{2.93}$$

The imaginary vector equation $\mathrm{j}(\mathbf{l}^{\text{o}} - \mathbf{r}_i) \times \mathbf{d}_i^{\text{o}} = 0,\ \mathrm{j} = \sqrt{-1}$, follows from the real vector equation in (2.93) are equivalent to

$$\mathbf{d}_i^{\text{o}} = \delta_i^{\text{o}}(\mathbf{r}_i - \mathbf{l}^{\text{o}})/(\lambda_i^{\text{o}} - \nu_i),\ ((\lambda_i^{\text{o}} - \nu_i)^2 - |\mathbf{r}_i - \mathbf{l}^{\text{o}}|)^2\delta_i^{\text{o}} = 0, \tag{2.94}$$

in the opposite case ($\lambda_i^{\text{o}} = \nu_i$ for a certain i in(2.92)) for $i = k$ we obtain $\mathbf{l}^{\text{o}} = \mathbf{r}_i$, with the result that inequalities (2.94) and (2.88) become incompatible. The optimal decision vector can be written in the form (2.90), where $\mu_i^{\text{o}} = \delta_i^{\text{o}}/(\lambda_i^{\text{o}} - \nu_i)$ is nonnegative in accordance with the inequalities $\delta_i^{\text{o}} \geq 0,\ \lambda_i^{\text{o}} > \nu_i$, and (2.92), while $\mathbf{l}^{\text{o}}$ is determined by the set $\{\mu_i^{\text{o}}\}$ in accordance with the fact that $\sum_{i=0}^{m} \mathbf{d}_i^{\text{o}} = 0$. The second equation in (2.94) implies that inequalities (2.92) become amplitudes for the values of i for which $\delta_i^{\text{o}} = \mu_i(\lambda - \nu_i) \neq 0$. Multiplying (2.92) by $\sum_{i=0}^{m} \mu_i^{\text{o}}$ and finding λ in the form

$$\lambda_i^{\text{o}} = \left(1 + \sum_{i=0}^{m} \mu_i^{\text{o}}\nu_i\right) \Big/ \sum_{i=0}^{m} \mu_i^{\text{o}}$$

from the condition that $\sum_{i=0}^{m} \delta_i^{\text{o}} = 1$ for

$$\delta_i^{\text{o}} = \mu_i^{\text{o}}(\lambda - \nu_i) = \mu_i^{\text{o}}|\mathbf{r}_i - \mathbf{l}^{\text{o}}| = |\mathbf{d}_i^{\text{o}}|$$

we get condition (2.89) for determining $\{\mu_i^{\text{o}}\}$. Since Tr $L^{\text{o}} = \lambda_i^{\text{o}}$, the maximal intensity of (2.85) is equal to (2.91). The proof of the theorem is complete. ■

Note that the equalities in (2.89) must be true for at least two indices i and k, since there is no such set $\{\mu_i\}$, $\mu_i \neq 0$, for only one subscript i that satisfies the i-th inequality. For every pair $\mathbf{r}_i$ and $\mathbf{r}_k$ for which (2.88) is valid there is a unique solution of the i-th and k-th equalities in (2.89) with $\mu_j = 0$ for all $j \neq i,\ k$ with $\mu_i > 0,\ \mu_k > 0$:

$$\mu_i = (|\mathbf{r}_k - \mathbf{r}_i| + \nu_k - \nu_i)^{-1},\ \mu_k = (|\mathbf{r}_i - \mathbf{r}_k| + \nu_i - \nu_k)^{-1},$$

but such a set $\{\mu_i\}$ may not satisfy the other inequalities in (2.89) for $j \neq i,\ k$. If there exists a pair $\mathbf{r}_i, \mathbf{r}_k$ for which all the inequalities in (2.89) are valid at $\mu_j \neq 0$ only when $j = i, k$, then the optimal decision vectors $\mathbf{d}_j^{\text{o}}$ are zero for $j \neq i, k$ (see (2.90)) and

$$\mathbf{d}_i^{\text{o}} = (\mathbf{r}_i - \mathbf{r}_k)/2|\mathbf{r}_i - \mathbf{r}_k|,\ \mathbf{d}_k^{\text{o}} = (\mathbf{r}_k - \mathbf{r}_i)/2|\mathbf{r}_k - \mathbf{r}_i|.$$

Here the optimal decision operators $D_i^{\text{o}} = |\mathbf{d}_i^{\text{o}}| + \hat{\mathbf{d}}_i^{\text{o}}$ are orthogonal and correspond to an error intensity

$$\varkappa^{\text{o}} = \frac{1}{2}(\nu_i + \nu_k) + \frac{1}{2}|\mathbf{r}_i - \mathbf{r}_k|.$$

In the case where the optimal operators D_i^{o} are nonzero for more than two i's, they define a nonorthogonal decomposition of unity in the two-dimensional space $\mathcal{E} = \mathbb{C}^2$. We will not try to find a general analytical solution to the system of equation (2.89) with $\mu_i^{\mathrm{o}} \neq 0$ for more than two i's; rather, we will give a geometric interpretation of such a solution.

2.3.4 Geometric Representation of 2-d Patterns

Let us represent the Hermitian operators (2.86) by points $r = (\nu, x, y, z) = (\nu, \mathbf{r})$ in the four-dimensional Minkowski space $\mathbb{R}^{1+3}$. To every nonnegative operator there corresponds a point inside the light cone $\nu = |\mathbf{r}|$. In these terms *a priori* neither the k-th nor the i-th hypothesis is dominant at $R_i = (\nu_i + \hat{\mathbf{r}}_i)/2$ and $R_k = (\nu_k + \hat{\mathbf{r}}_k)/2$ if and only if the interval $r_i - r_k = (\nu_i - \nu_k, \mathbf{r}_i - \mathbf{r}_k)$ is spacelike. In accordance with (2.92), point $l^{\mathrm{o}} = (\lambda_i^{\mathrm{o}}, \mathbf{l}^{\mathrm{o}})$, which represents the operator $L^{\mathrm{o}} = (\lambda^{\mathrm{o}} + \hat{\mathbf{l}}^{\mathrm{o}})$, is the apex of the four-dimensional cone

$$\mathcal{C}(l) = \{r = (\nu, \mathbf{r}) : \nu - \lambda_i^{\mathrm{o}} + |\mathbf{r} - \mathbf{l}^{\mathrm{o}}| = 0\} \tag{2.95}$$

covering all the points $r_i = (\nu_i, \mathbf{r}_i)$ and containing the subset $\{r_{j_\alpha}\} \subset \{r_i\}$ of the boundary points r_{j_α} satisfying (2.92). On the other hand, the optimal points l are only those whose projections $\mathbf{l}$ belong to the convex hull of the boundary subset of the spatial projections $\mathbf{r}_{j_\alpha}, \alpha = 0, \ldots, s,\ s \leq m$:

$$\sum_{\alpha=0}^{s} \mathbf{r}_{j_\alpha} \pi_{j_\alpha} = \mathbf{l},\ \sum_{\alpha=0}^{s} \pi_{j_\alpha} = 1, \tag{2.96}$$

where, in accordance with (2.90), $\pi_{j_\alpha} = \mu_{j_\alpha} / \sum_{\alpha=0}^{s} \mu_{j_\alpha} \geq 0$ ($\mu_i = 0$ if r_i is covered by cone (2.95): $\nu_i + |\mathbf{r}_i - \mathbf{l}| \leq \lambda$). We will say that the subset $\{r_{j_\alpha}\}$ has an apex if the points r_{j_α} lie on the cone: $r_{j_\alpha} \subset \mathcal{C}(l)$ with an apex l whose spatial projection $\mathbf{l}$ is a point on the convex hull $\{\mathbf{r}_{j_\alpha}\}$. In these terms Theorem 6 can be formulated as follows:

Theorem 7 *To solve the problem of optimal recognition of points* $r_i = (\nu_i, \mathbf{r}_i),\ i = 0, \ldots, m$, *separated by spacelike intervals* (2.88), *it is necessary and sufficient to find a subset* $\{\mathbf{r}_{j_\alpha}\}$ *with an apex* l^{o} *belonging to a cone that covers all other points of set* $\{r_i\}$, *that is, to specify a subset of vectors* $\mathbf{r}_{j_\alpha} \subset \{\mathbf{r}_i\},\ \alpha = 0, \ldots, s$ *whose convex hull contains vector* $\mathbf{l}^{\mathrm{o}}$ *with respect to which the sum* $|\mathbf{r}_{j_\alpha} - \mathbf{l}^{\mathrm{o}}| + \nu_{j_\alpha}$ *is the constant* λ_i^{o}*:*

$$|\mathbf{r}_{j_\alpha} - \mathbf{l}^{\mathrm{o}}| + \nu_{j_\alpha} = \lambda_i^{\mathrm{o}},\ \alpha = 0, \ldots, s, \tag{2.97}$$

while $|\mathbf{r}_i - \mathbf{l}^{\mathrm{o}}| + \nu_i \leq \lambda_i^{\mathrm{o}}$ *for all other indices* $i \in \{j_\alpha\}$. *The optimal decision operators are represented by points on the cone*

$$d_i^{\mathrm{o}} = (\delta_i^{\mathrm{o}}, \mathbf{d}_i^{\mathrm{o}}),\ \ \delta_i^{\mathrm{o}} = |\mathbf{d}_i^{\mathrm{o}}|,$$

with spatial vectors

$$\mathbf{d}_i^{\mathrm{o}} = \pi_i^{\mathrm{o}}(\mathbf{r}_i - \mathbf{l}^{\mathrm{o}}) \Big/ \sum_{i=0}^{m} \pi_i \mathbf{r}_i, \tag{2.98}$$

where $\pi_i^o = 0$ *for* $i \notin \{j_\alpha\}$*, and* $\left\{\pi_{j_\alpha}^o, \alpha = 0, \ldots, s\right\}$ *is any nonnegative solution to the system of equations* (2.96)*. The minimal intensity in this case is*

$$\varkappa^o = \sum_{i=0}^{m} (\nu_i + |\mathbf{r}_i - \mathbf{l}^o|)\pi_i. \tag{2.99}$$

Note that every pair of point r_i, r_k separated by a spacelike interval defines, via two equations from (2.97), $j_\alpha = i, k$, a set of point $\mathbf{l} \in \mathbb{R}^{1+3}$ whose difference of distances to the points $\mathbf{r}_i$ and $\mathbf{r}_k$ is constant:

$$|\mathbf{l}^o - \mathbf{r}_k| - |\mathbf{l}^o - \mathbf{r}_i| = \nu_i - \nu_k. \tag{2.100}$$

These points lie on one of the two sheets of the hyperboloid of revolution with foci at $\mathbf{r}_i$ and $\mathbf{r}_k$ and eccentricity

$$\varepsilon = \frac{|\mathbf{r}_i - \mathbf{r}_k|}{|\nu_i - \nu_k|} > 1.$$

Here, if $\nu_i = \nu_k$, the hyperboloid (2.100) becomes a plane normal to the segment $\mathbf{r}_i\pi_i + \mathbf{r}_k\pi_k$ $(\pi > 0,\ \pi_i + \pi_k = 1)$ at point $(\mathbf{r}_i + \mathbf{r}_k)2$, while if $\nu_i \neq \nu_k$, we select the sheet in whose plane lies the focus with the higher intensity, ν_i or ν_k. Obviously, if the subset $\{r_{j_\alpha}\}$ has an apex l, the spatial projection $\mathbf{l}$ is the common point of all the hyperboloids (2.100) corresponding to all the pairs of the set $\{r_{j_\alpha}\}$ that belong to the convex hull $\{\mathbf{r}_{j_\alpha}\}$. We will call this point $\mathbf{l}$ the center of $\{\mathbf{r}_{j_\alpha}\}$ representing L^o is unique, which means that the center of $\{\mathbf{r}_{j_\alpha}\}$ is unique, too. It can easily be shown that for every vector $\mathbf{l}$ of the convex hull $\{\mathbf{r}_{j_\alpha}\}$ the system of linear equations (2.96) has a unique solution $\left\{\pi_{j_\alpha}^o\right\}$ if and only if vectors $\mathbf{r}_{j_\alpha} - \mathbf{r}_{j_0}$, $\alpha = 1, \ldots, s$, are linearly independent.

2.3.5 Optimal and Simplex Solutions

The reader will recall that a convex hull of a set $\{\mathbf{r}_{j_\alpha}\}$ of points $\mathbf{r}_{j_\alpha}$, $\alpha = 0, 1, 2, 3,$ is called an s-simplex (a segment if $s = 1$, a triangle if $s = 2$, a tetrahedron if $s = 3$, and so on) if the vectors $\mathbf{r}_{j_0}, \mathbf{r}_{j_\alpha}$, $a = 1, \ldots, s$, are linearly independent. It is well-known that each s-dimensional face (an s-face) of an n-simplex $(n \geq s)$ is a simplex, too. We will call a subset that generates a simplex convex hull a simplex subset.

Theorem 8 *The problem of optimal recognition of polarizations* $\{\mathbf{r}_i, i = 0, \ldots, m\}$ *always has a solution that can be described by the simplex set* $\left\{\mathbf{d}_{i_\alpha}^o, \alpha = 0, \ldots, s\right\}, s \leq m$*, of the nonzero vectors* (2.98) *corresponding to the simplex subset* $\{\mathbf{r}_{j_\alpha}\} \subseteq \{\mathbf{r}_i\}$ *with a center at* $\mathbf{l}^o$ *and a maximal sum*

$$\nu_{j_\alpha} + |\mathbf{r}_{j_\alpha} - \mathbf{l}^o| = \max_{i=0,\ldots,m} \{\nu_i + |\mathbf{r}_i - \mathbf{l}^o|\}.$$

This solution is unique if and only if the s-simplex generated by subset $\{\mathbf{r}_{j_\alpha}\}$ *is an s-face of the convex hull of all vectors* $\mathbf{r}_{j_0}, \ldots, \mathbf{r}_{j_m}$ *with a common center* $\mathbf{l}^o$*.*

Proof. By Theorem 7, the solution to the problem considered here is reduced to finding the cone (2.95) that covers all points $\{\mathbf{r}_i\}$ and has an apex $\mathbf{l}^{\text{o}}$ with a projection lying inside the convex hull of projections $\{\mathbf{r}_{j_\alpha}\}$ of the tangency points r_{j_α}. Obviously, there is always such a cone. Let $n \le m$ be the number of tangency points $r_{j_\alpha}, \alpha = 0, \dots, n$. If the subset $\{\mathbf{r}_{j_\alpha}, \alpha = 0, \dots, s\}$ $(s \le m)$ is a simplex set, the validity of Theorem 2.80 is obvious. If this subset is not a simplex, then the convex hull $\{\mathbf{r}_{j_\alpha}\}$ can be partitioned into several simplexes with a common vertex $\mathbf{r}_{j_0}$ via diagonal planes $(\mathbf{r}_{j_0}, \mathbf{r}_{j_\alpha}, \mathbf{r}_{j_\beta})$ or diagonal lines $(\mathbf{r}_{j_0}, \mathbf{r}_{j_\alpha})$ when all the vectors $\mathbf{r}_{j_\alpha}$ are coplanar. Hence, the center $\mathbf{l}^{\text{o}}$ is an interior point of one of the s-simplexes $(s < n \le m)$ with apexes $\mathbf{r}_{j_\alpha}, \alpha = 0, \dots, s$, which are the projections of the tangency points and determine the unique positive solution $\left\{\pi^{\text{o}}_{j_\alpha}\right\}$ of system (2.96). The set $\left\{\mathbf{d}^{\text{o}}_{j_\alpha}\right\}$ of nonzero vectors (2.98) is a simplex set if and only if the set $\{\mathbf{r}_{j_\alpha}\}$ is a simplex and determines the optimal solution with maximal quality (2.97) and minimal error intensity (2.99). When center $\mathbf{l}^{\text{o}}$ is an interior point of a nonsimplex convex hull of the projections of tangency points, the optimal simplex solution is not unique (the partition into simplexes is note unique) and there are also optimal nonsimplex solutions. But if point $\mathbf{l}^{\text{o}}$ is a boundary point of the convex hull, that is, an interior point of an s-face, the optimal solution is unique if the face is an s-simplex. ∎

Corollary 9 *To solve the problem of optimal testing of several hypotheses in the two-dimensional space $\mathcal{E} = \mathbb{C}^2$, it is sufficient to limit oneself to $s + 1 \le 4$ solutions $j_0, \dots, j_s$ corresponding to a simplex subset of hypotheses $\mathbf{r}_{j_0}, \dots, \mathbf{r}_{j_s}$. Each such solution procedure can be realized in an indirect measurement described by an orthogonal decomposition in the observation space $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$.*

Indeed, in the three-dimensional space $\mathbb{R}^3$ there is not a single simplex subset $\mathbf{r}_{j_0}, \dots, \mathbf{r}_{j_s}$ for $s > 3$ and, therefore, for every m there always exists an optimal decomposition in the two-dimensional space $\mathcal{E}$ consisting of $s + 1 \le 4$ nonzero decision operators $D^{\text{o}}_{j_\alpha} = \delta^{\text{o}}_{j_\alpha} + \hat{\mathbf{d}}^{\text{o}}_{j_\alpha}$ of rank one. It is well known that each nonorthogonal decomposition of unity in operators $D_{j_0}, \dots, D_{j_s}$ of rank one can be extended to an orthogonal decomposition in an $(s+1)$-dimensional space $\mathcal{E} \subset \mathcal{H}$. Hence, we can limit ourselves to the four-dimensional measurement space $\mathcal{H}$, which can always be represented as the tensor product of two-dimensional spaces $\mathcal{E} : \mathcal{H} = \mathcal{E} \otimes \mathcal{E}$, corresponding to the composition of two identical systems.

Note that the optimal solution may be degenerate (in the sense that a hypothesis $\mathbf{r}_i$ may correspond to $D_i = 0$) even if the set $\mathbf{r}_0, \dots, \mathbf{r}_m$ is a simplex set $(m \le 3)$, for example, at $m = 2$, $\nu_0 = \nu_1 = \nu_2$, if the vectors $\mathbf{r}_0, \mathbf{r}_1$ and $\mathbf{r}_2$ form an obtuse triangle.

In conclusion of this section we will consider two particular cases.

(1) *Optimal recognition of pure polarization.* Here the polarizations are normalized to prior intensities: $|\mathbf{r}_i| = \nu_i$, with the representative points $r_i = (\nu_i, r_i)$ belonging to the cone $\nu = |\mathbf{r}|$. Expression (2.97), which determines the subset of points $\mathbf{r}_{j_\alpha}$ of tangency of this cone and the covering cone (2.95), has the form

$$|\mathbf{r}_{j_\alpha} - \mathbf{l}^{\text{o}}| + |\mathbf{r}_{j_\alpha}| = \lambda^{\text{o}}_i. \tag{2.101}$$

In relation to $\mathbf{r}_{j_\alpha}$, this is the equation of an ellipsoid of revolution with foci at 0 and $\mathbf{l}^{\mathrm{o}}$ and eccentricity $\varepsilon = |\mathbf{l}^{\mathrm{o}}|/\lambda_i^{\mathrm{o}} < 1$. In accordance with (2.92), all the other point $\mathbf{r}_i \notin \{\mathbf{r}_{j_\alpha}\}$ must lie inside the ellipsoid. Hence, the problem of optimal recognition of pure polarization is reduced to finding the ellipsoid described about points $\{\mathbf{r}_i\}$ with foci at 0 and $\mathbf{l}^{\mathrm{o}}$, where $\mathbf{l}^{\mathrm{o}}$ is an interior point of the convex hull of the points of tangency $\{\mathbf{r}_{j_\alpha}\}$. The quality $\varkappa^{\mathrm{o}}$ of the optimal solution is equal to the length of the major axis of the ellipsoid, λ_i^{o}.

(2) *Optimal recognition of equi-intensity polarizations.* The priori intensities $\nu_i = \nu_0,\ i = 1, \ldots, m$, and the corresponding points are points of the hyperplane $\nu = \nu_0$. The density operators $R_i = (\nu + \mathbf{r}_i)/2$, all having the same trace ν_0, are represented by normalized vectors, $|\mathbf{r}_i| \leq \nu_0$. The intersection of the covering cone (2.95) and the hyperplane $\nu = \nu_0$ is a sphere $|\mathbf{r} - \mathbf{l}^{\mathrm{o}}| = \rho$ of radius $\rho = \lambda_i^{\mathrm{o}} - \nu_0$. Hence, the problem of optimal recognition of equiprobable polarizations is reduced to finding a sphere described about all points $\mathbf{r}_i : |\mathbf{r} - \mathbf{l}^{\mathrm{o}}| \leq \rho$ with radius ρ and centered at $\mathbf{l}^{\mathrm{o}}$, the center belonging to the convex hull of the tangency points $\mathbf{r}_{j_\alpha} : |\mathbf{r}_{j_\alpha} - \mathbf{l}^{\mathrm{o}}| = \rho$. The radius $\rho = \lambda_i^{\mathrm{o}} - \nu_0$ determines the maximal intensity (2.91):

$$\varkappa^{\mathrm{o}} = \rho + \nu_0 = \lambda_i^{\mathrm{o}} \tag{2.102}$$

($\rho \leq \nu_0$ since $|\mathbf{r}_i| \leq \nu_0$ for all i's). The minimum of intensity (2.102) is obtained at $\rho = \nu_0 : \varkappa^{\mathrm{o}} = 2\nu_0$. This corresponds to the typical equiprobable case $|\mathbf{r}_i| = 1$, when there is at least one simplex subset $\{\mathbf{r}_{j_\alpha}\}$ for which the center $\mathbf{l}^{\mathrm{o}} = 0$ is an interior point of the simplex.

Chapter 3

Efficient Estimation of Wave Patterns

In this section we develop the noncommutative theory of efficient measurements and optimal estimation of unknown parameters of wave patterns as applied to problems of sound and visual pattern recognition. We consider two variants of the lower bounds for the variance of the measured parameters, the variants being based on noncommutative generalizations [35], [18], [52] of the Rao-Cramér inequality [49], and introduce the notion of canonical states [17], for which we derive generalized uncertainty relations similar to the quantum mechanical uncertainty relations [30], [36], [18]. We then establish the necessary and sufficient conditions for efficient measurements, conditions that extend the conditions of efficiency of quantum mechanical measurements obtained in [17], [18] to the case of classical wave signals and fields. We formulate the necessary and sufficient conditions for the optimality of generalized measurements of wave patterns, conditions that generalize the respective conditions for quantum systems, see for example [37]. Finally, we investigate the structure of optimal covariant measurements for symmetric wave patterns, which in the case of quantum symmetric fields has been studied in [36], [13]. The exposition is based largely on the works of Belavkin [?], [13], [15], [18].

3.1 Wave Patterns Variances and Invariant Bounds

We will consider two variants of the lower bound for the variance of parameters of wave patterns, both based on noncommutative generalizations [35], [18], [52] of the Rao-Cramér inequality. In contrast to [35], [52], these bounds are represented in a form invariant under diffeomorphisms, the form will be used to obtain generalized uncertainty relations and efficient measurements of canonical parameters.

3.1.1 Classical Variance Bound

In Section 2.2 we considered the problem of recognizing pure or mixed wave patterns taken from a given finite or denumerable standard family. Generally, sound and visual patterns may contain unknown parameters that run through an infinite set of continuous values $\theta \in \Theta$ of finite or denumerable dimensionality. For example, we may not know the mean frequency and the moment when the sound signal appears or the mean position and the wave number of the visual pattern, or we may *a priori* have no information on the expected amplitudes of the oscillations in the given finite or denumerable family of standard modes.

It is natural to estimate the unknown parameters by the intensity distributions in the representations in which the wave packets with distinct values of θ are clearly separated; for example, the frequency and position can be calculated as the mean values on the coordinate representation, while the mean time of arrival of a signal and the wave number of a wave packet can be calculated in the momentum representation (but not vice versa).

We will call a family of wave packets described in a Hilbert space $\mathcal{H}$ by amplitudes $\{\psi_\theta\}$ resolvable in a representation defined by the decomposition of unity $I = \int M(dx)$ on a given Borel space X if there exists a measurable map $\hat{\theta} : X \to \Theta$ satisfying the condition

$$\int \hat{\theta}(x)\mu_\theta(dx) = \theta \int \mu_\theta(dx) \quad \forall \theta \in \Theta, \tag{3.1}$$

where $\mu_\theta(dx) = (\psi_\theta \mid M(dx)\psi_\theta)$ is the respective intensity distribution on X. Thus, the resolvability of the family $\{\psi_\theta\}$ means that it is possible to calculate the unknown $\theta \in \Theta$ in a given representation given the observed distribution of the μ_θ as the mean values of a function $\hat{\theta}(x)$ known as the unbiased estimator of parameters θ.

It is natural to define the quality of the resolvability of family $\{\psi_\theta\}$ by the size of the variance of the unbiased estimator $\hat{\theta}$, assuming that the quality for a given θ is all the higher the smaller the standard deviation from θ in the distribution induced on Θ by the measure μ_θ of the wave packet ψ_θ. To find the lower bound for this variance, we can use the classical Rao-Cramér inequality [49] if the measure μ_θ possesses the appropriate differentiability properties in θ. For the sake of simplicity we take the case of one parameter $\theta \in \mathbb{R}$. If we assume that there exists a second moment for the logarithmic derivative $\hat{\gamma}_\theta = \partial \ln \mu_\theta / \partial\theta$, which is the Radon-Nikodym derivative of measure $\mu'_\theta = \partial\mu_\theta/\partial\theta$ defined by

$$\mu_\theta(dx)\hat{\gamma}_\theta(x) = \partial\mu_\theta(dx)/\partial\theta, \tag{3.2}$$

we can easily obtain the inequality

$$\int (\hat{\theta}(x) - \theta)^2 \mu_\theta(dx) \cdot \int (\hat{\gamma}_\theta(x) - \gamma_\theta)^2 \mu_\theta(dx) \geq J_\theta^2, \tag{3.3}$$

where $J_\theta = \int \mu_\theta(dx)$ is the total intensity of the wave packet ψ_θ, and γ_θ is the mean value of $\hat{\gamma}_\theta$:

$$J_\theta \gamma_\theta = \int \hat{\gamma}_\theta(x)\mu_\theta(dx) = J'_\theta = 2\,\mathrm{Re}(\psi_\theta \mid \psi'_\theta).$$

Inequality (3.3), which implies the inverse proportionality of the standard deviation $\sigma_{\hat{\theta}} \geq 1/\sigma_{\hat{\gamma}}$ or variances

$$\sigma^2_{\hat{\theta}} = \int (\hat{\theta}(x) - \theta)^2 \mu_\theta(dx)/J_\theta,\ \sigma^2_{\hat{\gamma}} = \int (\hat{\gamma}_\theta(x) - \theta)^2 \mu_\theta(dx)/J_\theta \tag{3.4}$$

follows in an obvious manner from the Schwarz inequality if we allow for the fact that the right-hand side can be represented, in accordance with (3.1), in the form of the square of the scalar product

$$J_\theta = \int \hat{\theta}(x)\mu'_\theta(dx) - \theta J'_\theta = \int (\hat{\theta}(x) - \theta)(\hat{\gamma}_\theta(x) - \gamma_\theta)\mu_\theta(dx).$$

In a more general situation, where the estimated parameters $\theta = [\theta^i]_{i=1}^m$ are differentiable functions $\theta(\alpha)$ of unknown parameters $\alpha = [\alpha^k]_{k=1}^n$ of the density operators of mixed wave patterns $S(\alpha)$, we can easily obtain a matrix Rao-Cramér inequality that is invariant with respect to the choice of the state parameters:

$$R \geq D\sigma^{-1}_{\hat{\gamma}\hat{\gamma}} D^\intercal, \tag{3.5}$$

where $D = \left[\partial\theta^i/\partial\alpha^k\right]$ is the Jacobian matrix of the $\alpha \mapsto \theta$ transformation, $R = \sigma_{\hat{\theta}\hat{\theta}}$ is the covariance matrix

$$R^{ik}(\alpha) = \int (\hat{\theta}^i(x) - \theta^i(\alpha))(\hat{\theta}^k(x) - \theta^k(\alpha))\mu(\alpha,\ dx) \tag{3.6}$$

of unbiased estimators $\hat{\theta}^i(x)$ with respect to $\mu(\alpha,\ dx) = \mathrm{Tr}\ S(\alpha)M(dx)$,

$$\int \hat{\theta}^i(x)\mu(\alpha,\ dx) = \theta^i(\alpha)J(\alpha),\ J(\alpha) = \mathrm{Tr}\,S(\alpha), \tag{3.7}$$

and $\sigma_{\hat{\gamma}\hat{\gamma}}$ is a similar covariance matrix for the logarithmic derivatives $\hat{\gamma}_k = \partial \ln \mu(\alpha)/\partial\alpha^k$, $k = 1, \ldots, n$. We will derive the inequality for the general noncommutative case.

3.1.2 Symmetric Nonclassical Bound

The lower bound of inequality (3.5) depends, naturally, on the choice of the representation determined by the method of measurement. by using the noncommutative analog of the logarithmic derivative introduced by Helstrom, we can obtain a more exact bound for the variances of the unbiased estimators that does not depend on the choice of representation.

If we assume that the family $\{S_\theta\}$ of the trace class density operators is strongly differentiable in θ in a certain region Θ, we can define a symmetric logarithmic derivative by the following equation:

$$\hat{g}_\theta S_\theta + S_\theta \hat{g}_\theta = 2S'_\theta. \tag{3.8}$$

It is easy to show (see [31]) that if $|\mathrm{Tr}\,(S'_\theta \hat{x})|^2 \leq c\mathrm{Tr}\,(S_\theta \hat{x}^2)$ for every Hermitian operator $\hat{x}$, the solution to this equation exists and is unique, with $\mathrm{Tr}\,\left(S_\theta \hat{g}^2_\theta\right) < \infty$.

Let us consider the operator

$$\hat{x} = \int \hat{\theta}(x) M(dx), \quad \mathrm{Tr}\,(S_\theta \hat{x}) = \theta J_\theta,$$

determined by the unbiased estimator $\hat{\theta}$ for a fixed measurement M. Since

$$\int (\hat{\theta}(x) - \theta)^2 \mu_\theta(dx) = \mathrm{Tr}\left(S_\theta \int (\hat{\theta}(x) - \theta)^2 M(dx)\right),$$

$$\mathrm{Tr}\left(S_\theta \int (\hat{\theta}(x) - \hat{x}) M(dx)(\hat{\theta}(x) - \hat{x}) + (\hat{x} - \theta)^2\right) \geq \mathrm{Tr}\,[S_\theta(\hat{x} - \theta)^2],$$

it is sufficient to find the lower bound of the variance $\sigma^2_{\hat{x}}$ of operator $\hat{x}$. By analogy with the commutative case we have

$$\begin{aligned} J_\theta &= \mathrm{Tr}\,(\hat{x}S'_\theta - \theta J'_\theta) = \mathrm{Tr}\,[(\hat{x} - \theta)S'_\theta] = \tfrac{1}{2}\mathrm{tr}\,[(\hat{x} - \theta)(\hat{g}_\theta S_\theta + S_\theta \hat{g}_\theta)] \\ &= \tfrac{1}{2}\mathrm{Tr}\,[S_\theta((\hat{x} - \theta)(\hat{g}_\theta - \gamma_\theta) + (\hat{g}_\theta - \gamma_\theta)(\hat{x} - \theta))] \\ &= \langle \hat{x} - \theta | \hat{g}_\theta - \gamma_\theta \rangle_+ . \end{aligned}$$

Thus, the total intensity J_θ is equal to the symmetrized scalar product with respect to S_θ of the operators $\hat{x} - \theta$ and $\hat{g}_\theta - \gamma_\theta$, where $\gamma_\theta = J'_\theta / J_\theta$ is the mean value of the operator $\hat{g}_\theta$ of the logarithmic derivative, and $J'_\theta = \mathrm{Tr}\,(S_\theta \hat{g}_\theta) = \mathrm{Tr}\,S'_\theta$. Applying the Schwarz inequality

$$\left|\langle \hat{x} - \theta \mid \hat{g}_\theta - \gamma_\theta \rangle_+\right|^2 \leq \langle \hat{x} - \theta \mid \hat{x} - \theta \rangle_+ \langle \hat{g}_\theta - \gamma_\theta \mid \hat{g}_\theta - \gamma_\theta \rangle_+ ,$$

we arrive at the sought inequality:

$$\sigma^2_\theta \geq \mathrm{Tr}\,[S_\theta(\hat{x} - \theta)^2]/J_\theta \equiv \sigma^2_x \geq J_\theta / \mathrm{Tr}\,[S_\theta(\hat{g}_\theta - \gamma_\theta)^2] \equiv 1/\sigma^2_{\hat{g}\theta}. \tag{3.9}$$

Thus, the variance of any unbiased estimator cannot be smaller than the inverse variance of the operator of the logarithmic derivative (3.8):

$$\sigma^2_{\hat{g}\theta} = \mathrm{Tr}\,[S_\theta(\hat{g}_\theta - \gamma_\theta)^2]/J_\theta. \tag{3.10}$$

A similar result can be obtained in the case where there are several parameters $\theta = [\theta^i]^m_{i=1}$ for the estimator $\hat{\theta}(x) = [\hat{\theta}^i(x)]^m_{i=1}$ satisfying the unbiasness conditions (3.1), which when met make matrix (3.5) the covariance matrix of estimators $\hat{\theta}^i$, and the mean square error at a fixed R_θ assumes the minimal value.

For the covariance matrix R_θ, Helstrom [35] has established the lower bound by assuming that the operator function $S_\theta = S(\theta)$ is differentiable and using the concept of the operators $\hat{g}_i$ of partial symmetrized logarithmic derivatives of the functions $S(\theta)$ in θ^i. He defined these operators by the following equations:

$$\hat{g}_i S_\theta + S_\theta \hat{g}_i = 2(\partial S_\theta / \partial \theta^i). \tag{3.11}$$

As in the classical case [49], this bound is defined by the matrix $G_\theta = [G_{ik}(\theta)]$ of the covariances of the solutions $\hat{g}_i = \hat{g}_i(\theta)$ of (3.1). This matrix for noncommutative $\hat{g}_i$ is taken in symmetrized form

$$G_{ik}(\theta) = \frac{1}{2} \mathrm{Tr} S_\theta (\hat{g}_i \hat{g}_k + \hat{g}_k \hat{g}_i) \tag{3.12}$$

(the mathematical expectations of Tr $(S_\theta \hat{g}_i(\theta))$ are equal to zero). The corresponding inequality has the form

$$R_\theta \geq G_\theta^{-1}, \quad \theta \in \Theta \tag{3.13}$$

and is understood as the nonnegative definiteness of matrix $[R^{ik}(\theta) - G^{ik}(\theta)]$, where $G^{ik}(\theta)$ are the elements of the inverse matrix $G_\theta^{-1} : G^{ij}(\theta)\, G_{jk}(\theta) = \delta^i_k$. Inequality (3.13) is the noncommutative analog of the Rao-Cramér inequality [49]. The matrix G_θ plays the role of a metric tensor locally defining the distance $\sigma(\theta, \theta + d\theta) = G_{ik}(\theta)\, d\theta^i\, d\theta^k$ in the parameters space Θ, similar to the Fisher information distance in classical statistics.

We now turn to a more general situation where the state parameters are not the measured parameters θ^i but other parameters $\alpha = \{\alpha^k, k = 1, \ldots, n\}$, $S = S(\alpha)$. The parameters θ^i are differentiable functions $\theta^i = \theta^i(\alpha)$ of the unknown parameters. The respective generalized Helstrom inequality (3.13) represents a bound for the covariance matrix $R = R(\alpha)$ of the estimators $\hat{\theta}^i$ in a form invariant with respect to the choice of the variables of states $S(\alpha)$,

$$R \geq D G^{-1} D^{\intercal}, \tag{3.14}$$

where $D = \left[\partial \theta^i / \partial \alpha^k\right]$, and $G = G(\alpha)$ is the covariance matrix (3.12) of the operators $\hat{g}_k = \hat{g}_k(\alpha)$ of symmetrized logarithmic derivatives of the operator function $S(\alpha)$ in α^k.

Inequality (3.14), which is equivalent to inequality (3.13) only if $m = n$ and matrix $D = D(\alpha)$ is nonsingular, can be verified by a line of reasoning similar to the one that will lead us to inequality (3.17) (see Section 3.2.4).

Inequality (3.22) can be reduced to the classical Rao-Cramér inequality only where the family $\{S(\alpha)\}$ is commutative. For noncommutative families other generalizations [50] of the Rao-Cramér inequality are possible, generalizations that are based on other definitions of logarithmic derivatives and that lead to other lower bounds for R differing from the invariant Helstrom bound $DG^{-1}D^{\intercal}$. For real-valued parameters α these generalizations may serve equally well as analogs of the Rao-Cramér inequality and coincide only if $\{S(\alpha)\}$ constitutes a commutative family, in which case they are reduced to the classical Rao-Cramér inequality. However, in the event of complex-valued parameters α a special invariant generalization of the Rao-Cramér inequality becomes especially important. It is based on the notions of right and left logarithmic derivatives and was suggested independently by Belavkin [18] and Yuen and Lax [52].

From now on we shall consider complex parameters $\alpha^k = \alpha_1^k + \mathrm{j}\alpha_2^k$, $\alpha = \{\alpha^k\} \in \mathbb{C}^n$. The estimated parameters $\theta^i = \theta^i(\alpha, \alpha)$, $i = 1, \ldots, m$, are assumed to be functions independently differentiable in α and $\bar{\alpha}$.[1] Let us define the non-Hermitian logarithmic derivatives of $S = S(\alpha, \alpha)$ thus:

$$S\hat{h}_k = \partial S/\partial\bar{\alpha}^k, \quad \hat{h}_k^* S = \partial S/\partial\alpha^k, \quad k = 1, \ldots, n. \tag{3.15}$$

The operators $\hat{h}_k = \hat{h}_k(\alpha, \bar{\alpha})$ are called the right derivatives with respect to $\bar{\alpha}^k$ (and the operators $\hat{h}_k^*$ the left derivatives with respect to α) and have zero mathematical expectations. The covariance matrix $H = [H_{ik}]$ of

$$H_{ik}(\alpha, \bar{\alpha}) = \mathrm{Tr}\,[S(\alpha, \bar{\alpha})\hat{h}_i\hat{h}_k^*] \tag{3.16}$$

is Hermitian and, assuming it is nonsingular, defines a positive definite metric $\mathrm{d}s^2 + H_{ik}\,\mathrm{d}\bar{\alpha}^i\mathrm{d}\alpha^k$ in a complex domain $\mathcal{O} \subset \mathbb{C}^n$ of the unknowns $\alpha \in \mathcal{O}$.

Suppose that a joint measurement of the parameters θ^i is described by a decomposition of unity that defines the estimator $\hat{\theta}$. This estimator is represented by a vector quantity that, in general, assumes complex values $x = \{x^i\} \in \mathbb{C}^m$, is represented by a conditional distribution $\mu(dx \mid \alpha, \alpha) = \mathrm{Tr}\,[M(dx)S(\alpha, \bar{\alpha})]$, and satisfies the unbiasness condition $\left\langle \hat{\theta}^i \right\rangle = \theta^i(\alpha, \bar{\alpha})$. Then the mean square error of measurement is determined by the matrix $R = R(\alpha, \bar{\alpha})$ of covariances $R^{ik} = \left\langle (\hat{\theta}^i - \theta^i)(\hat{\theta}^k - \theta^k)^* \right\rangle$, for which the following inequality holds true:

$$R \geq DH^{-1}D^{\dagger}, \tag{3.17}$$

where $D = D(\alpha, \bar{\alpha})$, as in (3.14), is the matrix of the derivatives $\partial\theta^i/\partial\alpha^k$, and $D^{\dagger}$ is the respective Hermitian conjugate matrix. Even in the real case, that is $\hat{\theta}^i = \bar{\theta}^i$, inequality (3.17) leads to a lower bound that differs from the Helstrom bound (3.14). We will call the lower bound in (3.17) the right bound. Other bounds can also be considered, say, the left bound, which is based on the left logarithmic derivatives with respect to $\bar{\alpha}$. The proof of all such inequalities is similar to that of inequality (3.17), which is given in Section 3.2.4. The right

[1] The derivatives $\partial/\partial\bar{\alpha}$ and $\partial/\partial\alpha$ are defined in terms of the partial derivatives $\partial/\partial\alpha_1$ and $\partial/\partial\alpha_2$ in the common manner:

$$\begin{aligned}
\partial/\partial\alpha &= \frac{1}{2}(\partial/\partial\alpha_1 + j\partial/\partial\alpha_2), \\
\partial/\partial\bar{\alpha} &= \frac{1}{2}(\partial/\partial\alpha_1 + j\partial/\partial\alpha_2).
\end{aligned}$$

bound in (3.17) is invariant under replacement of derivatives with respect to α^k, by derivatives with respect to new variables $\beta^k = \beta^k(\alpha)$ only if the functions $\beta^k(\alpha)$ are analytic, that is, $\partial\beta^k/\partial\bar{\alpha}^i = 0$, and the matrix of derivatives $\partial\beta^k/\partial\alpha^i$ is nonsingular. Hence, the use of inequality (3.17) in invariant form $R \geq H^{-1}$, where, as in (3.13), we employ derivatives with respect to the estimated parameters θ^i (but, in contrast to (3.13), right derivatives rather than symmetrized are used), is inexpedient since the condition for the equivalence of these inequalities includes not only the condition that matrix D be nonsingular but the analyticity condition $\partial\theta^i/\partial\bar{\alpha}^k = 0$ as well (that is, the independence of functions $\theta^i(\alpha, \bar{\alpha})$ on $\bar{\alpha}$), which is not out initial assumption. A similar situation for complex-valued parameters α exists in the classical case.

3.2 Uncertainty Relations and Efficient Measurements

In this section we will introduce the notion of canonical families of wave patterns for whose canonical parameters we will establish uncertainty relations that generalize the quantum mechanical uncertainty relations obtained in the one-dimensional case by Helstrom [31] and in the case of multidimensional Lie algebra by Belavkin [18]. We will then find the limit of accuracy in estimating the canonical Lie parameters of wave patterns and prove that such limits are exact only for canonical signals for which there are efficient measurement or quasimeasurement procedures. The discourse will follow the scheme suggested in [18]; for examples of uncertainty relations for quantum systems the readers is advised to turn to [36],[31].

3.2.1 Canonical Families and Uncertainty Relations

In classical mathematical statistics an important role is played by canonical, or exponential, families of probability distributions, for which a special selection of parameters θ and α makes the Rao-Cramér bound exact. In Section 3.2.3 we will prove that in the noncommutative case a similar role is played by density operators of the form

$$S(\beta, \bar{\beta}) = \chi^{-1} \mathrm{e}^{\beta^k \hat{x}_k^*} S_0 \mathrm{e}^{\bar{\beta}^k \hat{x}_k^*}, \tag{3.18}$$

where the $\hat{x}_k, k = 1, \ldots, n$, are linearly independent operators in $\mathcal{H}$, which may be non-Hermitian ($\hat{x}_k^* \neq \hat{x}_k$) and may not commute with the conjugate operators ($\hat{x}_i \hat{x}_k^* \neq \hat{x}_k^* \hat{x}_i$), and $\chi = \chi(\beta, \bar{\beta})$ is the generating function of the moments of these operators in state S_0:

$$\chi(\beta, \bar{\beta}) = \mathrm{Tr}\, S_0 \mathrm{e}^{\bar{\beta}^k \hat{x}_k} \mathrm{e}^{\beta^k \hat{x}_k^*}, \tag{3.19}$$

which is finite ($\chi < \infty$) in a neighborhood of zero $\beta = 0$ of the complex space $\mathbb{C}^n$. The family of density operators (3.18) will be said to be canonical and the parameters β^k, canonically conjugate to the $\hat{x}_k$. In contrast to the commutative case, even for Hermitian operators $\hat{x}_k$ it is meaningful to assume that the conjugate parameters β^k may have complex values.

Of special interest is the case, which has no classical analog, of canonical states (3.18) where the β^k are imaginary and the $\hat{x}_k$ are Hermitian. The parameters $\theta^k = \operatorname{Im}\beta^k/(2\pi)$ acquire a dimensionality and meaning of quantities that are dynamically conjugate to the $\hat{x}_k$; for instance, if $\hat{x}$ is frequency, θ is time, if x is momentum, θ is displacement, if $\hat{x}$ is angular momentum, θ is the angle of rotation. The canonical states (3.18) at $\beta^k = 2\pi \mathrm{j}\theta^k$ assume the form

$$S_\theta = \mathrm{e}^{2\pi \mathrm{j}\theta^k \hat{x}_k} S_0 \mathrm{e}^{-2\pi \mathrm{j}\theta^k \hat{x}_k} \tag{3.20}$$

and are unitary equivalent to state S_0, which corresponds to a zero value of θ. It has been established that if we put $\alpha = \beta$ and apply inequality (3.17) to the canonical family (3.20), we arrive at the exact formulation of the generalized uncertainty principle for any pair of dynamically conjugate quantities $\hat{\theta}^k$ and $\hat{x}^k$, where the first quantity in the pair may not correspond to the Hermitian operator that meaningfully describes in $\mathcal{H}$ the measurement of this quantity.[2]

Differentiation (3.18) with respect to $\bar{\beta}^k$ and comparing the result with (3.1), we get

$$\hat{h}_i = \mathrm{e}^{-\bar{\beta}^k \hat{x}_k} \chi \frac{\partial}{\partial \bar{\beta}^i} (\chi^{-1} \mathrm{e}^{\bar{\beta}^k \hat{x}_k}) = \hat{x}_i(\bar{\beta}) - \theta_i, \tag{3.21}$$

where $\hat{x}_i(\bar{\beta}) = \mathrm{e}^{-\beta^k \hat{x}_k} \frac{\partial}{\partial \bar{\beta}^i} \mathrm{e}^{\bar{\beta}^k \hat{x}_k}$, and $\theta_i = \frac{\partial}{\partial \bar{\beta}^i} \ln \chi = \operatorname{Tr}[S\hat{x}_i(\bar{\beta})]$. Matrix (3.16), therefore, is the covariance matrix.

$$H_{ik} = \operatorname{Tr}[S(\hat{x}_i(\bar{\beta}) - \theta_i)(\hat{x}_k(\bar{\beta}) - \theta_k)^*] = \frac{\partial^2 \ln \chi}{\partial \bar{\beta}^i \partial \beta^k} \tag{3.22}$$

of the operators $\hat{x}_i(\bar{\beta})$ analytic in $\bar{\beta}$ and coinciding with $\hat{x}_i$ at $\beta = 0$. The inequality (3.17) in the neighborhood of point $\beta = 0$, therefore, can be written in the form of the uncertainty relation

$$R \gtrsim DS^{-1}D^\dagger, \tag{3.23}$$

which establishes the inverse proportionality between he matrix $S = [S_{ik}]$ of the covariances

$$S_{ik} = \operatorname{Tr}[S(\beta, \bar{\beta})(\hat{x}_i - \mu_i)(\hat{x}_k - \mu_k)^*] \tag{3.24}$$

of the operators $\hat{x}_i$, $\operatorname{Tr}[S(\beta, \bar{\beta})\hat{x}_i] = \mu_i$, and the covariance matrix R of the estimators $\hat{\theta}^i$ of the functions $\theta^i(\beta, \bar{\beta})$ of the conjugate parameters β and $\bar{\beta}$.

[2] The Heisenberg uncertainty principle is usually proved only for such dynamically conjugate quantities described by noncommutative operations $\hat{p}$ and $\hat{q}$ that satisfy, say, the commutation relations $[\hat{p}, \hat{q}] = 1/2\pi j$. The proof employs the well-known scalar inequality $\left\langle (\hat{p} - \langle \hat{p} \rangle)^2 \right\rangle \left\langle (\hat{q} - \langle \hat{q} \rangle)^2 \right\rangle \geq |\langle [\hat{p}, \hat{q}] \rangle|^2 / 4$, which is valid for any pair of operators $\hat{p}$ and $\hat{q}$. Strenthening and matrix multidimensional generalization of this inequality in terms of the covariance estimations of an arbitrary family of noncommutative operators are suggested in [7].

At $\theta = \beta$, (3.23) transforms into strict inequality in the entire domain $\mathcal{O} \ni \beta$. Putting $\theta = \mathrm{Im}\,\beta/(2\pi)$ and allowing for the fact that $\partial\theta/\partial\beta = 1/4\pi\mathrm{j}$, we obtain $\hat{x}^k = \hat{x}$ the generalized uncertainty relation

$$(2\pi)^2 R_\theta \geq (1/4) S^{-1} \tag{3.25}$$

in terms of the variances $R_\theta = \left\langle (\hat{\theta} - \theta)^2 \right\rangle_\theta$, with $S = \mathrm{Tr}\ [S_0(\hat{x} - \mu)^2]$, valid for any pair of dynamically conjugate quantities $\hat{\theta}$ and $\hat{x}$ defining the canonical family (3.20). For the quantum case of pure states $S_0 = |\psi_0)(\psi_0|$, the scalar inequality (3.13) by Helstrom [30] by a complicated procedure for calculating the matrix elements of the operators of symmetrized logarithmic derivatives.

In the multidimensional case, when the operators $\hat{x}_k$ are pairwise commutative (but not necessarily with $\hat{x}_k^*$ and S_0), the situation is the same: $\hat{x}_k(\bar{\beta}) = \hat{x}_k$ for every $\beta \in \mathcal{O}$ and inequality (3.23) is strict. The averages μ_k and the covariances (3.24) at $\hat{x}_k = \hat{x}_k^*$ and $\beta^k = 2\pi\mathrm{j}\theta^k$ are independent of θ and, therefore, coincide with the respective values at $\theta = \theta$: $\mu_k = \mathrm{Tr}\ (S_0 \hat{x}_k)$ and

$$S_{ik} = \mathrm{Tr}\ [S_0(\hat{x}_i - \mu_i)(\hat{x}_k - \mu_k)]. \tag{3.26}$$

The uncertainty relation (3.25) in this case acquires a matrix meaning: R_θ is the covariance matrix of estimators $\hat{\theta}^i$ of the canonical parameters belonging to a translation group in state S_θ, and S is the covariance matrix (3.26) of the generators $\hat{x}_k$ of the group, defining the lower bound $S^{-1}/16\pi^2$ for R_θ uniformly in every $\theta \in \mathbb{R}^n$.

We now take up the case of noncommutative $\{\hat{x}_k\}$. Suppose that the operators $\hat{x}_k$ form a Lie algebra:

$$\hat{x}_i \hat{x}_k - \hat{x}_k \hat{x}_i = C_{ik}^j \hat{x}_j, \tag{3.27}$$

where C_{ik}^j are structure constants. Here the operators $\hat{x}_i(\bar{\beta})$ in (3.21) are linear combinations of the generators $\hat{x}_i$:

$$\hat{x}_i(\bar{\beta}) = L^{-1}(-\bar{\beta})_i^j \hat{x}_j, \tag{3.28}$$

with $L(\xi) = \xi^k C_k (I - \mathrm{e}^{-\xi C})^{-1}$ an n-by-n matrix that exists in a neighborhood $\mathcal{O} \in \mathbb{C}^n$ of zero $\xi = 0$, and the $C_k = \left[C_{ik}^j\right]$ are the generators of the adjoint representation of the commutation relations (3.27). Expressing the covariance matrix H of the operators (3.28) in terms of the covariances (3.24) of the generators $\hat{x}_i$, we get instead of (3.23) the inequality

$$R \geq D L^\dagger S^{-1} L D^\dagger, \tag{3.29}$$

where $L = L(-\beta)$. In the case of (3.20), the family of S_θ is unitary homogeneous with respect to the Lie group with Hermitian operators $\hat{x}_k$ and canonical parameters θ^k. Similarly to (3.25), we obtain a more general relationship

$$(2\pi)^2 R_\theta \geq \frac{1}{4} L_\theta^\intercal S^{-1} L_\theta, \tag{3.30}$$

where $L_\theta = \theta^i G_i (I - \mathrm{e}^{-\mathrm{j}\theta^k G_k})^{-1}$, $G_k = 2\pi \mathrm{j} C_k$. Inequality (3.30) determines in the domain $\mathcal{O} \subset \mathbb{R}^n$ of the convergence of the series expansion

$$(I - \mathrm{e}^{-\theta^k G_k})^{-1} = \sum_{m=0}^{\infty} \mathrm{e}^{-m\theta^k G_k}, \quad \theta = \{\theta^i\} \in \mathcal{O},$$

the lower bound of the mean square error in estimating the canonical parameters of the unitary representation $\mathrm{e}^{2\pi \mathrm{j}\theta^k \hat{x}_k}$ of a Lie group.

3.2.2 Efficient Measurements and Quasimeasurements

In classical statistics, estimations whose covariance matrix assumes the minimal value and thus transforms, locally or globally, the Rao-Cramér inequality into an equality are known as efficient (locally or globally, respectively). In the noncommutative case, the concept of efficiency introduced by analogy with the classical concept looses its universality because the generalization of the Rao-Cramér inequality is not unique and the definitions of locally efficient estimates [35], [50], [52] based on different variants of this generalization are not equivalent. For this reason we distinguish between the efficient measurements (or estimates) for which the invariant Helstrom bound (3.14) is attained and those for which the right bound (3.17) is attained, with the former called Helstrom efficient and the latter, right efficient. As we show below, the notion of right efficiency is more universal: measurements that are Helstrom efficient are right efficient, but not vice versa. Let us first prove that Helstrom efficient estimates exist globally for canonical families of density operators (3.18) if the operators $\hat{x}_k$ are Hermitian and pairwise commutative and if for the estimated parameters θ we take the derivatives $\theta_k = \partial \ln \chi / \partial \varkappa^k$ of the generating function $\chi(\varkappa) = \mathrm{Tr}(S_0 \mathrm{e}^{\varkappa^k \hat{x}_k})$, where $\varkappa = \beta + \bar{\beta}$. The parameters θ_k selected in this manner coincide with the averages defined by the canonical subfamilies of the density operators,

$$S(\varkappa) = \chi^{-1}(\varkappa) \mathrm{e}^{\varkappa^k \hat{x}_k/2} S_0 \mathrm{e}^{\varkappa^k \hat{x}_k/2}, \tag{3.31}$$

with $\mathrm{Im}\, \beta^k = 0$:

$$\theta_k(\varkappa) = \mathrm{Tr}\, [S(\varkappa)\hat{x}_k] = \partial \ln \chi / \partial \varkappa^k. \tag{3.32}$$

Taking for parameters α^k the canonical parameters $\varkappa^k$ and differentiating the operator functions (3.31), we obtain the symmetrized logarithmic derivatives in $\varkappa^k : \hat{g}_k = \hat{x}_k - \theta_k$. Thus, the covariances (3.12) coincide with the covariances of operators $\hat{x}_k$,

$$G_{ik} = \mathrm{Tr}\, [S(\varkappa)(\hat{x}_i - \theta_i)(\hat{x}_k - \theta_k)] = \frac{\partial^2 \ln \chi}{\partial \varkappa^i \partial \varkappa^k}, \tag{3.33}$$

which are equal to the derivatives $D_{ik} = \partial \theta_i / \partial \varkappa^k$ defining matrix D in (3.14) for $\alpha = \varkappa$. Therefore, inequality (3.14) assumes the form $R \geq G$, or $[R_{ik} - G_{ik}] \geq 0$, where $R_{ik} = \left\langle (\hat{\theta}_i - \theta_i)(\hat{\theta}_k - \theta_k) \right\rangle$ are the covariance of the unbiased estimators $\hat{\theta}_k : \left\langle \hat{\theta}_k \right\rangle = \theta_k$. If for these estimators we take the results x_k of measurements of the observables $\hat{x}_k$ (which are compatible), then matrix R assumes the minimal value $R = G$. Thus, for the canonical families (3.31) with commutative operators $\hat{x}_k$ there exists a Helstrom efficient measurement of the functions (3.32) of the canonical parameters $\varkappa_k$, which is the usual compatible measurement of the observables x_k. The domain of this efficiency, obviously, coincides with the domain $\mathcal{O} \subset \mathbb{R}^n$ for which $\chi(\varkappa) < \infty, \varkappa \in 0$. It has been established that the converse is true in the following sense.

Let the estimators $\hat{\theta}_k$ (i.e. the results of a measurement) have averages $\theta_k(\alpha)$ and covariances $R_{ik}(\alpha)$ that are differentiable in a certain domain, and let the matrices $R = [R_{ik}\,(\alpha)], D = [\partial\theta/\partial\alpha^k]$ satisfy the conditions

$$\partial(R^{-1}D)^j_k/\partial\alpha^i = \partial(R^{-1}D^j_i)/\partial\alpha^k \tag{3.34}$$

(the regularity conditions). We can then introduce the canonical parameters $\varkappa^k = \varkappa^k(\alpha)$ defined uniquely by the derivatives $\partial\varkappa^i/\partial\alpha^k = (R^{-1}D)^i_k$ if we put $\varkappa^k(\alpha_0) = 0$ for a fixed α_0. It can easily be verified that for a family of density operators $S(\alpha)$ of canonical form (3.31), with $\varkappa^k = \varkappa^k(\alpha)$ differentiable functions possessing a nonzero Jacobian, the regularity conditions are met in an efficient measurement at $\theta_k\,(\alpha) = \partial\ln(\chi(\varkappa(\alpha))/\partial\varkappa^k$ such that $R_{ik} = G_{ik}(\varkappa(\alpha))$ and $(R^{-1}G)^i_k = \partial\varkappa^i/\partial\alpha^k$. Proof of the converse assertion that under the regularity conditions the global Helstrom efficiency comes into play only for canonical families (3.31) is given in Section 3.2.4 for a more general situation involving complex variables.

Hence we have proved the following

Theorem 10 *Under appropriate regularity conditions, inequality* (3.14) is *transformed into an equality in a certain domain* $\mathcal{O} \subset \mathbb{R}^n$ *if and only if the family of density operators* $S(\alpha)$ *has the canonical form* (3.14)*, where* $\hat{x}_k$*,* $k = 1, \ldots, n$*, are commutative Hermitian operators in* $\mathcal{H}$*, and the canonical parameters* $\varkappa^k$*,* $k = 1, \ldots, n$*, are functions of parameters* α *defined by the equations*

$$\partial\ln\chi/\partial\varkappa^k = \theta_k(\alpha), \quad k = 1, \ldots, n.$$

3.2.3 The Theorem Regarding the Canonical Density Operators

Suppose that in a certain domain $\mathcal{O} \subset \mathbb{C}^n$ the unbiased estimators $\hat{\theta}_k$ possess averages $\theta_k(\alpha, \bar{\alpha})$ and covariances $R_{ik}(\alpha, \bar{\alpha})$ that satisfy the regularity conditions (3.34), to which we adjoin the analyticity condition

$$\tfrac{\partial}{\partial\bar{\alpha}^k} R^{-1}D = 0. \tag{3.35}$$

Here we can introduce, as we did in Section 3.2.1, canonically conjugate parameters $\beta^k = \beta^k(\alpha)$ via the equations $\partial\beta^i/\partial\alpha^k = (R^{-1}D)^i_k$ and conditions $\beta^k(\alpha_0) = 0$ for a fixed $\alpha_0 \in \mathcal{O}$ with the functions $\beta^k\,(\alpha)$ being analytic in view of conditions (3.35).

Theorem 11 *Under the formulated regularity conditions, inequality* (3.17) *is transformed into an equality in a certain domain* $\mathcal{O} \subset \mathbb{C}^n$ *if and only if the family* $\{S(\alpha, \alpha), \alpha \in \mathcal{O}\}$ *has the canonical form* (3.18)*, with* $S_0 = S(\alpha_0, \bar{\alpha}_0)$ *for an* $\alpha_0 \in \mathcal{O}$*, the operators* $\hat{x}_k,\ k = 1, \ldots, n,$ *simultaneously possess in* $\mathcal{H}$ *the property of the right proper decomposition of unity*

$$I = \int M(dx), \quad \hat{x}_k M(dx) = x_k M(dx), \quad x = \{x_k\} \in \mathbb{C}^n, \tag{3.36}$$

and the parameters $\beta^k,\ k = 1, \ldots, n,$ *are analytic functions* $\beta^k(\alpha)$ *determined by the equations*

$$\partial \ln \chi / \partial \bar{\beta}^k = \theta_k(\alpha, \bar{\alpha}), \quad \alpha \in \mathcal{O}.$$

Optimal estimation is then reduced to a quasimeasurement of the non-Hermitian operators $\hat{x}_k$ *described by the decomposition of unity* (3.36)*, while the minimal mean square error is determined by the covariance matrix*

$$R_{ik} = \mathrm{Tr}\,[S(\hat{x}_i - \theta_i)(\hat{x}_k - \theta_k)^*]. \tag{3.37}$$

Proof. Sufficiency is proved in the same way as in Section 3.2.1. Employing the fact of invariance oft he right bound (3.17) under the analytic transformations $\alpha \to \beta$, we select for the variables α^k determining this bound the parameters β^k of the family of density operators (3.18). The elements $\partial\theta_i/\partial\beta^k$ of matrix D then coincide with the elements (3.22) of matrix H if we allow for the fact that $\theta_i = \partial \ln \chi / \partial \bar{\beta}^i$. Since according to (3.36) the operators $\hat{x}_k$ are commutative, $\hat{x}_i \hat{x}_k = \int x_i x_k M(dx) = \hat{x}_k \hat{x}_i$, we have $\theta_k = \mu_k$ and $H_{ik} = S_{ik}$, where the μ_k are the averages of the $\hat{x}_k$, and the S_{ik} are the covariances (3.24) of these operators. Hence, inequality (3.17) assumes the form $R \geq S$. What remains to be proved is that the measurement described by the decomposition of unity (3.36) leads to an estimation for which $R = S$ even when the operators are not commutative with the respective conjugates: $\hat{x}_i \hat{x}_k^* \neq \hat{x}_k^* \hat{x}_i$ (which occurs when decomposition (3.36) is nonorthogonal). To do this, it is sufficient to allow for the representation

$$\hat{x}_i = \int x M(dx), \quad \hat{x}_i \hat{x}_k^* = \int x_i \bar{x}_k M(dx), \quad x \in \mathbb{C}^n, \tag{3.38}$$

which is obtained by integrating the equations in (3.36) and the adjoint equation $M(dx)\hat{x}_k^* = \bar{x}_k M(dx)$. Thanks to (3.38) the covariances

$$R_{ik} = \int (x_i - \theta_i)(\bar{x}_k - \theta_k) \mathrm{Tr}\,[S M(dx)] \tag{3.39}$$

of the estimators $\hat{\theta}_k$ obtained as a result of a quasimeasurement of operators $\hat{x}_k$ coincide with the covariances S_{ik} of these operators, which proves the efficiency of this quasimeasurement for the density operators (3.18). The proof of the converse of Theorem 11 follows from the derivation of inequality (3.17) and will be discussed in Section 3.2.4. ■

3.2.4 Discussion and an Example

Thus, the condition of (right) efficiency requires the existence of commutative operators possessing a joint right spectral decomposition and playing the role of sufficient statistics, which it is natural to call right efficient. Here it is sufficient to restrict the discussion to the operators in the minimal subspace generated by the regions $S(\beta,\bar{\beta})\mathcal{H}$ with the density operators $S(\beta,\bar{\beta})$ for all $\beta(\alpha)\in\mathbb{C}^n$ for which $\alpha\in\mathcal{O}$. Even if we consider only the real values of parameters $\theta(\alpha,\bar{\alpha})$, optimal estimation can be described by non-Hermitian and noncommutative (with the conjugate) operators of the right-efficient statistics and, therefore, may not be Helstrom efficient. However, estimates that are Helstrom efficient correspond, according to Theorem 10, to the particular case of right efficiency where the $\hat{x}_k$ are Hermitian. If the operators $\hat{x}_k$ in (3.18) are non-Hermitian but commutative with the conjugate operators, the right-efficient estimates also coincide with complexified estimates, which are Helstrom efficient. However, commutativity $\hat{x}_k\hat{x}_i^* = \hat{x}_i^*\hat{x}_k$, may not occur either.

Example 12 *Let $\hat{x}_k = \varphi_k(\hat{a})$, where the φ_k are entire functions $\mathbb{C}^r \to \mathbb{C}$, and let $\hat{a} = \{\hat{a}_i,\ i = 1,\dots,r\}$ be the annihilation operators satisfying the commutation relations*

$$\hat{a}_i\hat{a}_k - \hat{a}_k\hat{a}_i = \hat{0}, \quad \hat{a}_i\hat{a}_k^* - \hat{a}_k^*\hat{a}_i = \delta_i^k\hat{1}.$$

Show the right efficient measurement for the parameters $\theta_k = \partial\ln\chi/\partial\bar{\beta}^k$ of the density operators (3.18).

It is well-known that the operators $\hat{a}$ have right eigenvectors $|\alpha)\in\mathcal{H},\ \alpha\in\mathbb{C}^r$, that define the nonorthogonal decomposition of unity

$$I = \int |\alpha)(\alpha| \prod_{i=1}^{r} \pi^{-1}\mathrm{d}\,\mathrm{Re}\,\alpha_i \mathrm{d}\,\mathrm{Im}\,\alpha_i, \quad \hat{a}_i|\alpha) = \alpha_i|\alpha).$$

It is obvious then that the operators $\hat{x} = \varphi(\hat{a})$ have a right proper decomposition of unity (3.36), where

$$M(dx) = \mathrm{d}x \int \delta(x-\varphi(\alpha))\ |\alpha)(\alpha| \prod_{i=1}^{r} \pi^{-1}\mathrm{d}\,\mathrm{Re}\,\alpha_i \mathrm{d}\,\mathrm{Im}\,\alpha_i$$

($\mathrm{d}x$ is the Lebesgue measure on $\mathbb{C}^n$, and $\delta(x-\varphi)$ is the Dirac delta function). Hence, optimal estimation of the parameters $\theta_k = \partial\ln\chi/\partial\bar{\beta}^k$ of the density operators (3.18) at $\hat{x} = \varphi(\hat{a})$ is right efficient and can be reduced to a coherent measurement and estimation of $\theta = \varphi(\alpha)$ by the result α. In the particular case where $\varphi(\alpha)$ is a linear function and S_0 is a Gaussian state this fact has been established in [12].

Note that along with right and left lower bounds one can consider other combined bounds via the factorization $\theta = \theta_+ + \theta_-$ by appropriately defining the right derivatives with respect to θ_+ and the left derivatives with respect to θ_-. An interesting question arising in this connection is whether the class of efficient estimations is exhausted by the estimations for which at least one such bound is attained.

Let us now consider the (right) efficiency of estimating the parameters β^k of the canonical families (3.18). The inequality (3.17) corresponding to this case with $\theta^k = \beta^k$ has the form $R \geq H^{-1}$, where H is the matrix of derivatives (3.22). Without loss of generality, we can assume that $\mathrm{Tr}\,(\hat{x}_k S_0) = 0$.

Theorem 13 *The inequality $R \geq H^{-1}$ transforms into an equality if and only if the operators $\hat{x}_k$ in* (3.18) *possess a right joint decomposition of unity* (3.36)*, the generating function of the moments* (3.19) *of these operators in state S_0 is Gaussian, $\chi(\beta, \bar{\beta}) = \exp\{\bar{\beta}^i H_{ik} \beta^k\}$, with H_{ik} independent of β and $\bar{\beta}$ and linear functions $y^k = (H^{-1})^{ki} x_i$ of the results x_k of joint quasimeasurement of observables $\hat{x}_k$ are selected for the estimators $\hat{\beta}^k$.*

Proof. Sufficiency of the above-formulated conditions for the existence of right-efficient estimation is obvious: the fact that matrix H coincides with the covariance matrix S of operators $\hat{x}_k$ implies that the covariance matrix $R = H^{-1} S H^{-1}$ is equal to H^{-1}. Necessity follows from the necessary conditions of right efficiency in Theorem 11, according to which the family $S(\beta, \bar{\beta})$ must have the form

$$S(\beta, \bar{\beta}) = \psi^{-1} \mathrm{e}^{\theta_k \hat{y}^{k*}} S_0 \mathrm{e}^{\bar{\theta}_k \hat{y}^k}, \tag{3.40}$$

where $\psi = \mathrm{Tr}\,[S_0 \mathrm{e}^{\bar{\theta}_k \hat{y}^k} \mathrm{e}^{\theta_k \hat{y}^{k*}}]$, $\beta^k = \partial\left(\theta_k \ln \psi\right) / \partial \theta_k$, and the operators $\hat{y}^k$ possess the joint right unity decomposition in the sense:

$$I = \int M(dx), \quad \hat{y}^k M(dy) = y^k M(dy), \quad y = \left(y^k\right) \in \mathbb{C}^n.$$

Comparing (3.31) with (3.18), we conclude that $\theta_k \hat{y}^k = \bar{\beta}^k \hat{x}_k$, whence

$$\theta_k = H_{ki} \bar{\beta}^i, \quad \psi(\theta, \bar{\theta}) = \chi(\beta, \bar{\beta}) = \bar{\beta}^i H_{ik} \beta^k, \quad \hat{y}^k = (H^{-1})^{ki} \hat{x}_i.$$

The proof of Theorem 13 is complete. ■

Proof of Inequality (3.17)

1. We start with the one-dimensional case. Let $\hat{x}$ be a non-Hermitian operator in $\mathcal{H}$ for which

$$\mathrm{Tr}\,[\hat{x} S(\alpha, \bar{\alpha})] = \theta(\alpha, \bar{\alpha}). \tag{3.41}$$

Differentiating (3.41) with respect to α and employing definition (3.15) and the normalization condition $\mathrm{Tr}\, S(\alpha, \bar{\alpha}) = 1$, according to which $\mathrm{Tr}\,(S \hat{h}^*) = 0$, we obtain

$$\partial \theta / \partial \alpha = \mathrm{Tr}\,[S(\hat{x} - \theta) \hat{h}^*].$$

Since the covariance $\mathrm{Tr}\,[S(\hat{x} - \theta)\hat{h}^*]$ obeys the Schwarz inequality

$$|\mathrm{Tr}\,[S(\hat{x} - \theta)\hat{h}^*]|^2 \leq \mathrm{Tr}\,[S(\hat{x} - \theta)(\hat{x} - \theta)^*] \mathrm{Tr}\,(S \hat{h} \hat{h}^*), \tag{3.42}$$

which reflects the fact that the determinant of the 2-by-2 covariance matrix Tr $(S\hat{h}_i\hat{h}_k^*)$, $i = 0, 1$, with $\hat{h}_0 = (\hat{x} - \theta)$ and $\hat{h}_1 = \hat{h}$, is nonnegative, we can write

$$\text{Tr }[S(\hat{x} - \theta)(\hat{x}^* - \bar{\theta}) \geq |\partial\theta/\partial\alpha|^2/\text{Tr }(S\hat{h}\hat{h}^*). \tag{3.43}$$

This inequality, obviously, specifies the lower bound on the variance of the estimation of parameter $\theta = \theta\,(\alpha, \bar{\alpha})$ in the class of ordinary measurements described by normal operators $\hat{x}$. But since the normality condition, $\hat{x}\hat{x}^* = \hat{x}^*\hat{x}$, was not used in deriving (3.43), this bound is the lower one for the variance of any estimators $\hat{\theta}$ obtained as a result of arbitrary generalized measurements described in $\mathcal{H}$ by decompositions of unity $I = \int M(dx)$, $x \in \mathbb{C}$ that may be nonorthogonal. Indeed, the nonnegative definiteness

$$(\hat{x} - x)M(dx)(\hat{x} - x)^* \geq 0 \quad (M \geq 0) \tag{3.44}$$

implies

$$\int |x - \theta|^2 M(dx) \geq (\hat{x} - \theta)(\hat{x} - \theta)^*, \tag{3.45}$$

where $\hat{x} = \int xM(dx)$, and $\theta = \text{Tr }(S\hat{x})$. Taking the mathematical expectations of both sides of (3.44), allowing for the fact that the variance R of the estimator $\hat{\theta} = x$ is equal to Tr $S \times \int |x - \theta|^2 M(dx)$, and combining the result with (3.43), we find that

$$R \geq \text{Tr }[S(\hat{x} - \theta)(\hat{x} = \theta)^*] \geq |D|^2/H, \tag{3.46}$$

where $D = \partial\theta/\partial\alpha$, and $H = \text{Tr }(S\hat{h}\hat{h}^*)$. This proves inequality (3.17) for the one-dimensional case.

2. The equality in (3.46) occurs if, first, the averages of both sides of (3.45) coincide and if, second, the Schwarz inequality transforms into an equality. Actually, the first requirement establishes an equality in (3.44). Specifically, we have the following.

Lemma 14 *Suppose that the ranges $S(\alpha, \bar{\alpha})\mathcal{H}$ of the density operators from a family $\{S(\alpha, \alpha),\ \alpha \in \mathcal{O}\}$ generate the entire space $\mathcal{H}$. Then the fact that* Tr $(SA) = 0$ *for every nonnegative definite operator A in $\mathcal{H}$ and all $\alpha \in \mathcal{O}$ implies that $A = 0$.*

Proof. It is sufficient to prove that in $\mathcal{H}$ there is no vector χ of the form $\chi = S^{1/2}\psi$ for which $(\chi|A|\chi) \neq 0$. But this follows from the well-known inequality

$$\text{Tr }(S^{1/2}AS^{1/2}) \geq (\psi|S^{1/2}AS^{1/2}|\psi),$$

which is true for every nonnegative A at $(\psi \mid \psi) = 1$.

Applying this result to the operator A that is equal to the difference between the right- and left-hand sides of (3.45), we find that under the lemma's hypothesis the equality in (3.45) occurs only if

$$(\hat{x} - x)M(dx)(\hat{x} - x)^* = 0, \quad \text{i.e.} \quad \hat{x}M(dx) = xM(dx).$$

This proves that right-efficient estimation in a certain region $\mathcal{O} \ni \alpha$ exists if there is an operator of minimal sufficient statistics, $\hat{x}$, possessing a right proper decomposition of unity in the subspace generated by the subspaces $S(\alpha, \bar{\alpha})\mathcal{H}$. In the real case, $x \in \mathbb{R}$, such an operator $\hat{x}$ is obviously Hermitian.

The second requirement for equality to occurs in (3.46) is equivalent to the condition of linear dependence, $s\hat{h} = \bar{\lambda} S(x - \theta)$, where $\lambda = D/R$ if the first condition for equality in (3.45) is met. Extending this condition over the entire region $\mathcal{O} \ni \alpha$ in which the analyticity condition (3.25), $\partial\lambda/\partial\bar{\alpha} = 0$, is assumed to hold true, we arrive at the equation

$$\partial S/\partial\bar{\alpha} = \bar{\lambda} S(\hat{x} - \theta), \; \partial S/\partial\alpha = (\hat{x} - \theta)^* S\lambda$$

in $S = S(\alpha, \bar{\alpha})$. Its solution combined with the boundary condition $S(\alpha_0, \bar{\alpha}_0) = S_0$ has the canonical form (3.18), where $\beta(\alpha) = \int_{\alpha_0}^{\alpha} \lambda(\alpha)$ is an analytic function, and $\hat{x}$ is the operator of right-efficient statistics. This proves that in the one-dimensional case the existence of right-efficient estimation requires that the density operators $S(\alpha, \bar{\alpha})$ be canonical. This condition is formulated in Theorem 11. For the real case, $\hat{x}^* = \hat{x}$, this fact also proves the necessity in Theorem 10.

3. The multidimensional generalization can be carried out if for $\hat{x} - \theta$ and $\hat{h}$ we take the sums $(\hat{x}^i - \theta^i)\bar{\eta}_i$ and $\hat{h}_k \bar{\xi}^k$, where η_i, $i = 1, \ldots, m$ and ξ^k, $k = 1, \ldots, n$, are complex numbers. If we allow for the fact that here

$$\text{Tr}\,[S(\hat{x} - \theta)\hat{h}^*] = \bar{\eta}_i(\partial\theta^i/\partial\alpha^k)\xi^k,$$

then from (3.42) at $\xi^k = (H^{-1}D^\dagger)^{ki}\eta_i$ we arrive at the inequality

$$R^{ik}\bar{\eta}_i\eta_k \geq \text{Tr}\,[S(\hat{x}^i - \theta^i)(\hat{x}^k - \theta^k)^*\eta_i\eta_k] \geq (DH^{-1}D^\dagger)^{ik}\bar{\eta}_i\eta_k$$

valid for an arbitrary $\hat{x}^i$ for which Tr $(S\hat{x}^i) = \theta^i$. Putting $\hat{x}^i = \int x^i M(dx)$, where $\int M(dx) = I$, $\hat{x} \in \mathbb{C}^m$, is the decomposition of unity describing the estimator $\hat{\theta}^i = x^i$, and applying inequality (3.45) with $\hat{x} = \hat{x}^i\bar{\eta}_i$ and $\theta = \theta^i\bar{\eta}_i$, we obtain for the matrix R of covariances of $\hat{\theta}^i$ the first inequality in (3.46), which in view of the arbitrariness of η_i yields (3.17).

Inequality (3.46) transforms into an equality at $\alpha \in \mathcal{O}$ only when $\hat{x}^i M(dx) = x^i M(dx)$ and $\partial S/\partial\bar{\alpha}^k = \bar{\lambda}_{ki} S(x^i - \theta^i)$, where $\lambda_{ik} = (R^{-1}D)_{ik}$, whence, if we allow for the regularity conditions λ_{ik}, we arrive at (3.18). ■

3.3 Optimal and Covariant Estimation of Wave Patterns

In this section we will consider the necessary and sufficient conditions for the optimality of measuring sound and visual patterns by the criterion of mean square error in parameter estimation and by the maximal intensity criterion. To avoid substantiation of the operator integrals involved in the discussion this is done in [37], we interpret them as operator-valued Radon measures. The solution to the optimal measurement problem will be found for homogeneous families of wave patterns for which it coincides with optimal covariant measurements of the corresponding parameters of quantized fields, with the latter measurements introduced in [13].

3.3.1 Optimal Measurements

The problems of optimal estimation of continuous wave parameters constitute essentially multialternative problems with an infinite-dimensional solution space (or manifold) X. without loss of generality, we can assume that the information parameter space Θ coincides with X equipped with measure $\mathrm{d}\lambda$. Let us assume that a wave signal, which in general is described by a density operator S, depends in a continuous manner on real- or complex-valued random parameters $\theta = (\theta_1, \ldots, \theta_n),\ S = S_\theta$, having a given *a priori* distribution $P(d\theta)$. The deviation of the estimate $x \in X$ from θ is penalized by an integrable cost function $c_x(\theta)$ of the form, say, $(x-\theta)^2$. On X we must find an optimal quasimeasurement that (a) is described by an operator-valued measure $M(dx)$, (b) determines the decomposition of unity in the Hilbert space $\mathcal{H}$, and (c) minimizes the mean estimation cost

$$\langle c \rangle = \iint \mu_\theta(dx) c_x(\theta) P(d\theta) = \int \mathrm{Tr}\ R_x M(dx),$$

where $\mu_\theta(dx) = \mathrm{Tr}\ M(dx) S_\theta$ is the observed intensity distribution on X for a given θ, and $R_x = \int c_x(0)\ S_\theta P(d\theta)$ is the operator of the mean cost $x \in X$. We will now formulate the necessary and sufficient conditions for the optimality of solution M° to this extremal problem, which in [37] were introduced to estimate the parameters of quantum states. This will be done in a manner similar to that of Theorem 5:

Theorem 15 *The lower bound*

$$\inf_{M \geq 0} \left\{ \int \langle R_x, M(dx) \rangle \,\Big|\, \int M(dx) = I \right\} \tag{3.47}$$

is attained on measure M° if and only if for almost all $x \in X$ there exists a minorant operator $\Lambda^\circ \leq R_x$ such that

$$(R_x - \Lambda^\circ) M^\circ(dx) = 0 \quad \forall x \in X. \tag{3.48}$$

The operator Λ° is a trace class operator, or $\mathrm{Tr}\ \Lambda^\circ = \langle \Lambda^\circ, I \rangle < \infty$, *that determines the solution to the duality problem*

$$\sup_{\Lambda} \{ \langle \Lambda, I \rangle \,|\Lambda \leq R_x,\ x \in X \} \tag{3.49}$$

for which conditions (3.48) *are also necessary and sufficient* (*if we allow for the fact that $M^\circ \geq 0$ and $\int M^\circ(dx) = E$*).

Proof. For a proof of this theorem as well as for the existence conditions for a solution see [37].

Allowing for the fact that the operators $M(dx)$ can be decomposed into operators of the form $|\chi_x)(\chi_x|\, \mathrm{d}\lambda(x)$, where the χ_x are the generalized elements of space $\mathcal{H}$, we find that the problem of optimal estimation of wave parameters will be solved if and only if we can find a family of reference waves, $\{\chi_x\}$ satisfying the completeness condition

$$\int |\chi_x)(\chi_x|\, \mathrm{d}\lambda(x) = I \tag{3.50}$$

and a Hermitian operator Λ for which

$$R_x = \Lambda \geq 0, \quad (R_x - \Lambda)\chi_x = 0, \quad x \in X. \tag{3.51}$$

Note that, in contrast to problems of signal discrimination, in problems of parameter estimation the commutative case $R_x R_{x'} = R_{x'} R_x$, which can be reduced to the classical case, is of no practical interest and will not be discussed here.

The solution of problem (3.51) poses no fundamental difficulties in the case of a single unknown real-valued parameter $\theta(\chi = \mathbb{R}^1)$ and a quadratic penalty function

$$C_x(\theta) = (x - 0)^2.$$

The mean estimation cost operator

$$R_x = \int (x - \theta)^2 S_\theta P(d\theta)$$

in the case of (3.51) can be represented, via three Hermitian operators

$$R^{(k)} = \int \theta^k S_\theta P(d\theta), \quad k = 0, 1, 2, \tag{3.52}$$

in the form

$$\begin{aligned} R_x &= x^2 R^{(0)} - 2x R^{(1)} + R^{(2)} \\ &= (\hat{x} - x) R^{(0)} (\hat{x} - x) + R^{(2)} - \hat{x} R^{(0)} x, \end{aligned}$$

where $\hat{x}$ is an operator satisfying the equation

$$\hat{x} R^{(0)} + R^{(0)} \hat{x} = 2R^{(1)}. \tag{3.53}$$

and for χ_x take the complete orthogonal system of generalized eigenvectors determining the spectral decomposition of the Hermitian operator $\hat{x}$,

$$(\hat{x} - x)\chi_x = 0, \hat{x} = \int x|\chi_x)(\chi_x|\, \mathrm{d}x,$$

the conditions (3.51) are satisfied in an obvious manner:

$$(\hat{x} - x) R^{(0)} (\hat{x} - x) \geq 0, \quad (\hat{x} - x) R^{(0)} (\hat{x} - x)\chi_x = 0.$$

Thus, the solution of the parameter estimation problem by criterion (3.51) is reduced to measuring operator $\hat{x}$ satisfying (3.53). The result of such a measurement, x, leads to the minimal error $\langle c \rangle = \mathrm{Tr}\ \Lambda^{o}$ equal to the a posteriori variance

$$\sigma^2 = \mathrm{Tr}\ (R^{(2)} - \hat{x} R^{(0)} x).$$

■

As an example, let us consider the estimation of the amplitude of a coherent signal of known shape received against a background of Gaussian noise. The density operator of the corresponding mode has the Gaussian form

$$S(\theta) = \int |\alpha)(\alpha| \bar{n}^{-1} \exp\left\{ -\frac{|\alpha - \theta|^2}{\bar{n}} \right\} \pi^{-1} \mathrm{d}\,\mathrm{Re}\,\alpha \mathrm{d}\,\mathrm{Im}\,\alpha, \qquad (3.54)$$

where θ is the amplitude, which assumes real values, $\theta \in \mathbb{R}^1$. We assume that amplitude θ has a Gaussian prior density

$$p(\theta) = (2\pi \bar{s})^{-1/2} \exp\{-\theta^2 / 2\bar{s}\},$$

where $\bar{s}$ is the prior variance, $\langle \theta^2 \rangle = \bar{s}$. It is then easy to find, via the formulas of Gaussian integration, the operators (3.52), which define the mean decision cost operator

$$R_x = \left(x - \frac{2\bar{s}}{2\bar{s} + \bar{n} + 1/2} Q \right) S \left(x - \frac{2\bar{s}}{2\bar{s} + \bar{n} + 1/2} Q \right) + \frac{2\bar{s}(\bar{n} + 1/2)}{2\bar{s} + \bar{n} + 1/2} S.$$

Here $Q = \int \mathrm{Re}\,\alpha |\alpha)(\alpha| \pi^{-1}\ \mathrm{d}\,\mathrm{Re}\,\alpha\ \mathrm{d}\,\mathrm{Im}\,\alpha$ is the operator of the "coordinate" of the harmonic oscillator representing this mode, and

$$S = \int |\alpha)(\alpha| \left((2\bar{s} + \bar{n})\bar{n} \right)^{-1/2} \exp\left\{ -\frac{(\mathrm{Re}\,\alpha)^2}{2\bar{s} + \bar{n}} - \frac{(\mathrm{Im}\,\alpha)^2}{\bar{n}} \right\} \pi^{-1}\ \mathrm{d}\,\mathrm{Re}\,\alpha\, \mathrm{d}\,\mathrm{Im}\,\alpha$$

is the density operator, with $\int S_\theta p(\theta)\ \mathrm{d}\theta = R^{(0)}$. Hence, optimal estimation of the amplitude of a Gaussian signal is reduced to measuring the coordinate operator Q, whose result q determines the optimal estimate

$$x = 2\bar{s} q / (2\bar{s} + \bar{n} + 1/2)$$

with a minimal mean square error

$$\sigma^2 = 2\bar{s}(\bar{n} + 1/2)(2\bar{s} + \bar{n} + 1/2).$$

3.3.2 Miltidimensional Optimal Estimation Problem

In the case $(n > 1)$ even for the quadratic quality criterion

$$C_x(\theta) = \sum_{j=1}^{n} (x_j - \theta_j)^2$$

the general solution to problem (3.51) is unknown. Only in the particular case where the operators $\hat{x} = \{\hat{x}_j\}$ obeying (3.53), with

$$R^k = R_j^k \equiv \int \theta_j^k S_\theta p(\theta)\, \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_n, \quad k = 0, 1, 2,$$

commute with each other ($x_j x_i = x_i x_j$), optimal estimation is reduced, obviously, to joint measurement of these operators.

In general, a good estimate of parameters θ_j can be obtained by an indirect measurement of noncommutative operators $\{\hat{x}_i\}$ (see Section 3.2.2). However, this estimate is not necessarily optimal, even if the indirect measurement is ideal.

For an example let us take the complex-valued one-dimensional case ($X = \mathbb{C}^1$), which can also be interpreted as the real-valued two-dimensional:

$$C_x(\theta) = |x - \theta|^2 = (\mathrm{Re}(x - \theta))^2 + (\mathrm{Im}(x - \theta))^2.$$

An exact solution to the problem of optimal estimation of a single parameter θ has been obtained in [30] for this case of a quadratic penalty function, a Gaussian state S_θ; of the form (3.54), with $\theta \in \mathbb{C}^1$, and a Gaussian prior probability density

$$p(\theta) = \bar{s}^{-1} \exp\{-|\theta|^2/\bar{s}\}, \quad \bar{s} = \left\langle |\theta|^2 \right\rangle. \tag{3.55}$$

The density $p(\theta)$ is normalized with respect to

$$\mathrm{d}\lambda(\theta) = \pi^{-1} \mathrm{d}\,\mathrm{Re}\,\theta \mathrm{d}\,\mathrm{Im}\,\theta.$$

In this case, by the standard formulas of Gaussian integration, we can easily find the mean decision cost operator

$$R_x = \left(x^* - \frac{\bar{s}}{\bar{n} + \bar{s} + 1} A^*\right) S \left(x - \frac{|\alpha|^2}{\bar{s} + \bar{n}}\right) \mathrm{d}\lambda(\alpha)$$

is the density operator $S = \int S_\theta p(\theta)\, \mathrm{d}\lambda(\theta)$. Assuming that

$$\Lambda = \frac{\bar{s}(\bar{n} + 1)}{\bar{s} + \bar{n} + 1} S, \quad \chi_x = c^{-1} \Big| \left(1 + \frac{\bar{n} + 1}{\bar{s}}\right) x \Big),$$

where $|\alpha)$, $\alpha = (1 + (\bar{n} + 1)/\bar{s})x$, are coherent vectors, and $c = \bar{s}/(\bar{s} + \bar{n} + 1)$ is a coefficient that can be found from condition (3.50) if we allow for the completeness of coherent states

$$\int |\alpha)(\alpha|\, \mathrm{d}\lambda(\alpha) = I,$$

and allowing for the equation $A|\alpha) = \alpha|\alpha)$, we find that conditions (3.51) are met:

$$(x^* - cA^*) S (x - cA) \geq 0, \quad (x^* - cA^*) S (x - cA) |c^{-1} x) = 0. \tag{3.56}$$

Thus, optimal estimation in the one-dimensional complex-valued quadratic-Gaussian case is reduced to a coherent measurement describing an ideal proper indirect measurement of the annihilation operator A whose result α determines the estimate

$$x = \frac{\bar{s}}{\bar{s} + \bar{n} + 1}\alpha$$

with a minimal error

$$\sigma^2 = \frac{\bar{s}(\bar{n} + 1)}{\bar{s} + \bar{n} + 1}.$$

This error is equal to the error of the appropriate classical problem of estimation in a Gaussian linear channel with a noise intensity of $\bar{n} + 1$. The quantity $\bar{n}$ (the mean number of the noise quanta) is determined by the noise proper in the wave channel, while the unity corresponds to the “effective noise” thanks to the inaccuracy in the ideal indirect measurement. The measurement noise of unit intensity can be interpreted as the noise produced by an ideal wave amplifier or as the noise produced by an ideal optical heterodyne.

3.3.3 Optimal Measurement of Wave States

Let X be a set of hypotheses concerning the states of a wave field, and $\{\hat{R}_x,\ x \in X\}$ the respective decomposable family of density operators $R_x = \bigoplus_n R_x^{(n)}$ in the Hilbert space $\mathcal{H} = \bigoplus_n \mathcal{H}^{(n)}$. Then the set of measurements described by operator-valued measures M on $X(M(\cdot) \geq 0,\ \int M(\mathrm{d}x) = I)$ of the decomposable form $M(dx) = \bigoplus_n M^{(n)}\ \mathrm{d}x$ is sufficient. The optimal strategy is described by the family of operator-valued measures $M^{(n)}$ on X, i.e. $M^{(n)}(\cdot) \geq 0,\ \int M^{(n)}(dx) = I^{(n)}$, defined independently in $\mathcal{H}^{(n)}$ for every n by the conditions

$$(R_x^{(n)} = \Lambda^{(n)})M^{(n)}(dx) = 0, \quad R_x^{(n)} \leq \Lambda^{(n)} \quad \forall z \in X \tag{3.57}$$

(the maximum intensity criterion). Here $\Lambda^{(n)}$ are Hermitian trace class operators in $\mathcal{H}^{(n)}$ that are nonnegative (for $R_x^{(n)} \geq 0$) and can be represented in the form

$$\Lambda^2 = \int R_x^{(n)} M^{(n)}(dx) R_x^{(n)}.$$

Problem (3.57) is incomparably simpler than the general problem of optimal discrimination of a family $\{R_x\}$ and for every n has a finite-dimension of space $\mathcal{H}^{(n)}$ if the signal space $\mathcal{L}$ is finite-dimensional. In what follows, the index n will be dropped.

Let $\mathcal{U}_x = R_x\mathcal{H}$ be the range of values of operators R_x in $\mathcal{H}$, let $\mathcal{U}(dx)$ be their algebraic sum for all $x \in dx$, and let $\mathcal{U} = \int \mathcal{U}(dx)$ be the sum of all the subspace $\mathcal{U}_x \subset \mathcal{H}$. Each nonnegative operator R_x can be represented in the form $R_x = \psi_x\psi_x^*$, where ψ_x is the operator from $\mathcal{U}_x$ into $\mathcal{U}$. The following conjectures are multidimensional generalizations of the appropriate assertions of Theorem 3 (for a discrete set X).

Theorem 16 (1) *Subspace $\mathcal{U}$ is sufficient for solving problem* (3.57)*. Every operator Λ satisfying conditions $\Lambda \geq R_x$ $\forall x \in X$ for $R_x \geq 0$ has an inverse Λ^{-1} in $\mathcal{U}$.*
(2) *The solution to problem* (3.57) *in the sufficient space $\mathcal{U}$ has the form*

$$M(dx) = \Lambda^{-1}\psi_x\hat{\mu}(dx)\psi_x^*\Lambda^{-1}, \text{ where } \Lambda = \int\Big(\int \psi_x\hat{\mu}(dx)\psi_x^*\Big)^{1/2} \tag{3.58}$$

and $\hat{\mu}$ is a measure on X whose values $\hat{\mu}$ (dx) are nonnegative operators in $\mathcal{U}_x$ defined by the conditions

$$(\psi_x^*\Lambda^{-1}\psi_x - I_x)\hat{\mu}(dx) = 0, \quad \psi_x^*\Lambda^{-1}\psi_x \geq I_x \quad \forall x \in X \tag{3.59}$$

(I_x is the identity element in $\mathcal{U}_x$).
(3) *If the subspace $\mathcal{U}(dx)$ does not intersect with the sum $\overline{\mathcal{U}(dx)}$ of all the remaining subspaces $\mathcal{U}_y$, $y \notin dx$, the operator $\hat{\mu}(dx)$ is strictly positive in $\mathcal{U}_x$ and is defined by the condition $\psi_x^*\Lambda^{-1}\psi_x = I_x$, $x \in X$.*

3.3.4 Waves with Group Symmetry

Equations (3.59) are considerably simpler than Eqn. (3.57) since the dimensionality of each operator equation in (3.59) is equal to rank $r(R_x)$. For the case where $r(R_x) = 1$ the solution has been found [14] under the condition that the square root of the correlation matrix $[\psi_x^*\psi_y]$ has equal diagonal elements. An analog of this condition in the general case where $r(R_x) \geq 1$ is the condition of group (say, cyclic in [12]) symmetry of the family $\{R_x\}$.

Let X be a homogeneous set with respect to a group G, that is, group G acts on X transitively, and let $U(g)$, $g \in G$, be a unitary representation of G in $\mathcal{H}$. The family $\{R_x, x \in X\}$ is said to be G-homogeneous (or G-invariant) if X is a homogeneous set with respect to group G and $U(g)R_{gx}^{-1}U^*(g) = R_x$.

Let G be a finite compact or locally compact group, dg be the left Haar measure on G, the family $\{R_x\}$ be homogeneous and continuous in $U(g)$. The following conjectures are true:

Theorem 17 (1) *The sufficient space $\mathcal{U}$ is a subspace in $\mathcal{H}$ cyclically generated by the family $\{U(g), g \in G\}$ over $\mathcal{U}_0 = R_{x_0}\mathcal{H}$, where x_0 is any element belonging to X. The operators R_x in $\mathcal{U}$ can be represented in the form*

$$R_x = U(g)\psi\psi^*U^*(g) \quad \forall g \in G,$$

where $U(g)$ is a subrepresentation induced in $\mathcal{U} \subset \mathcal{H}$, ψ an operator from $\mathcal{U}_0$ into $\mathcal{U}$, and G_x the left coset $G_x = \{g : gx_0 = x\}$ over the stationary subgroup $G_0 = G_{x_0}$ of element x_0.
(2) *The optimal strategy* (3.58) *has the covariant form*

$$M(dx) = \int_{G(dx)} U(g)\Lambda^{-1}\psi\hat{\mu}\psi^*\Lambda^{-1}U^*(g)\,\mathrm{d}g, \tag{3.60}$$

where $G(dx) = \bigcup_{x \in dx} G_x$ *the union of the* G_x *over all* $x \in dx$,

$$\Lambda = \Big(\int U(g)\psi\hat{\mu}\psi^* U^*(g)\,\mathrm{d}g \Big)^{1/2}$$

is the G*-invariant, and* $\hat{\mu}$ *a non-negative operator in* $\mathcal{U}_0$ *satisfying the conditions*

$$\Big(\psi^*\Lambda^{-1}\psi - \hat{I}\Big)\,\hat{\mu} = 0, \quad \psi^*\Lambda^{-1}\psi \leq \hat{I}. \tag{3.61}$$

(3) *If the representation* $U(g)$ *in* $\mathcal{U}$ *is topologically irreducible, then operator* Λ *is a multiple of the identity element* $\hat{I}$ *of space* $\mathcal{U}_0 : \Lambda = \lambda I$ *and operator* $\hat{\mu}$ *is proportional to the proper projector* $\hat{\pi}$ *of operator* $\hat{R} = \psi^*\psi$ *corresponding to its maximal eigenvalue* $\lambda : \hat{\mu} = \mu\hat{\pi}$, $(\hat{R} - \Lambda)\hat{\pi} = 0$.

The proportionality factor can be made equal to unity by appropriately renormalizing $\mathrm{d}g$. For the particular case where G is finite and its action on X is effective this result was obtained earlier in [39]. If G is an Abelian group, the case is trivial and can be of no interest.

(4) *Let* $U_\omega(g)$, $\omega \in \Omega$, *be the field of nonequivalent irreducible representations* $U_\omega(g)$ *in space* $\mathcal{H}_\omega$, *and* $\mathrm{d}\omega$ *the Plancherel measure. Then* (3.61) *can be represented in the form*

$$\int_\Omega \mathrm{Tr}_{\mathcal{H}_\omega}(\hat{R}_\omega\hat{\mu})^{1/2}\,\mathrm{d}\omega = \hat{\mu}, \quad \int_\Omega \mathrm{Tr}_{\mathcal{H}_\omega}(\hat{R}_\omega\hat{\mu})^{-1/2}\hat{R}_\omega\,\mathrm{d}\omega \leq \hat{I}, \tag{3.62}$$

where $\hat{R}_\omega = \int \hat{r}(g)U_\omega(g)\,\mathrm{d}g$ *is the Fourier transform of the operator correlation function* $\hat{r}(g) = \psi^*U^*(g)\psi$. *If the family* $\{U_\omega(\cdot)\}$ *is discrete and* $\mathrm{d}\omega$ *is the dimensionality of representations* $U_\omega(g)$ (*formally, if* $\mathcal{H}_\omega$ *is infinite-dimensional*), *then conditions* (3.62) *assume the form*

$$\mathrm{Tr}_{\mathcal{H}_\omega}(\hat{R}_\omega\hat{\mu})^{1/2}d\omega = \hat{\mu}, \quad \mathrm{Tr}_{\mathcal{H}_\omega}(\hat{R}_\omega\hat{\mu})^{-1/2}\hat{R}_\omega\;d\omega \leq \hat{I}. \tag{3.63}$$

If $\mathcal{U}_0$ is one-dimensional, (3.62) and (3.63) lead us to the following solutions:

$$\mu = \Big(\int_\Omega \mathrm{Tr}_{\mathcal{H}_\omega}\hat{R}_\omega^{1/2}\,\mathrm{d}\omega\Big)^2, \; \mu = \Big(\sum_\Omega \mathrm{Tr}_{\mathcal{H}_\omega}\hat{R}_\omega^{1/2}\;d\omega\Big)^{1/2},$$

which were found in [8] (that is, the case of group symmetry for pure states is "equidiagonal"). The rank of operators $\hat{R}_\omega$ determines the multiplicity of representations $U_\omega(g)$ in the representation $U(g)$ in $\mathcal{U}$.

(5) *Suppose that the multiplicity of representations* $U_\omega(g)$ *in the representation* $U(g)$ *is unity. Then the operators* $\hat{R}_\omega$ *are one-dimensional,* $\hat{R}_\omega = \psi_\omega^* \otimes \psi_\omega$. *and Eqs* (3.62) *and* (3.63) *assume the form*

$$\Big(\int \hat{S}_\omega\,\mathrm{d}\omega/c_\omega - \hat{I}\Big)\hat{\mu} = 0, \quad \Big(\sum \hat{S}_\omega\;d\omega/c_\omega - \hat{I}_\omega\Big)\hat{\mu} = 0,$$

where $\hat{S}_\omega = \mathrm{Tr}_{\mathcal{H}_\omega}\hat{R}_\omega$, *and* $c_\omega = [\mathrm{Tr}_{\mathcal{U}_0}(\hat{S}_\omega\hat{\mu})]^{1/2}$.

In particular, if all $\hat{S}_\omega$ are commutative, then operator $\hat{\mu}$ is a multiple of the proper projector $\hat{\pi}$ of operators $\hat{S}_\omega$, which corresponds to the eigenvalues λ_ω with maximal

$$\mu = \Big(\int \lambda_\omega^{1/2} \, \mathrm{d}\omega \Big)^2$$

(or $\mu = \Big(\sum \lambda_\omega^{1/2} \, d\omega \Big)^2$): $\hat{\mu} = \mu \hat{\pi}, (\hat{S}_\omega - \lambda\omega)\hat{\pi} = 0.$

3.3.5 Application to a Group Symmetry and Indeterminate Phase

To apply the above results to the case of a decomposable G-homogeneous family of density operators $S_x = \otimes_n R_x^{(n)}$ it is sufficient to supply all the spaces and operators in (3.57)-(3.62) with an index n and then sum over n. In particular, if the representations $U^{(n)}(g)$ in $\mathcal{H}^{(n)}$ are n-th tensor powers of the representation of $U(g)$ in $\mathcal{L}$, then the solution of the problem of optimal recognition of audio and optical fields is reduced to finding the irreducible representations $U_\omega(g)$ contained in $U^{(n)}(g)$. The operators $R_\omega^{(n)}$ determining (3.62) and (3.63) are

$$\hat{R}_\omega^{(n)} = \int \hat{r}^{(n)}(g) U_\omega(g) \, \mathrm{d}g$$

where

$$\hat{r}^{(n)}(g) = \psi^{(n)*} U^{(n)}(g^{-1}) \psi^{(n)}.$$

For example, if the states S_x are Gaussian, the family of signals $\{\varphi_x, x \in X\}$ is G-homogeneous: $U(g)\varphi_{gx} = \varphi_x$, and the correlation noise operator L (or N) is G-invariant: $U(g)LU^*(g) = L$, then the family of the $R^{(n)}$ operators is also G-homogeneous with respect to the appropriate tensor powers $U^{(n)}(g)$ of the representation $U(g)$ in the subspace $\mathcal{U}$ generated by vector $\varphi = \varphi_{x_0}$ for a certain $x_0 \in X$. In this manner we can find the exact solution to the following problems: resolution of several nonorthogonal partially coherent signals or fields that form a homogeneous family of permutations with respect to a certain group (symmetric groups $S(r)$ and their subgroups), estimation of the time lag of pulsed signals and the carrier frequency in quasiperiodic signals (cycle Z groups), joint measurement of the duration and the frequency of a wave packet (the symplectic group), separate or joint measurement of momenta and position of quantum systems (and ensembles of such systems) with r degrees of freedom (the $Z(r)$ groups), detection of photon polarization and electron spin (the $SU(2)$ group), detection of complex signals and fields with equal intensities of rank r against a thermal background (the $SU(r)$ groups and their subgroups), and the like.

Afterword

Optimal wave tomography as mathematical wave-pattern recognition theory, emerged in the 70's first in connection with the problems of optimal estimation and hypothesis testing in quantum theory, is a new pattern recognition theory. The key problems in this theory, mathematical design of the optimal dynamic analyzer discriminating between a given family of pure or mixed *a priori* unknown wave patterns, is thoroughly studied here. Like the problem of optimal quantum measurement it cannot be tackled by the methods of classical mathematical statistics which is not concerned how these date should be obtained from the physical wave states in an optimal way. We extended here the results of optimal quantum measurement theory obtained in [46]–[50] into the direction of wave, rather than particle statistical estimation and hypothesis testing theory, naturally including into the wave tomography not only quantum matter waves but also classical wave patterns like optical and acoustic waves. The developed methods are applied to the problems of mathematical design of optimal wave analyzer discriminating the visual and sound patterns. Thus, Hilbert space and operator methods, developed first in quantum theory of optimal quantum measurement, are found to be equally useful in the classical wave theory where the possible observations are restricted to only intensity distributions of waves, i.e. when the wave states are not the allowed observables, as they are not the observables for individual particles in the quantum theory. It has been shown that all the attributes of quantum measurement theory such as complementarity, entanglements or Heisenberg uncertainty relations have also an exact reflection in the wave pattern recognition theory.

Acknowledgement

The first author (VPB) acknowledges EEC support through the ATESIT project IST-2000-29681 which allowed to cover the expenses for retyping and typesetting this book in TeX.

Bibliography

[1] S.A. Akhmanov and A.S. Chirkin, *Statistical Phenomena in Nonlinear Optics* (Moscow, Moscow Univ. Press, 1971) (in Russian).

[2] P.A. Bakut and S.S. Shehurov, *Problems of Information Transmition,* **4**, No.1, 77–82 (1968).

[3] V.P. Belavkin and B.A. Grishanin, *Problems of Information Transmition,* **8**, No. 3, 103–109 (1972).

[4] V.P. Belavkin, *Radio Engin. Elektron. Physics*, **17**, No.12, 2028–2032 (1972).

[5] V.P. Belavkin, *Radio Engin. Elektron. Physics*, **17**, No.12, 2032–2038 (1972).

[6] V.P. Belavkin, Candidate's Dissertation [in Russian], Moscow State University (1972).

[7] V.P. Belavkin and B.A. Grishanin, *Problems of Information Transmition*, **9**, No.3, 209–215 (1973).

[8] V.P. Belavkin and R.L. Stratonovich, *Radio Engin. Elektron. Physics*, **18**, No.9, 1349–1354 (1973).

[9] V.P. Belavkin, *Problems Control Inform. Theory*, **3**, 47–62 (1974).

[10] V.P. Belavkin and A.G. Vancjan, *Radio Engin. Elektron. Physics*, **19**, No.7, 1397–1401 (1974).

[11] V.P. Belavkin, *Problems Control Inform. Theory*, **4**, No.3, 241–257 (1975).

[12] V.P. Belavkin, *Radio Engin. Elektron. Physics*, **20**, No.6, 1177–1185 (1975).

[13] V.P. Belavkin, *in: Reports of 6th USSR Conference on Coding Theory and Information Transfer*, Vol **6**, (Moscow-Tomsk, 1975) pp 13–18 (in Russian).

[14] V.P. Belavkin, *Stochastics* No.**1**, 315–345 (1975).

[15] V.P. Belavkin, *Zarubezhnaya Radio Elektronika*, No.**5**, 3–29 (1975).

[16] V.P. Belavkin, *Radio Engin. Elektron. Physics*, **21**, No.1, 78–86 (1976).

[17] V.P. Belavkin, *in: Proc. of 4th International Symposium on Information Theory* (Moscow-Leningrad 1976) pp 91–95.

[18] V.P. Belavkin, Translated from *Teoret. Mat. Fizika.*, Plenum, No.**3**, 213–222 (1976).

[19] V.P. Belavkin, *in: Reports of 7th USSR Conference on Coding Theory and Information Transfer*, Vol **7**, (Moscow-Vilnus, 1978) pp 23–28 (in Russian).

[20] V.P. Belavkin, *Problems Control Inform. Theory*, **7**, No.5, 345–360 (1978).

[21] V.P. Belavkin, *in: Optimization Techniques* (Proc. of the 9th IFIP Conference on Optimization Techniques, Warsaw, September 4-8, 1979), Part I, *eds. K. Iracki, K. Malanowski and S. Walukiewicz* (Berlin, Springer, 1980) pp 141–149.

[22] V.P. Belavkin, *Radio Engin. Elektron. Physics*, **25**, No.7, 1445–1453 (1980).

[23] N. Bohr, *Naturwissenschafter* **16**, 245–257 (1928); J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Berlin: Springer, 1932).

[24] H.J. Caulfield (ed.), *Handbook of Optical Holography* (New York, Academic Press, 1979) p. 230.

[25] J.D. Gabor, *Proc. Reg. Phys. Soc. London*, Section B, **64**, 449 (1951).

[26] R.J. Glauber, *Phys. Rev.*, **130**, 2529–2539 (1963).

[27] B.A. Grishanin, *Radio Engineer. Elektron. Physics*, **18**, No.4, 572–577 (1973).

[28] B.A. Grishanin and R.L. Stratonovich, *Problemy Peredači Informacii*, **6**, No. 3, 15–23 (1970).

[29] P. Halmos, *Summa Brasil. Math.*, **2**, 125–134 (1950).

[30] C.W. Helstrom, *Quantum Detection and Estimation Theory* (New York, Academic Press, 1976).

[31] C.W. Helstrom, *Information and Control* **10**, 254–291 (March, 1975).

[32] C.W. Helstrom, *J. Opt. Soc. Amer.*, **60**, 659–666 (1970).

[33] C.W. Helstrom, *J. Opt. Soc. Amer.*, **59**, 164–175 (1969).

[34] C.W. Helstrom, *J. Statist. Phys* **1** 231–252 (1969).

[35] C.W. Helstrom, *Phys. Lett.* **25A**, 101–102 (1967).

[36] A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam, North-Holland, 1982).

[37] A.S. Holevo, *in: AMS Translation Proc. Steklov Math. Inst.*, Issue 3, (1978).

[38] A.S. Holevo, *in: Proc. 2nd Japan-USSR Symp. Prob. Theo.*, Vol.**1** (Kyoto, Japan, Aug 1972) pp 22-40.

[39] A.S. Holevo (Kholevo), *J. Multivariate Anal.* No.**3**, (1973).

[40] R.S. Kennedy, *M.I.T. Res. Lab. Electron. Quart. Progr. Rep.* **110** 142–146 (July 15, 1973).

[41] J.R. Klauder and E.C.G. Sudarshan, *Fundamentals of Quantum Optics* (New York, W.A. Benjamin, 1968).

[42] A.A. Kuriksha, *Quantum Optics and Optical Location* (Moscow, Sov. radio, 1973) (in Russian).

[43] D.G. Luenberger, *Optimization by Vector Space Methods* (New York, Wiley, 1969).

[44] V.P. Maslov, *Théorie des perturbations et méthodes asymptotiques* (Paris, Gauthier-Villars, 1972).

[45] L.L. Myasnikov and E.N. Myasnikova, *Automatic Recognition of Sound Patterns* (Leningrad, Energiya, 1970) (in Russian).

[46] J. von Neumann, *Mathematical Foundation of Quantum Mechanics* (Princeton, New Jersey, Princeton University Press, 1955).

[47] M.A. Neumark, *C.R. Acad. Sci. d'URSS*, **41**, 359–361 (1943).

[48] Yu.P. Pyt'ev, *Kibernetika*, No.**3**, 126–134 (1973).

[49] C.R. Rao, *Linear Statistical Inference and Its Application* (New York, Wiley, 1965).

[50] R.L. Stratonovich, *Stochastics*, **1**, 87–126 (1973).

[51] H.P. Yuen and M. Lax, *IEEE Trans. Inform. Theory* **IT-21**, 125–134 (March, 1975).

[52] H.P. Yuen and M. Lax, *IEEE Trans. Inform. Theory* **IT-19**, No.6, 740–750 (1973).

[53] H.P. Yuen and M. Lax, *Proc. IEEE* **58**, 1770–1773 (1970).