Regularization methods for model- and detector-independent estimation of distributions (the deconvolution problem)

V. B. Anikeev and V. P. Zhigunov
Institute of High Energy Physics, Protvino

Fiz. Elem. Chastits At. Yadra **24**, 989–1055 (July–August 1993)

Modern high-statistics experiments made to measure cross sections and nucleon structure functions include a procedure for correcting the experimental histograms for the limited acceptance and finite resolution of the detector. It is obvious that this must be a model-independent procedure. A survey is given of the following correction methods: the least-squares method (pseudoinversion method), Tikhonov's method, the spectral-window method, the iteration method, the correction-factor method, the maximum-entropy method, etc. For all methods, the statistical and systematic errors in the estimates are analyzed in the framework of a single model that relates the experimental histogram to the studied distribution. The use of the methods is considered for estimating continuous distributions and structure functions. Numerical experiments are used as illustrations.

INTRODUCTION

The aim of many experiments in high-energy physics is to measure differential cross sections of the kinematic variables that describe the final state after a collision of particles or their decay. At the present time, the volume of data obtained in these experiments is so great that it is not the statistical error in the data but the systematic errors in the experiments that become the decisive factor. One of the main sources of systematic errors in experimental data is the finite resolution of the detectors and their acceptance. Therefore, the direct experimental data are detector-dependent estimates of the measured differential cross sections. It is therefore necessary to "correct" the experimental data for the detector acceptance and resolution. In such "corrections" one often uses model arguments, and these, in their turn, lead to model-dependent estimates of the cross sections. It follows from what we have said that the problem has become very topical in recent years and has a very general nature. Its solution has been considered in numerous studies, but in them there has been no general formulation of the problem; it has not been recognized that the problem belongs to the class of improperly posed problems, the solution of which leads either to an unacceptably large statistical error in the estimate of the cross sections or to a systematic error when regularization methods are used to suppress the buildup of measurement errors. As a consequence, there has been no analysis of the existence and magnitude of this error in the "correction" of experimental data. It is obvious that the method for "correcting" the data of an experiment and its analysis must be chosen at the stage when the experiment is planned, since it is only after this that one can see how "good" the estimate of the cross section may be for given detector resolution and intended statistics of the experiment.

Therefore, in this review each method is examined in accordance with the following scheme:

formulation of the mathematical model relating the studied distribution to the experimental data;

the "correction" of the experimental data, which is in fact the solution of an improperly posed inverse problem determined by the chosen model;

different methods of "correcting" the data, which are different methods of regularizing the problem;

the different regularization methods lead to different estimates of the studied cross section;

for the estimate it is necessary to characterize its bias (systematic error) and noise component (statistical error).

We formulate a model of the measurement process,[1] since it will be used in all sections of the review. A large proportion of the presented material is invariant with respect to the number of kinematic variables on which the considered differential cross section depends. Therefore, for brevity we shall speak of a single variable, which we denote by x. By reconstructing the event from the directly measured quantities in the detector and estimating x, we obtain instead of its true value some other value x'. This process can be described as a whole by a conditional probability density $K(x'|x)$, which we shall call the resolution of the experiment with respect to the variable x. In the literature, the standard deviation $\sigma(x)$ is often called the resolution with respect to x, and it is implicitly assumed that $K(x'|x)$ is a normal distribution:

$$N[x'|x,\sigma^2(x)]=[\sqrt{2\pi}\sigma(x)]^{-1}\times\exp\{(x'-x)^2/2\sigma^2(x)\}.$$

The function $K(x'|x)$ is calculated by the Monte Carlo method using a program that simulates the operation of the detector and the reconstruction of the events. We shall consider this question in more detail later.

Other characteristics of the experiment are the function $E(x)$, which is the probability of detection by the detector of an event with given x (detector acceptance or efficiency), and $\tilde{E}(x')$, which is the probability of inclusion of an event with the obtained x' in the final statistics of the experiment after cutoffs with respect to the kinematic variables have been made in the analysis of events.

As a result of these detector distortions, we obtain in place of the studied normalized distribution

 1063-7796/93/040424-29$10.00

$p(x)=d\sigma/dx/\sigma_{\text{tot}}$ the experimental distribution

$$\tilde{p}(x')=c\int dx\tilde{E}(x')K(x'|x)E(x)p(x), \tag{1}$$

where c is a normalization constant, and the integral is over all allowed values of x at the given energy of the beam particle. The integration of $p(x)$ with kernel $K(x'|x)\geqslant 0$ possesses a filtering property and leads to smoothing of the distribution $p(x)$ and loss of its fine structure. The presence of $E(x)$, $\tilde{E}(x')\leqslant 1$ leads to additional distortions. Therefore, the difference between the true distribution $p(x)$ and the experimental one $\tilde{p}(x')$ (the systematic error) may be quite appreciable and exceed the statistical error, and this necessitates a "correction" of the experimental data.

These data are the number of events in the channels of the histogram with respect to the value of x, which we denote by Y_n. It follows from the expression (1) that their mathematical expectations are[1)]

$$\bar{Y}_n=L\int_{x_n}^{x_{n+1}}dx'\tilde{E}(x')\int dxK(x'|x)E(x)\frac{d\sigma}{dx},$$

where L is the integrated luminosity of the experiment. Introducing the notation

$$\begin{aligned}&S(x)=Ld\sigma/dx,\\&A_n(x)=\int_{x_n}^{x_{n+1}}dx'\tilde{E}(x')K(x'|x)E(x),\end{aligned} \tag{2}$$

we can finally express the mathematical model of the experiment in the form[1]

$$Y_n=\int dxA_n(x)S(x)+\varepsilon_n,\quad n=1,...N, \tag{3}$$

where ε_n is the fluctuation of the number of events Y_n in channel n of the histogram. With regard to ε_n, we assume that $\bar{\varepsilon}_n=0$, $\varepsilon_n\varepsilon_m=K_{nm}\approx\delta_{nm}/Y_n$. The functions $A_n(x)$ have a transparent meaning: $A_n(x)$ is the probability that an event with given n is recorded in channel n of the histogram.

The functions $A_n(x)$ can be calculated only by the Monte Carlo method, which is as follows: N events are realized for given x; for each event, "measurements" are obtained by using a detector-simulation program; the value of x' is calculated by using a reconstruction program; the number of events k in bin n of the histogram with respect to x' is counted; the value of $A_n(x)=k/N$ is calculated.[2)]

It is natural to interpret the expression (3) for the measurement of Y_n as the scalar product of the functions $S(x)$ and $A_n(x)$. Introducing this scalar product in the space of the functions $S(x)$, we transform it into a Hilbert space. Then any function $S(x)$ can be expressed as a sum of two components: $S_A(x)$, which belongs to the subspace spanned by the system of functions $A_n(x)$,

$$S_A(x)=\sum_{n=1}^{N}C_nA_n(x), \tag{4}$$

and $S_A^{\perp}(x)$, the component orthogonal to this subspace:

$$\int dxA_n(x)S_A^{\perp}(x)=0. \tag{5}$$

It follows that the component $S_A^{\perp}(x)$ cannot be estimated on the basis of Y_n alone, since it follows from (3) and (5) that Y_n does not depend on it. The function $S_A(x)$ is often called the "measured" ("visible") component of $S(x)$, and $S_A^{\perp}(x)$ is called the "unmeasured" ("invisible") component.

From these considerations there follows an important conclusion: If in the framework of the model (3) $S_A^{\perp}(x)\not\equiv 0$, then the finiteness of the number of measurements leads to a bias in the estimate of the function $S(x)$.

In the framework of the model (3), we consider in Sec. 1 linear estimates of $S(x)$, and in Sec. 2 we discuss nonlinear estimates, using both *a priori* information on the positivity of $S(x)$ (maximum-entropy principle, etc.) and no such information—the correction-factor method. Section 3 is devoted to estimation of nucleon structure functions using the data of deep inelastic lepton scattering, where we shall give a generalization of the model (3) to the simultaneous estimation of two functions.

1. LINEAR ESTIMATES OF CROSS SECTIONS

In presenting methods and numerical illustrations of them, we shall largely follow Ref. 2. Of the two examples used in Ref. 2 to compare the methods, we shall use here the historically earlier one:[1]

$$S(x)=C\left[A_1\frac{G_1^2}{(x-x_1)^2+G_1^2}+A_2\frac{G_2^2}{(x-x_2)^2+G_2^2}\right],$$

$$E(x)=1-\frac{x-x_1}{x_2-x_1+2G_2},\quad \tilde{E}(x')=1,$$

$$K(x'|x)=N(x'|x,\sigma^2),$$

where $x\in[4,16]$, $A_1=2$, $A_2=1$, $G_1=G_2=1$, $x_1=10$, $x_2=14$, $\sigma=1.5$, and C was chosen using the condition $\int dxS(x)=10^4$ (see Fig. 1a).

General properties of linear estimates

In the general case, the linear estimate $\hat{S}(x)$ has the form

$$\hat{S}(x)=\sum_{n=1}^{N}a_n(x)Y_n. \tag{6}$$

It follows from (4) and (5) that for reasonable estimates that do not use *a priori* information $a_n(x)$ must lie in the subspace spanned by $A_n(x)$.

If we use the relation (3), then $\hat{S}(x)$ can be represented as follows:

$$\hat{S}(x)=\sum_n a_n(x)\int dx'A_n(x')S(x')+\sum_n a_n(x)\varepsilon_n. \tag{7}$$

Taking the average (7) over the errors of the measurements, we obtain

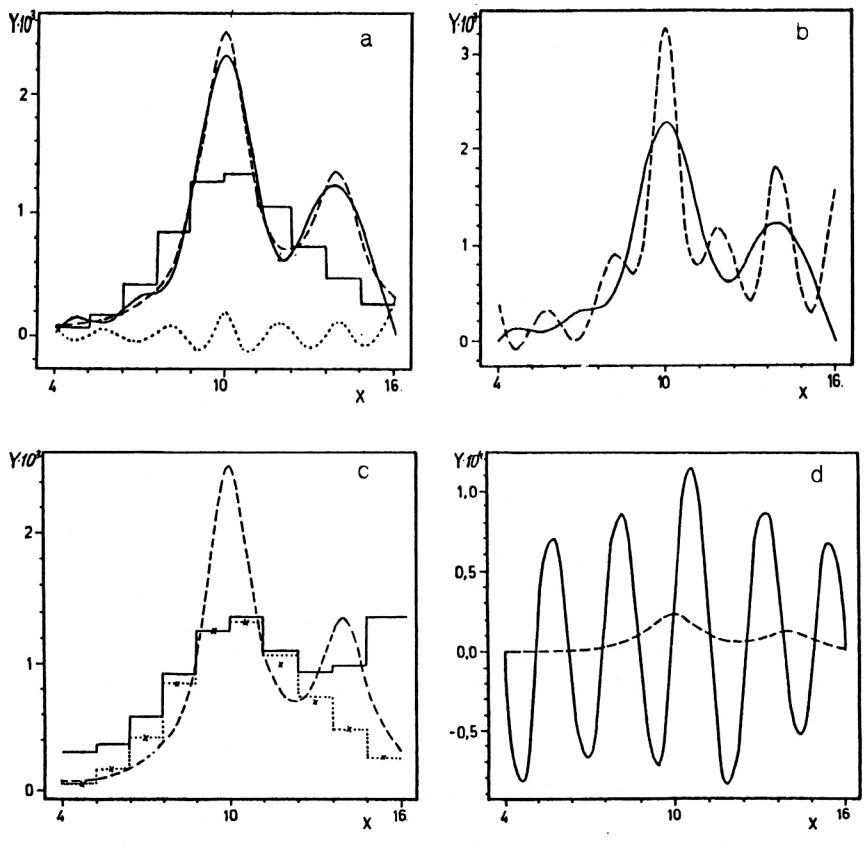

FIG. 1. a) The histogram is the distribution of $(d\bar{Y}/dx)_n$ ($\varepsilon_n=0$), the continuous curve is $S_A(x)$, the broken curve is the function $S(x)$, and the dotted curve is the "unmeasured" component $S_A^{\perp}(x)$; b) the continuous curve is $S_A(x)$, and the broken curve is $S_A(x)+5S_A^{\perp}(x)$; c) the histogram is the step estimate $\hat{S}_0$, the broken curve is the function $S(x)$, the dotted histogram is the distribution of $(d\bar{Y}/dx)_n$ ($\varepsilon_n=0$), and the markers are the modeled distribution of $(d\bar{Y}/dx)_n$ ($\varepsilon_n\neq 0$), $\mathrm{Sp}\, A_R^0=3.03$, $\mathrm{Sp}\, \Sigma^0=1.86\cdot 10^4$; d) the continuous curve is the least-squares estimate $\hat{S}^+(x)$ ($\varepsilon_n\neq 0$), and the broken curve is the function $S(x)$; $\mathrm{Sp}\, A_R^+=10$, $\mathrm{Sp}\, \Sigma^+=1.1\cdot 10^9$.

$$\bar{\hat{S}}(x)=\sum_n a_n(x)\int dx' A_n(x')S(x')$$

$$=\int dx' A_R(x,x')S(x'), \tag{8}$$

where the function

$$A_R(x,x')=\sum_n a_n(x)A_n(x') \tag{9}$$

can be called the residual instrumental function.[1]

It follows from (8) that if $A_R(x,x')$ does not possess the property

$$\int dx' A_R(x,x')S_A(x')=S_A(x), \tag{10}$$

then the estimate $\hat{S}(x)$ will possess an additional bias due to the choice of the particular method of estimation [the specific choice of the functions $a_n(x)$ in (6)].

The noise component of the estimate $\hat{S}(x)$ is characterized, as follows from (7) and (8), by an error operator:[1]

$$\Sigma(x,x')\equiv\overline{[\hat{S}(x)-\bar{\hat{S}}(x)][\hat{S}(x')-\bar{\hat{S}}(x')]}$$

$$=\sum_{n,m} a_n(x)K_{nm}a_m(x'). \tag{11}$$

Thus, the linear estimate of the general form (6) can be interpreted as the result of a "new experiment" with new instrumental function $A_R(x,x')$ and noise component $\xi(x)=\Sigma_n a_n(x)\varepsilon_n$ with the error operator (11).

We consider the mean square of the distance between $S_A(x)$ and $\hat{S}(x)$:

$$\bar{D}^2=\overline{\int dx[S_A(x)-\hat{S}(x)]^2}$$

$$=\overline{\int dx\{[S_A(x)-\bar{\hat{S}}(x)]+[\bar{\hat{S}}(x)-\hat{S}(x)]\}^2}$$

$$=\int dx[S_A(x)-A_R S]^2+\mathrm{Sp}\,\Sigma.$$

It can be seen from this that it is convenient to characterize the noise component of the estimate $\hat{S}(x)$ as a whole by the value of $\mathrm{Sp}\,\Sigma$. In what follows, we shall see that the resolution of the "new experiment" [respectively, the bias of the estimate $\hat{S}(x)$] can be conveniently characterized by the value of $\mathrm{Sp}\, A_R$.[3]

Least-squares method

We consider first of all the properties of estimates based on the least-squares method. On the basis of the arguments about the "measured" component $S(x)$, we shall seek an estimate of $S(x)$ in the form[4,6]

$$\hat{S}(x)=\sum_n C_n A_n(x)$$

from the condition of a minimum of the functional

$$\Phi=\sum_{n,m}\left[Y_n-\int dxA_n(x)\hat{S}(x)\right]W_{nm}$$
$$\times\left[Y_m-\int dxA_m(x)\hat{S}(x)\right],$$

which in this case takes the form

$$\Phi=\sum_{n,m}\left(Y_n-\sum_k G_{nk}C_k\right)W_{nm}\left(Y_m-\sum_l G_{ml}C_l\right), \quad (12)$$

where $W=K^{-1}$, and $G_{nk}=\langle A_n|A_k\rangle=\int dxA_n(x)A_k(x)$ is the Gram matrix of the functions $A_n(x)$. Differentiation of the functional (12) with respect to C_i leads to the following system of equations for the vector C:

$$G^TWG\hat{C}=G^TWY. \quad (13)$$

Since the matrices G and W are positive definite, the solution of this system of equations can be written in the form

$$\hat{C}=G^{-1}Y.$$

Therefore, the least-squares estimate of $S(x)$ is

$$\hat{S}^+(x)=\sum_{n,m}A_n(x)G^{-1}_{nm}Y_m. \quad (14)$$

It is readily seen that $\int dxA_n(x)\hat{S}^+(x)=Y_n$ and that the functional (12) vanishes.

A more formal method[4] of obtaining the estimate (14) is based on the theory of integral equations.[5] Minimization of the least-squares functional

$$\Phi=\sum_n\left[\tilde{Y}_n-\int dx\tilde{A}_n(x)\hat{S}(x)\right]^2,$$

where $\tilde{Y}_n=\sqrt{W_n}$, $\tilde{A}_n(x)=\sqrt{W_n}A_n(x)$ for $W_{nn}=W_n\delta_{nm}$, with respect to the function $S(x)$ leads to an integral equation with degenerate kernel:

$$\sum_n\tilde{A}_n(x)\int dx'\tilde{A}_n(x')S(x')=\sum_m Y_m\tilde{A}_m(x).$$

The general solution of this equation has the form

$$\hat{S}(x)=\sum_{n=1,m=1}^{N}\frac{1}{\lambda_n}U_n(x)\langle U_n|\tilde{A}_m\rangle Y_m+S^{\perp}_A(x),$$

where $S^{\perp}_A(x)$ is any solution of the corresponding homogeneous equation, $U_n(x)$ are eigenfunctions, and λ_n are eigenvalues of the kernel of the equation

$$\sum_n\tilde{A}_n(x)\int dx'\tilde{A}(x')U_n(x')=\lambda_nU_n(x).$$

The linearly independent solutions $S^{\perp}_A(x)$ form a countable set orthogonal to the functions $U_n(x)$. It follows from the last equation that $U_n(x)$ are linear combinations of the functions $\tilde{A}_n(x)$ and, therefore, the first term of the solution $\hat{S}(x)$ [the "measured" component of $S(x)$] is a linear combination of the original instrumental functions $A_n(x)$.

The residual instrumental function $A_R(x,x')$ in the least-squares method,

$$A^+_R(x,x')=\sum_{n,m}A_n(x)G^{-1}_{nm}A_m(x'), \quad (15)$$

possesses the maximum resolution that can be achieved, since it does not distort the "measured" component of $S(x)$ [the relation (10) is satisfied]:

$$\int dx'A^+_R(x,x')S_A(x')$$
$$=\int dx'\sum_{n,m}A_n(x)G^{-1}_{nm}A_m(x')\times\sum_l C_lA_l(x')$$
$$=\sum_{n,m,l}C_lA_n(x)G^{-1}_{nm}G_{ml}=\sum_l C_lA_l(x)=S_A(x). \quad (16)$$

This means that the least-squares estimate $\hat{S}^+(x)$ is an unbiased estimate of the projection of $S(x)$ onto the subspace spanned by the system of functions $A_n(x)$, i.e., the subspace that is accessible for "measurement" by the experiment: $\overline{\hat{S}^+(x)}=S_A(x)$ (see Fig. 1a).

The estimate (14) can be obtained as a solution of the variational problem

$$\min_{S(x)}\Phi=\int dxS^2(x)$$

subject to the condition that $\langle A_n|S\rangle=Y_n$, $n=1,...,N$. The expression (14) is called[6,7] a pseudosolution (pseudoinversion), and $\Sigma_nA_n(x)G^{-1}_{nm}$ is a pseudoinvertible operator for the system of equations $\langle A_n|S\rangle=Y_n$ if the scalar product in the $S(x)$ space is defined as $\langle\tilde{S},\tilde{S}\rangle=\int dx\tilde{S}(x)\tilde{S}(x)$ and in the space of measurements as $\langle\tilde{Y},\tilde{Y}\rangle=\Sigma_{n,m}\tilde{Y}_nW_{nm}\tilde{Y}_m$.

The fulfillment of (10) for $A^+_R(x,x')$ actually means that $A^+_R(x,x')$ in the least-squares method is the projector onto the "measured" part of the space of functions $S(x)$. Indeed, denoting it by P_A, one can show that $P^2_A=P_A$. The orthogonal (to it) projector onto the "measured" part of the space of functions $S(x)$ is

$$P^{\perp}_A(x,x')=\delta(x-x')-P_A(x,x').$$

We calculate the characteristic of the resolution introduced earlier:

$$\mathrm{Sp}\,A^+_R=\int dx\sum_{n,m}A_n(x)G^{-1}_{nm}A_m(x)=N.$$

This agrees with the natural assertion that the greater the number of measurements, the greater the dimension of the "measured" space and the better, in principle, one can construct the estimate $\hat{S}(x)$ on the basis of the measurements Y_n.

The error operator and its trace for the least-squares estimate are given by

$$\Sigma^+(x,x')=\sum_{n,m,l,k}A_n(x)G^{-1}_{nm}K_{ml}G^{-1}_{lk}A_k(x'),$$
$$\mathrm{Sp}\,\Sigma^+=\mathrm{Sp}\,KG^{-1}. \quad (17)$$

We consider from a general point of view the noise component of the estimate $\hat{S}^+(x)$. For sufficiently large values of the nondiagonal elements of the matrix G [the functions $A_n(x)$, $A_m(x)$, $n\neq m$, overlap strongly, and this is something that usually happens in practice], it may be poorly conditioned. A consequence of this is the presence

of small eigenvalues λ of the matrix G. This may lead to an unacceptably strong noise in $\hat{S}^+(x)$, as can be seen from the following arguments.

We introduce the eigenvectors $|g^{(l)}\rangle$ and eigenvalues λ_l of the Gram matrix G:

$$G|g^{(l)}\rangle=\lambda_l|g^{(l)}\rangle, \quad \langle g^{(l)}|g^{(m)}\rangle=\delta_{lm}. \tag{18}$$

Using the theorem on the spectral representation of G^{-1}, we can write

$$G^{-1}=\sum_{l=1}^{N}\frac{|g^{(l)}\rangle\langle g^{(l)}|}{\lambda_l}. \tag{19}$$

Substituting (19) in the expression (14) for $\hat{S}^+(x)$, we obtain

$$\hat{S}^+(x)=\sum_{n,m,l}A_n(x)g_n^{(l)}g_m^{(l)}\frac{(\bar{Y}_m+\varepsilon_m)}{\lambda_l}. \tag{20}$$

It can be seen from this that the noise component of $\hat{S}^+(x)$ is proportional to ε_m/λ_l, which for small λ_l may be very large. This can also be seen from the expression (17) for $\mathrm{Sp}\,\Sigma$ if we consider the special case $K_{nm}=\sigma^2\delta_{nm}$: $\mathrm{Sp}\,\Sigma=\Sigma_l\sigma^2/\lambda_l$.

For our numerical example, $\lambda_{\min}=0.105\cdot10^{-5}$, $\lambda_{\max}=0.821$. The addition to the exact values $\bar{Y}_n$ of relatively small statistical fluctuations ε_n (see Fig. 1c) leads to dramatic growth of the noise component in the estimate $\hat{S}^+(x)$ (see Fig. 1d).

Thus, although the estimate $\hat{S}^+(x)$ has a finite noise component ($\lambda_{\min}\neq0$), its large value makes it necessary to use regularization methods developed for the solution of improperly posed problems.

A general method for suppressing the noise component of $\hat{S}^+(x)$ is to filter $\hat{S}^+(x)$ through a noise-suppressing filter $U(x,x')$: $\hat{S}(x)=\int dx'U(x,x')\hat{S}^+(x')$. If $UP_A^{\perp}=0$, then one can show that $\hat{S}=US_A$.[8] Moreover, any estimate can be expressed as the result of filtering $\hat{S}^+(x)$ through $U(x,x')=A_R(x,x')$ of the given estimate. Indeed, using (6), (9), and (14), we have

$$A_R\hat{S}^+=\sum_n a_n(x)\int dx'A_n(x')\sum_{m,k}A_m(x')G^{-1}_{mk}Y_k$$

$$=\sum_n a_n(x)Y_n. \tag{21}$$

We shall consider the most widely used methods for choosing $A_R(x,x')$ [or $a_n(x)$] later.

One sometimes uses a very simple estimate of $\hat{S}_0(x)$ in the form of a step function that takes into account the detection efficiency $E(x)$ but ignores the deviation of the resolution $K(x,x')$ from a δ function. For comparison, we obtain expressions for $S_0(x)$ and its characteristics $A_R^0(x,x')$, $\Sigma^0(x,x')$, $\mathrm{Sp}\,A_R^0$, $\mathrm{Sp}\,\Sigma^0$.

We seek $\hat{S}_0(x)$ in the form

$$\hat{S}_0(x)=\sum_n a_n\chi_n(n)$$

in accordance with the least-squares method from the condition of a minimum of the functional

$$\Phi=\sum_{n,m}\left[Y_n-\int dx\chi_n(x)E(x)\sum_k a_k\chi_k(x)\right]W_{nm}$$

$$\times\left[Y_m-\int dx\chi_m(x)E(x)\sum_l a_l\chi_l(x)\right],$$

where χ_n is the characteristic function of the nth bin of the experimental histogram:

$$\chi_n(x)=\begin{cases}1, & x\in[x_n,\ x_{n+1})\\ 0, & x\notin[x_n,\ x_{n+1})\end{cases}.$$

Writing $\int dx\chi_n(x)E(x)\chi_n(x)=C_n$, we have

$$\hat{S}_0(x)=\sum_n\chi_n(x)C_n^{-1}Y_n,$$

$$A_R^0(x,x')=\sum_n\chi_n(x)C_n^{-1}A_n(x'),$$

$$\mathrm{Sp}\,A_R^0=\sum_n C_n^{-1}\int_{x_n}^{x_{n+1}}dxA_n(x),$$

$$\Sigma^0(x,x')=\sum_{n,m}\chi_n(x)C_n^{-1}W_{nm}^{-1}C_m^{-1}\chi_m(x'),$$

$$\mathrm{Sp}\,\Sigma^0=\sum_n C_n^{-2}\Delta x_nK_{nn}, \quad \Delta x_n=x_{n+1}-x_n.$$

Figure 1c gives $\hat{S}_0(x)$ and the values of $\mathrm{Sp}\,A_R^0$ and $\mathrm{Sp}\,\Sigma^0$. Comparison of these last quantities with the values of $\mathrm{Sp}\,A_R^+$ and $\mathrm{Sp}\,\Sigma^+$ shows how improvement of the resolution in the estimate $\hat{S}^+(x)$ can lead to a dramatic increase of the noise component. Figure 1c illustrates how allowance for the inefficiency $E(x)$ in the step estimate can lead to an absurd result in the case of bad resolution $K(x,x')$: The large values of Y_n resulting from the bad resolution in the last two bins, corrected for the low efficiency in them, lead to an overshoot in $\hat{S}_0(x)$.

Spectral-window method

As a regularization method, the spectral-window method reduces to suppression of the contribution of small λ_l in (19) by the introduction of a factor $f(\lambda_l)$ under the summation sign:[6]

$$\tilde{G}^{-1}=\sum_{l=1}^{N}f(\lambda_l)\frac{|g^{(l)}\rangle\langle g^{(l)}|}{\lambda_l}, \tag{22}$$

such that $f(\lambda)\leqslant1$ and $f(\lambda)$ is small for small eigenvalues. For the further analysis of the resolution and noise of the estimate obtained from $\hat{S}^+(x)$ by replacing G^{-1} by $\tilde{G}^{-1}$, it is convenient to introduce the functions

$$\Psi_l(x)=\frac{1}{\sqrt{\lambda_l}}\sum_n g_n^{(l)}A_n(x). \tag{23}$$

Using (18), one can show that $\Psi_l(x)$ form an orthonormal basis in the "measured" space:

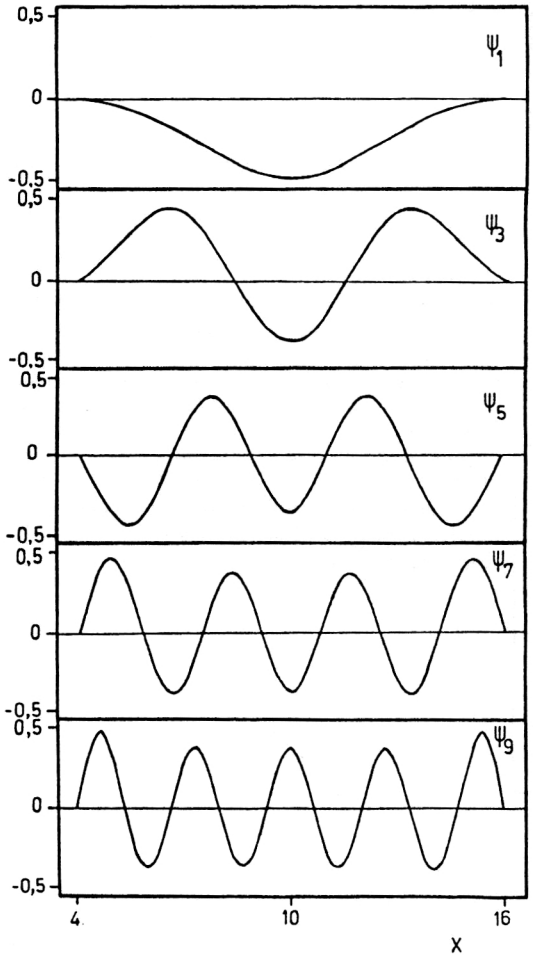

FIG. 2. The functions Ψ_1, Ψ_3, Ψ_5, Ψ_7, Ψ_9 of an orthogonal basis in the measured space.

$$\langle \Psi_l | \Psi_k \rangle = \sum_{n,m} \int dx \frac{1}{\sqrt{\lambda_l}} g_n^{(l)} A_n(x) \frac{1}{\sqrt{\lambda_k}} g_m^{(k)} A_m(x)$$

$$= \sum_{n,m} \frac{1}{\sqrt{\lambda_l \lambda_k}} g_n^{(l)} G_{nm} g_m^{(k)}$$

$$= \sum_{n} \frac{1}{\sqrt{\lambda_l \lambda_k}} g_n^{(l)} \lambda_k g_n^{(k)} = \delta_{lk}.$$

Figure 2 shows $\Psi_n(x)$ for $n=1$, 3, 5, 7, 9.

Making the substitution $G^{-1} \to \tilde{G}^{-1}$ in the expression (14) for $\hat{S}^+(x)$, we obtain

$$\hat{S}(x) = \sum_l \frac{1}{\sqrt{\lambda_l}} \Psi_l(x) f(\lambda_l) \langle g^{(l)} | Y \rangle. \tag{24}$$

From this expression we can see that

$$A_R(x,x') = \sum_l \Psi_l(x) f(\lambda_l) \Psi_l(x'),$$

$$\mathrm{Sp}\, A_R = \sum_l f(\lambda_l) < \mathrm{Sp}\, A_R^+ = N,$$

$$\Sigma(x,x') = \sum_{l,n,m,k} \Psi_l(x) \frac{1}{\sqrt{\lambda_l}} f(\lambda_l) g_n^{(l)} \times K_{nm} g_m^{(k)} f(\lambda_k) \frac{1}{\sqrt{\lambda_k}} \Psi_k(x'),$$

$$\mathrm{Sp}\, \Sigma = \sum_{l,m,n} \frac{1}{\lambda_l} f^2(\lambda_l) g_n^{(l)} K_{nm} g_m^{(l)}. \tag{25}$$

For $K_{nm} = \sigma^2 \delta_{nm}$ we have

$$\mathrm{Sp}\, \Sigma = \sum_l^N f^2(\lambda_l) \frac{\sigma^2}{\lambda_l} < \mathrm{Sp}\, \Sigma^+ = \sum_l \frac{\sigma^2}{\lambda_l}.$$

The last assertion ($\mathrm{Sp}\, \Sigma < \mathrm{Sp}\, \Sigma^+$) also holds, of course, for a general form of the matrix K.

Thus, the spectral-window method leads, on the one hand, to a deterioration of the resolution A_R, i.e., the estimate $\hat{S}(x)$ will have an additional bias compared with $\hat{S}^+(x)$ as a result of the regularization. On the other hand, the noise component of $\hat{S}(x)$ is smaller than the noise component of $\hat{S}^+(x)$.

It is obvious that for $f(\lambda)=1$ the expressions (24) and (25) are the expressions for the least-squares method in the spectral representation, and hence

$$A_R^+(x,x') = \sum_{l=1}^N \Psi_l(x) \Psi_l(x'). \tag{26}$$

We consider the simplest form for $f(\lambda)$: $f(\lambda)=1$ for $\lambda \geqslant \lambda_{N_R}$, $N_R < N$ and $f(\lambda)=0$ for $\lambda < \lambda_{N_R}$. Assuming that λ_l are arranged in descending order, we have

$$A_R(x,x') = \sum_{l=1}^{N_R} \Psi_l(x) \Psi_l(x'). \tag{27}$$

Thus, whereas A_R^+ is the projector onto the entire visible space, A_R is the projector onto the subspace of it spanned by the functions $\Psi_1,\ldots,\Psi_{N_R}$. The choice of N_R is a compromise between a poorer resolution in the estimate (24) and the magnitude of its noise component. We shall consider the choice of N_R in more detail in the section devoted to the choice of the regularization parameter.

In Fig. 3a we compare the estimate $\hat{S}(x)$ by the spectral-window method with $S(x)$, and in Fig. 4a we compare $A_R(x_0,x)$ in the spectral-window method with $f(\lambda)$ taken as a step function with $A_R^+(x_0,x)$ and $A_R^0(x_0,x)$.

If we solve the more complicated generalized eigenvalue problem

$$G\tilde{g}^{(l)} = \tilde{\lambda}_l K \tilde{g}^{(l)}, \quad \langle \tilde{g}^{(l)} | K | \tilde{g}^{(m)} \rangle = \delta_{nm},$$

we can introduce a different orthonormal basis in the "measured" space:

$$\tilde{\Psi}_i(x) = \frac{1}{\sqrt{\lambda_l}} \sum_n \tilde{g}_n^{(l)} A_n(x).$$

It has certain advantages over the basis $\Psi_l(x)$. The Fourier coefficients of $\hat{S}^+(x)$ in the basis $\tilde{\Psi}_l(x)$ are

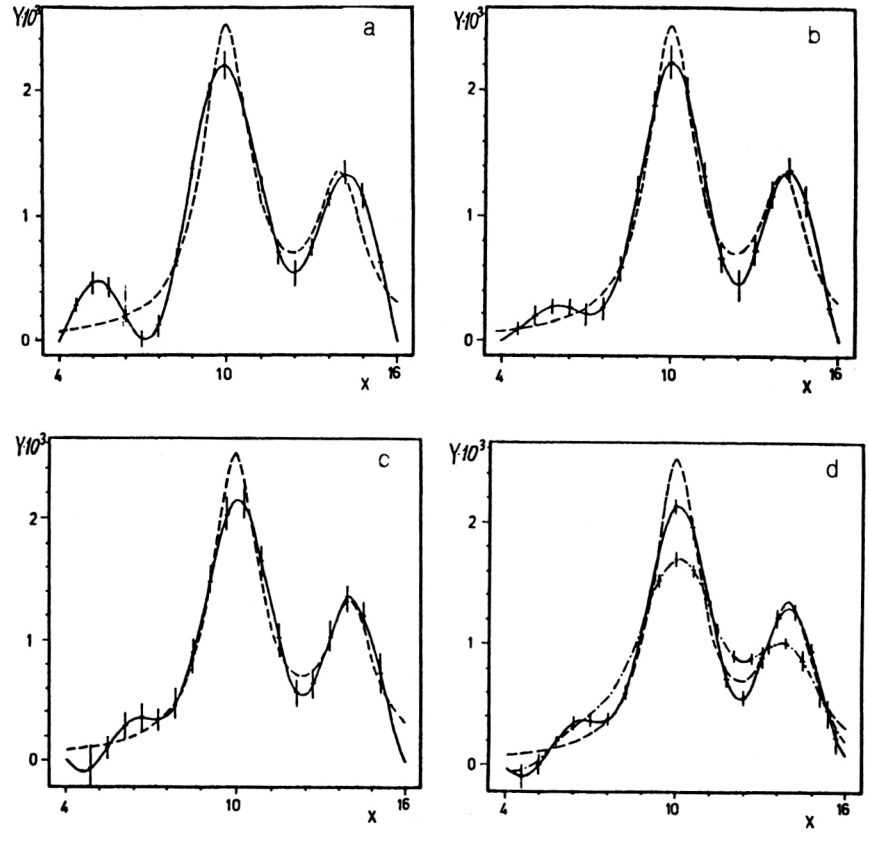

FIG. 3. Comparison of the estimates $\hat{S}$ that minimize $\overline{D^2}$ for different methods. The continuous curve is the estimate $\hat{S}(x)$ ($\varepsilon_n \neq 0$), and the broken curve is the function $S(x)$. As an illustration, we give at some points as errors the values of $\Sigma(x,x)^{1/2}$. a) Spectral-window method, Sp $A_R=5.0$, Sp $\Sigma=6.78\cdot 10^4$; b) iteration method, Sp $A_R=5.49$, Sp $\Sigma=1.11\cdot 10^5$; c) Tikhonov's method, Sp $A_R=5.82$, Sp $\Sigma=1.54\cdot 10^5$; d) dual method: $\sigma=0.3$ for the continuous curve, Sp $A_R=5.70$, Sp $\Sigma=1.58\cdot 10^5$; $\sigma=1.0$ for the chain curve, Sp $A_R=4.11$, Sp $\Sigma=4.97\cdot 10^4$.

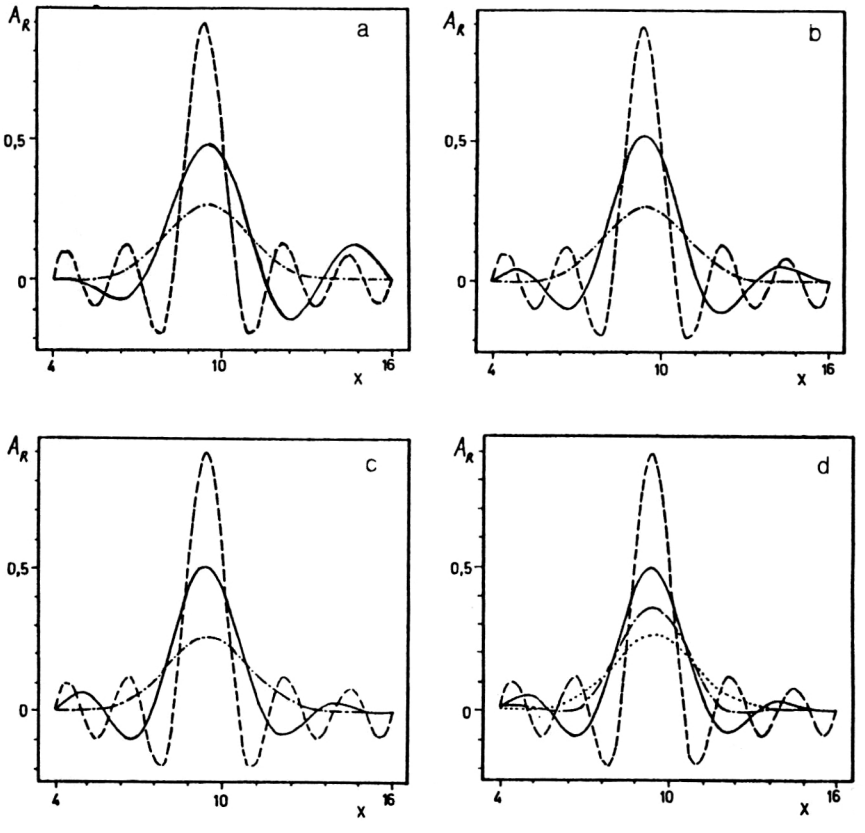

FIG. 4. Comparison of $A_R(x_0,x)$ (continuous curve) with $A_R^+(x_0,x)$ (broken curve) and with $A_R^0(x_0,x)$ (chain curve) for different methods, with $x_0=9.4$. a) Spectral-window method, b) iteration method, c) Tikhonov's method, d) dual method ($\sigma=0.3$ and $\sigma=1.0$).

$$\langle\tilde{\Psi}_l|\hat{S}^+\rangle=\sum_{nmk}\frac{1}{\sqrt{\lambda_l}}\tilde{g}_n^{(l)}G_{nm}G_{mk}^{-1}Y_k=\sum_n\frac{1}{\sqrt{\lambda_l}}\tilde{g}_n^{(l)}Y_n.$$

Therefore, the estimate (24) can be written in the form

$$\hat{S}(x)=\sum_l\frac{f(\lambda_l)}{\sqrt{\lambda_l}}\tilde{\Psi}_l(x)\langle\tilde{g}^{(l)}|Y\rangle.$$

The coefficients of this expansion are not correlated and possess unit variance:[3)]

$$\overline{\langle\tilde{g}^{(l)}|\varepsilon\rangle\langle\varepsilon|\tilde{g}^{(k)}\rangle}=\langle\tilde{g}^{(l)}|K|\tilde{g}^{(k)}\rangle=\delta_{lk},$$

and the expressions (25) take the form

$$A_R(x,x')=\sum_l\tilde{\Psi}_l(x)f(\tilde{\lambda}_l)\tilde{\Psi}_l(x'),$$

$$\mathrm{Sp}\,A_R=\sum_l f(\tilde{\lambda}_l),$$

$$\Sigma(x,x')=\sum_l\tilde{\Psi}_l(x)\frac{f^2(\tilde{\lambda}_l)}{\tilde{\lambda}_l}\tilde{\Psi}_l(x'),$$

$$\mathrm{Sp}\,\Sigma=\sum_l\frac{f^2(\tilde{\lambda}_l)}{\tilde{\lambda}_l}.$$

Iteration method

We consider the use of a different regularization method—the iteration method[6,9]—to suppress the noise in $\hat{S}^+(x)$. In the spirit of our approach, we shall use it to construct an approximate expression for G^{-1}. For the construction of $\hat{S}^+(x)$ it was required to solve the system of equations (13), which is equivalent to the system

$$GC=Y. \tag{28}$$

The simplest iterative scheme of solution of (28), leading to a linear estimate of $S(x)$, is

$$C^{(i+1)}=G^{(i)}-\tau(GC^{(i)}-Y), \tag{29}$$

where i is the number of the iteration, and τ is a relaxation parameter, the choice of which we shall consider later. If as the zeroth approximation we take $C^{(0)}=0$, then we can show that

$$C^{(i)}=\tau\sum_{j=0}^{i-1}(1-\tau G)^jY. \tag{30}$$

Thus, in the iterative scheme (29) the linear estimate at the ith iteration has the form

$$\hat{S}^{(i)}(x)=\sum_{n,m}\tau\sum_{j=0}^{i-1}A_n(x)(I-\tau G)^j_{nm}Y_m$$

$$=\sum_{n,m}A_n(x)\tilde{G}^{-1}_{nm}Y_m. \tag{31}$$

Here, we again denote the approximate value of G^{-1}, which in this scheme is $\tau\sum_{j=0}^{i-1}(I-\tau G)^j$, by $\tilde{G}^{-1}$. The expressions for $A_R(x,x')$ and $\Sigma(x,x')$ have the form of (15) and (17) with G^{-1} replaced by $\tilde{G}^{-1}$. However, the expressions for $\mathrm{Sp}\,A_R$ and $\mathrm{Sp}\,\Sigma$ are changed:

$$\begin{aligned}\mathrm{Sp}\,A_R&=\mathrm{Sp}\,\tilde{G}^{-1}G,\\ \mathrm{Sp}\,\Sigma&=\mathrm{Sp}\,\tilde{G}^{-1}K\tilde{G}^{-1}G.\end{aligned} \tag{32}$$

This scheme for obtaining the iterative estimate becomes more transparent if we again go over to the spectral representation of $\tilde{G}^{-1}$ and the basis $\Psi_l(x)$ in the "measured" space. Using the well-known representation for a function F of a matrix G,

$$F(G)=\sum_l F(\lambda_l)|g^{(l)}\rangle\langle g^{(l)}|,$$

we can write down for the ith iteration

$$\tilde{G}^{-1}=\tau\sum_l^N\sum_{j=0}^{i-1}(1-\tau\lambda_l)^j|g^{(l)}\rangle\langle g^{(l)}|. \tag{33}$$

The sum over j is a sum of terms of a geometric progression, and this enables us to express (33) in the form

$$\tilde{G}^{-1}=\sum_l\frac{1}{\lambda_l}[1-(1-\tau\lambda_l)^i]|g^{(l)}\rangle\langle g^{(l)}|. \tag{34}$$

It follows from this that $\tilde{G}^{-1}$ will converge to G^{-1} if $0<\tau\lambda_{\max}<2$. However, to reduce the bias of the estimate $\hat{S}^{(i)}$ for a finite number of iterations, it is expedient to choose τ in accordance with the condition $\tau\cong 1/\lambda_{\max}$. In this case, the function $f_i(\lambda)=[1-(1-\tau\lambda)^i]$ at large λ is near unity, while for small λ it is near zero. Using (34) for $\tilde{G}^{-1}$, we can express $\hat{S}^{(i)}(x)$, $A_R(x,x')$, $\mathrm{Sp}\,A_R$, $\Sigma(x,x')$, $\mathrm{Sp}\,\Sigma$ for the iteration method in the form (24)–(25) with the function $f_i(\lambda)$ defined above.

Thus, the iteration method again leads to a stabilization of the estimate $\hat{S}(x)$ with respect to the input noise ε_n but at the same time it makes the resolution poorer. The choice of the number of iterations N_i (which plays the role of a regularization parameter in this case) is a compromise between the magnitude of the noise in the estimate and its resolution. In Fig. 3b, we compare the estimate (31) in the iteration method with $S(x)$, and in Fig. 4b we compare $A_R(x_0,x)$ in the iteration method with $A_R^+(x_0,x)$ and $A_R^0(x_0,x)$.

Tikhonov's regularization method

In this method,[10] the estimate of $S(x)$ is determined from a functional that generalizes the least-squares functional:

$$\Phi=\sum_{n,m}(Y_n-\bar{Y}_n)W_{nm}(Y_m-\bar{Y}_m)$$

$$+\alpha\int dx\left[\frac{d^kS(x)}{dx^k}\right]^2, \tag{35}$$

where the second term is called the stabilizer, and $\alpha>0$ is a regularization parameter.

We consider the case $k=0$. Variation of (35) with respect to $S(x)$ leads to the following equation for $\hat{S}(x)$:

$$\alpha\hat{S}(x)+\sum_{n,m}A_n(x)W_{nm}\int dx'A_m(x')\hat{S}(x')$$

$$=\sum_{n,m}A_n(x)W_{nm}Y_m. \tag{36}$$

Therefore, $\hat{S}(x)$ can be written in the form

$$\hat{S}(x)=\frac{1}{\alpha}\sum_n A_n(x)\sum_m W_{nm}\left[Y_m-\int dx' A_m(x')\hat{S}(x')\right]$$

$$\equiv\sum_n A_n(x)\Delta_n. \tag{37}$$

Substituting (37) in the definition of Δ_n, we arrive at a system of equations for Δ_n:

$$\Delta=\frac{1}{\alpha}W[Y-G\Delta],$$

which can be represented in the two equivalent forms

$$[\alpha I+WG]\Delta=WY,\quad [\alpha K+G]\Delta=Y,$$

and accordingly the solution can be expressed in the two forms

$$\Delta=[\alpha I+WG]^{-1}WY, \tag{38}$$

$$\Delta=[\alpha K+G]^{-1}Y. \tag{39}$$

As follows from (37) and, for example, (39), the estimate $\hat{S}(x)$ obtained by Tikhonov's regularization method is, as in the previous methods, obtained from $\hat{S}^+(x)$ by replacing G^{-1} by the approximate value

$$\tilde{G}^{-1}=[\alpha K+G]^{-1}. \tag{40}$$

Using the notation (40), we can write

$$\hat{S}(x)=\sum_{n,m}A_n(x)\tilde{G}^{-1}_{nm}Y_m,$$

$$A_R(x,x')=\sum A_n(x)\tilde{G}^{-1}A_m(x'), \tag{41}$$

$$\mathrm{Sp}\,A_R=\mathrm{Sp}\,G\tilde{G}^{-1},$$

$$\Sigma(x,x')=\sum_{n,m,l,k}A_n(x)\tilde{G}^{-1}_{nm}K_{ml}\tilde{G}^{-1}_{lk}A_k(x'),$$

$$\mathrm{Sp}\,\Sigma=\mathrm{Sp}\;\tilde{G}^{-1}K\tilde{G}^{-1}G.$$

If we assume $K_{nm}=\sigma^2\delta_{nm}$ and go over to the spectral representation (40),

$$\tilde{G}^{-1}=\sum_l[\alpha\sigma^2+\lambda_l]^{-1}|g^{(l)}\rangle\langle g^{(l)}|$$

$$=\sum\frac{f(\lambda_l)}{\lambda_l}|g^{(l)}\rangle\langle g^{(l)}|,$$

where $f(\lambda)=\lambda/(\alpha\sigma^2+\lambda)$, then the expressions (41) can be expressed in the form (24), (25) with replacement of the previous $f(\lambda)$ by the new one. Accordingly, the conclusions drawn concerning the spectral-window method will also hold for Tikhonov's method. In Fig. 3c, we compare the estimate by Tikhonov's method with $S(x)$, and in Fig. 4c we compare $A_R(x_0,x)$ in Tikhonov's method with $A^+_R(x_0,x)$ and $A^0_R(x_0,x)$.

We consider the case $k=2$, following Ref. 1. In this case, the variation of (35) with respect to $S(x)$ leads to the following equation for $\hat{S}(x)$:

$$\alpha\frac{d^4\hat{S}}{dx^4}+\sum_{n,m}A_n(x)W_{nm}\int dx' A_m(x')\hat{S}(x')$$

$$=\sum_{n,m}A_n(x)W_{nm}Y_m \tag{42}$$

with the boundary conditions

$$\frac{d^2\hat{S}}{dx^2}\delta\left(\frac{d\hat{S}}{dx}\right)\Big|_a^b=\frac{d^3\hat{S}}{dx^3}\delta(\hat{S})\Big|_a^b=0.$$

If we know $S|_{x=a}=dS/dx|_{x=a}$, then, requiring $d^2\hat{S}/dx^2|_{x=b}=d^3\hat{S}/dx^3|_{x=b}=0$, we can find the solution of (42) as follows. For these boundary conditions, there exists a Green's function for the equation

$$\frac{d^4F(x,x')}{dx^4}=\frac{d^4F(x,x')}{dx'^4}=\delta(x-x'),$$

and it has the form

$$F(x,x')=\begin{cases}(x-a)^2(3x'-x-2a)/6, & x\leqslant x',\\ (x'-a)^2(3x-x'-2a)/6, & x>x'.\end{cases} \tag{43}$$

Multiplying Eq. (42) by the Green's function (43) and integrating over x, we obtain

$$\hat{S}(x)=\frac{1}{\alpha}\sum_{n,m}U_n(x)W_{nm}\left[Y_m-\int dx' A_m(x')\hat{S}(x')\right]$$

$$\equiv\sum_n U_n(x)\Delta_n,$$

where

$$U_n(x)=\int dx' F(x,x')A_n(x'). \tag{44}$$

Proceeding as in the case for $k=0$, we obtain

$$\Delta=[\alpha K+M]^{-1}Y, \tag{45}$$

where

$$M_{nm}=\int dx dx' A_n(x)F(x,x')A_m(x').$$

Thus, in this case the estimate $\hat{S}(x)$ and its characteristics have the form

$$\hat{S}(x)=\sum_{n,m}U_n(x)[\alpha K+MM]^{-1}_{nm}Y_m,$$

$$A_R(x,x')=\sum_{n,m}U_n(x)[\alpha K+M]^{-1}_{nm}A_m(x'),$$

$$\mathrm{Sp}\,A_R=\mathrm{Sp}[\alpha K+M]^{-1}\langle A|U\rangle, \tag{46}$$

$$\Sigma(x,x')=\sum U_n(x)[\alpha K+M]^{-1}_{nm}K_{mi}[\alpha K+M]^{-1}_{ij}U_j(x'),$$

$$\mathrm{Sp}\,\Sigma=\mathrm{Sp}[\alpha K+M]^{-1}K[\alpha K+M]^{-1}\langle U|U\rangle.$$

The fact that the estimate in this case is a linear combination of the functions $U_n(x)$, and not $A_n(x)$, can be explained as follows. The stabilizer for $k=2$ is not the L_2 norm, which we used earlier, but

$$\|S\|^2=\int dx[d^2S/dx^2]^2,$$

which corresponds to a scalar product in the form

$$\langle\tilde{S}|S\rangle=\int dx(d^2\tilde{S}/dx^2)(d^2S/dx^2).$$

Then, assuming that $S_A^{\perp}=dS_A^{\perp}/dx=0$ for $x=a$, and using the boundary conditions for $\hat{S}_A(x)$ and the condition of orthogonality of $S_A(x)$ and $S_A^{\perp}(x)$ for this scalar product, we can write

$$\langle S_A|S_A^{\perp}\rangle=\int dxS_A^{\perp}(d^4S_A/dx^4)=0.$$

This condition will be satisfied if d^4S_A/dx^4 is a linear combination of the functions $A_n(x)$, as Eq. (42) requires, and $\hat{S}(x)$ is, accordingly, a linear combination of the functions $U_n(x)$.

The expressions analogous to (46) for the boundary conditions

$$d^2\hat{S}/dx^2|_{x=a}=d^3\hat{S}/dx^3|_{x=a}=d^2\hat{S}/dx|_{x=b}$$
$$=d^3\hat{S}/dx^3|_{x=b}=0$$

can be found in Ref. 1. As can be seen from the numerical example considered in Ref. 1, the estimate $\hat{S}(x)$ is very close to the solution for $k=0$.

In Ref. 11 a different approach was considered for $k=2$. It is based on representation of $\hat{S}(x)$ in the form

$$\hat{S}(x)=\sum_{i=1}^{I}a_ib_i(x),$$

where $b_i(x)$ are B splines, with $I\leqslant N$.[4)] Substitution of this expression in the least-squares functional gives

$$\Phi_{\text{LSM}}=\sum_{n,m}\left(Y_n-\sum_i A_{ni}a_i\right)W_{nm}\left(Y_m-\sum_j A_{mj}a_j\right)$$
$$=a^TA^TWAa-a^TA^TWY-YWAa+YWY,$$

where the matrix A is equal to $A_{ni}=\langle A_n|b_i\rangle$. We introduce the eigenvectors and eigenvalues of the matrix $G=A^TWA$:

$$Gg^{(l)}=\lambda_l^{(l)},\quad \langle g^{(l)}|g^{(m)}\rangle=\delta_{lm},\quad l=1,...,I.$$

Substituting in the first term of the functional the spectral representation of the matrix G,

$$G_{nm}=\sum_l g_n^{(l)}\lambda g_m^{(l)},$$

we rewrite it in the form

$$a^TGa=\sum_{l,i,j}ag_i^{(l)}\lambda g_j^{(l)}a_j=\sum_l\tilde{a}_l^2,$$

where the new, $\tilde{a}_l$, and old, a_i, parameters are related by

$$\tilde{a}_l=\sqrt{\lambda_l}\sum_i g_i^{(l)}a_i,\quad a_i=\sum_l\frac{1}{\sqrt{\lambda_l}}g_i^{(l)}\tilde{a}_l.$$

Substitution in the functional of the expression for a_i in terms of $\tilde{a}_l$ and minimization of the functional with respect to $\tilde{a}_l$ gives for them the estimate

$$\hat{\tilde{a}}_l=\sum_{l,n,m}\frac{1}{\sqrt{\lambda_l}}g_i^{(l)}A_{ni}W_{nm}Y_m.$$

It can be shown that the covariance matrix of the estimates $\hat{\tilde{a}}_l$ is equal to the unit matrix I. It follows from this that the variance of a_i,

$$D(a_i)=\sum_l(g_i^{(l)})^2/\lambda_l,$$

becomes unacceptably large in the presence of small λ_l. Therefore, it is necessary to use a regularization. The author of Ref. 11 proposes that one should find an estimate of the parameters a_i from Tikhonov's functional for $k=2$:

$$\Phi=\Phi_{\text{LSM}}+\frac{\alpha}{2}\int dx[S''(x)]^2=\Phi_{\text{LSM}}+\frac{\alpha}{2}a^TCa.$$

After transition in the stabilizer to the variables $\tilde{a}_l$, we can write

$$a^TCa=\sum\tilde{a}g_i^{(l)}\frac{1}{\sqrt{\lambda_l}}C_{ij}\frac{1}{\sqrt{\lambda_m}}g_j^{(m)}\tilde{a}_m\equiv\tilde{a}^TC_1\tilde{a}.$$

Introducing the eigenvectors and eigenvalues of the matrix C_1,

$$C_1f^{(l)}=\mu_lf^{(l)},\quad \langle f^{(l)}|f^{(m)}\rangle=\delta_{lm},\quad l=1,...,I,$$

we write a^TCa in the form

$$a^TCa=\sum_{l,m,i}\tilde{a}_lf_l^{(i)}\mu_if_m^{(i)}\tilde{a}_m\equiv\sum_i\tilde{\tilde{a}}_i\mu_i\tilde{\tilde{a}}_i,$$

where

$$\tilde{\tilde{a}}_i=\sum_i f_l(i)\tilde{a}_l,\quad \tilde{a}_l=\sum_i f_l^{(i)}\tilde{\tilde{a}}_i,\quad \sum_l\tilde{a}_l^2=\sum_i\tilde{\tilde{a}}_i^2.$$

Substitution of $\tilde{a}_l$, expressed in terms of $\tilde{\tilde{a}}_i$, in the remaining terms of the functional Φ and differentiation of it with respect to $\tilde{\tilde{a}}_i$ gives

$$\hat{\tilde{\tilde{a}}}_i=\sum_{l,j,n,m}(1+\tau\mu_i)^{-1}f_l^{(i)}\frac{1}{\sqrt{\lambda_l}}g_j^{(l)}A_{nj}W_{nm}Y_m$$
$$\equiv(1+\tau\mu_i)^{-1}\hat{\tilde{\tilde{a}}}_i(\tau=0).$$

Thus, the regularized solution is close to the unregularized one for $\tau\mu_i\ll 1$ and is close to zero for $\tau\mu\gg 1$, suppressing the contribution of the strongly noise-generating $\hat{\tilde{\tilde{a}}}_i(\tau=0)$.

The author does not introduce a residual instrumental function A_R, asserting that by the choice of α the systematic error can be made much smaller than the statistical error. This, of course, is true, but at the same time the noise component of the estimate may become unacceptably large. In addition, it remains an open question whether or not $b_i(x)$ belongs to the "measured" space.

The numerical example of this study was used here to illustrate unfolding by the maximum-entropy method. The estimates $\hat{S}(x)$ and their statistical errors in the two methods were found to be quite close to each other.

Dual method

Conceptually, this approach follows the studies of Refs. 7, 12, and 13 and is as follows. Let $A_0(x,x')$ be a resolution function with the desired properties. We introduce a "distance" between $A_0(x,x')$ and the function $A_R(x,x')$, which has the form (9) (Hilbert–Schmidt norm):

$$\|A_R-A_0\|_2^2$$

$$=\int dxdx'[A_R(x,x')-A_0(x,x')][A_R(x',x)-A_0(x',x)]$$

$$=\int dx\sum_{n,m}a_n(x)G_{nm}a_{m(x)}-2\int dxa_n(x)b_n(x)+B, \tag{47}$$

where

$$b_n(x)=\int dx'A_n(x')A_0(x,x'),$$

$$B=\int dxdx'A_0(x,x')A_0(x',x).$$

We now seek $a_n(x)$ from the condition of a minimum of the "distance" between A_R and A_0 under the condition that Sp Σ [see (11)] is equal to γ_1. In accordance with Lagrange's method, this problem reduces to the determination of the stationarity point of the functional

$$\Phi=\|A_R-A_0\|_2^2+\lambda[\text{Sp }\Sigma-\gamma_1]. \tag{48}$$

Varying (48) with respect to $a_n(x)$ and differentiating (48) with respect to λ, we obtain

$$\sum_m[\lambda K+G]_{nm}a_m(x)=b_n(x), \tag{49}$$

$$\text{Sp }\Sigma=\gamma_1. \tag{50}$$

If not all $b_n(x)$ are equal to zero, i.e., the "measured" spaces $A_0(x,x')$ and those of the problem (3) overlap, then (49) has the solution

$$a_n(x)=\sum_m[\lambda K+G]_{nm}^{-1}b_m(x). \tag{51}$$

Substituting $a_n(x)$ from (51) in (6), we obtain the estimate $\hat{S}(x)$, from which we can find an expression for Sp Σ containing λ. Substituting this expression in (50), we obtain for λ the equation

$$\text{Sp}[\lambda K+G]^{-1}K[\lambda K+G]^{-1}\langle b|b\rangle=\gamma_1. \tag{52}$$

Using for $\hat{S}(x)$ the expression (6), we have [Sp Σ is determined by the expression (52)]

$$\hat{S}(x)=\sum_n b_n(x)[\lambda K+G]_{nm}^{-1}Y_m.$$

$$A_R(x,x')=\sum_n b_n(x)[\lambda K+G]_{nm}^{-1}A_m(x'), \tag{53}$$

$$\text{Sp }A_R=\text{Sp}[\lambda K+G]^{-1}\langle A|b\rangle,$$

$$\Sigma(x,x')=\sum_{n,m,i,j}b_n(x)[\lambda K+G]_{nm}^{-1}K_{mi}$$

$$\times[\lambda K+G]_{ij}^{-1}b_j(x').$$

We take as $A_0(x,x')$ the operator $A_R^+(x,x')$. Using the definition of $b_n(x)$, we can show that in this case $b_n(x)=A_n(x)$, and the estimate (53) is equal to the estimate (41) of Tikhonov's regularization method if the Lagrangian multiplier λ is identified with the regularization parameter α. In this way, Tikhonov's regularization method acquires a new interpretation when it is used to solve the problem (3).

On the other hand, using the definition of $b_n(x)$ we can write

$$\hat{S}(x)=\int dx'A_0(x,x')\sum_{n,m}A_n(x')[\lambda K+G]_{nm}^{-1}Y_m,$$

i.e., $\hat{S}(x)$ is the instrumental response $A_0(x,x')$ to the estimate of $S(x)$ by Tikhonov's method.

The proposed approach to obtaining the estimate $\hat{S}(x)$ has a symmetric formulation: We shall seek $a_n(x)$ from the condition of a minimum of Sp Σ under the condition that $\|A_R-A_0\|_2^2=\gamma_2$. Such a problem reduces to the search for the stationary point of the functional

$$\Phi=\text{Sp }\Sigma+\lambda[\|A_R-A_0\|_2^2-\gamma_2]. \tag{54}$$

Variation of (54) with respect to $a_n(x)$ and differentiation of (54) with respect to λ leads to the equations

$$\sum_m\left[\frac{1}{\lambda}K+G\right]_{nm}a_m(x)=b_n(x), \tag{55}$$

$$\|A_R-A_0\|_2^2=\gamma_2. \tag{56}$$

Comparing (49) and (55), we see that the functions $a_m(x)$ in the two formulations differ only in the equation for the Lagrangian multipliers. Therefore, the conclusions that we drew concerning $\hat{S}(x)$ in (46) are also true for $\hat{S}(x)$ in (54).

For the numerical illustration, we chose $A_0(x,x')$ in the form of a Gaussian function: $A_0(x,x')=N(x'|x,\sigma^2)$. Figure 3d gives two estimates of (51) with $\sigma=1.0$ and $\sigma=0.3$. It can be seen from the figure that the estimate with $\sigma=1.0$ has a large systematic error. This is explained by the fact that in this case $b_n(x)$ in (53) are too "broad" to describe the behavior of $S(x)$. Therefore, in the other figures we also give an estimate and its characteristics with σ much smaller than the σ in the original resolution function $K(x|x')$. Note that in this case the estimate differs little from the estimate by Tikhonov's method for $k=0$. In Fig. 4d, we compare $A_R(x_0,x)$ in this case with $A_R^+(x_0,x)$ and $A_R^0(x_0,x)$.

On the optimality of Tikhonov's method

Given the existence of several methods of estimating $S(x)$, it is natural to ask which is the best of them in accordance with some criterion. As the criterion, we consider the mean distance between the "measured" component $S_A(x)$ and the required estimate $\hat{S}(x)$:

$$\overline{D_A^2}=\int dx\overline{[S_A(x)-\hat{S}(x)]^2}$$

$$=\int dx[S_A(x)-A_RS]^2+\mathrm{Sp}\,\Sigma. \qquad (57)$$

The value of $\overline{D_A^2}$ is determined by the systematic error [first term of (57)] and the statistical error [second term of (57)]. Minimization of (57) with respect to A_R would lead to the optimum estimate $\hat{S}(x)$. However, since the systematic error contains the unknown function $S_A(x)$, this problem cannot be solved.

However, it is possible to find the optimum A_R under weaker conditions. We find an expression that majorizes $\overline{D_A^2}$. Using the expressions

$$A_R^+=\sum_{n,m}A_n(x)G^{-1}_{nm}A_m(x),$$

$$S(x)=S_A(x)+S_A^{\perp}(x),$$

$$A_R^+S=A_R^+S_A=S_A,$$

$$A_R=\sum_n a_n(x)A_n(x),$$

$$\mathrm{Sp}\,\Sigma=\sum_{n,m}\int dx a_n(x)K_{nm}a_m(x),$$

which we obtained earlier, we can write

$$\overline{D^2}=\int dx[(A_R^+-A_R)S_A]^2+\mathrm{Sp}\,\Sigma$$

$$\leqslant\|A_R^+-A_R\|_2^2\|S_A\|^2+\mathrm{Sp}\,\Sigma\equiv M(A_R),$$

where $\|A_R^+-A_R\|_2^2$ is the operator norm (Hilbert–Schmidt norm), which is consistent with the norm

$$\|S_A\|^2=\int dx[S_A(x)]^2.$$

We now find an expression for A_R from the condition of a minimum of $M(A_R)$. Substituting in $M(A_R)$ the explicit expressions for A_R^+, A_R, and Sp Σ, we obtain

$$M(A_R)=\left[N-2\sum_n\int dx a_n(x)A_n(x)\right.$$

$$\left.+\sum_{n,m}\int dx a_n(x)G_{nm}a_m(x)\right]\|S_A\|^2$$

$$+\sum_{n,m}\int dx a_n(x)K_{nm}a_n(x). \qquad (58)$$

Varying (58) with respect to $a_n(x)$, we obtain

$$\sum_n\left[\frac{1}{\|S_A\|^2}K+G\right]_{nm}a_m(x)=A_n(x). \qquad (59)$$

Substituting the expression for $a_n(x)$ from (59) in the expression for A_R, we obtain

$$A_R(x,x')=\sum_{n,m}A_n(x)\left[\frac{1}{\|S_A\|^2}K+G\right]^{-1}_{nm}A_m(x'). \qquad (60)$$

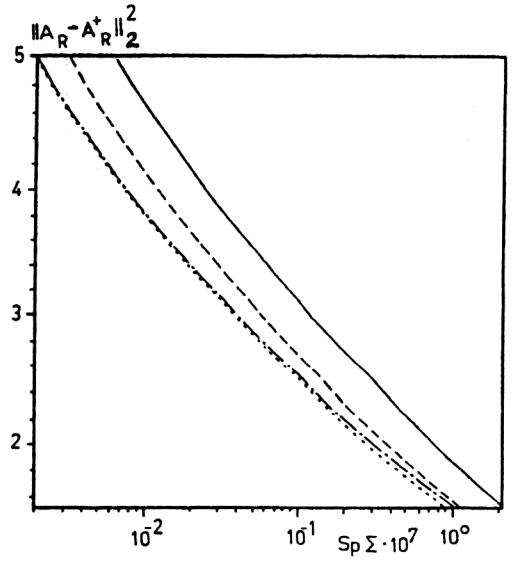

FIG. 5. Dependence of $\|A_R^+-A_R\|_2^2$ on Sp Σ. The continuous curve represents the spectral-window method, the broken curve is for the iteration method, the dotted curve is for Tikhonov's method, and the chain curve is for the dual method.

Comparing (60) with A_R from (41), we see that the A_R that minimizes $M(A_R)$ is the A_R of Tikhonov's method for $k=0$ when $\alpha=1/\|S_A\|^2$.

However, will this property of minimality of the majorant for Tikhonov's method be preserved for other values of α? It is interesting to compare the values of $M(A_R)$ for the different methods for the same value of the statistical error Sp Σ. If we set Sp $\Sigma=\gamma$, then, since $\|S_A\|^2$ is a constant, the minimum of $M(A_R)$ means the minimum of $\|A_R^+-A_R\|_2^2$. Hence, the A_R that gives such a minimum is a stationary point of the Lagrange function

$$\Phi=\|A_R^+-A_R\|_2^2+\lambda[\mathrm{Sp}\,\Sigma-\gamma].$$

However, it was shown in the previous section in the analysis of the functional (48) that this will be A_R for Tikhonov's method for $k=0$. The numerical calculations given in Fig. 5 illustrate this assertion.

Thus, Tikhonov's method is optimal from the point of view of the following criterion: a minimum of the majorant of the systematic error for a given value of the statistical error.

On the choice of the regularization parameter

By the regularization parameter α we shall here understand $1/N_R$ in the spectral-window method, the inverse $1/N_i$ of the number of iterations in the iteration method, α in Tikhonov's method, and λ in the dual method. It is obvious that the optimal value of α will be the one that minimizes $\overline{D^2}$. It was for this value that we gave the estimates $\hat{S}(x)$ in Fig. 3. In a real experiment, it is impossible to calculate $\overline{D^2}$, and one can rely only on analysis of the following quantities: Sp Σ, the characteristic of the noise component; Sp A_R or $\|A_R^+-A_R\|_2^2$, the characteristics of the resolution (systematic error); and

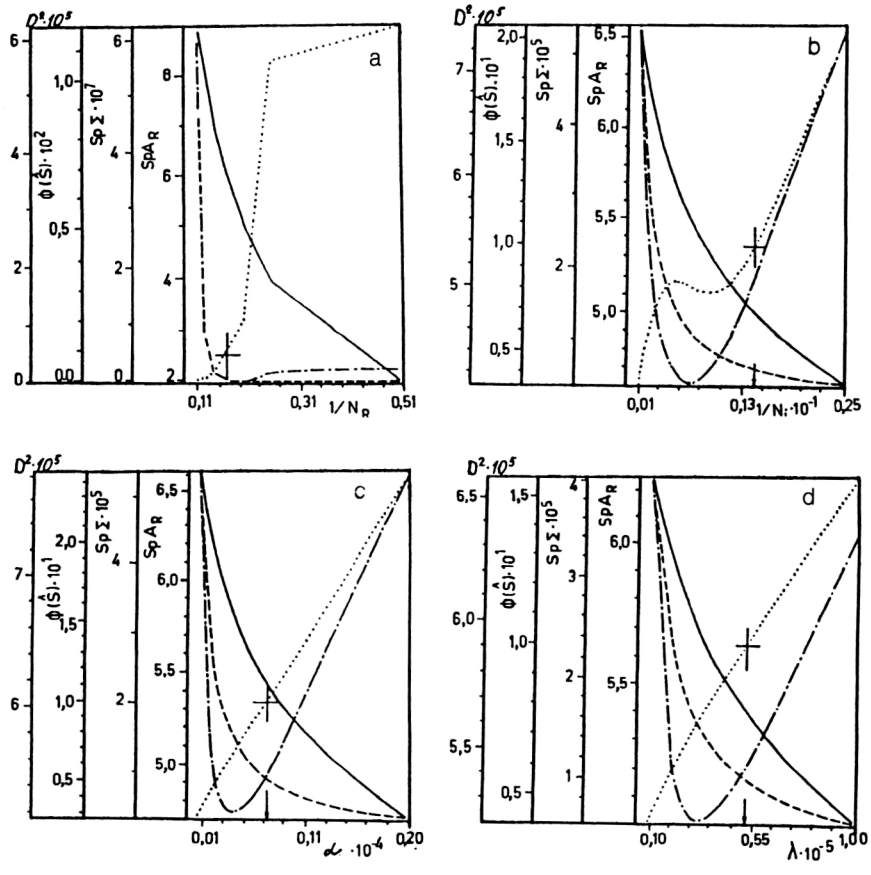

FIG. 6. Dependences of $\mathrm{Sp}\,A_R$ (continuous curve), $\mathrm{Sp}\,\Sigma$ (broken curve), $\Phi(\hat{S})$ (dotted curve), and $\overline{D^2}$ (chain curve) on the regularization parameter. The cross in the diagram indicates the point $\Phi(\hat{S})=N$, and the arrow shows the corresponding value of the regularization parameter. a) Spectral-window method, b) iteration method, c) Tikhonov's method, d) dual method.

$$\Phi(\hat{S})=\sum_{n,m}[Y_n-\bar{Y}_n(\hat{S})]W_{nm}[Y_m-\bar{Y}_m(\hat{S})], \qquad (61)$$

the characteristic of the degree of agreement between $\hat{S}(x)$ and the experimental data.

Graphs of these quantities are shown in Fig. 6.[5)] It can be seen from them that for the considered model only the dependence of $\mathrm{Sp}\,\Sigma$ on α has a singularity in the region of the minimum value of $\overline{D^2}$. In the majority of considered cases, $\mathrm{Sp}\,\Sigma$ begins to increase sharply near the optimum values of the regularization parameter, and this can be exploited in its choice.

The choice of the regularization parameter on the basis of the residual principle[14] is well known. In our treatment, this means choice of the value of α on the basis of the condition $\Phi(\hat{S})=N$. It can be seen from Fig. 6 that except for the version in Fig. 6a (spectral-window method) the residual principle leads to a value of α that lies to the right of the optimum value of α. The estimates $\hat{S}(x)$ corresponding to this principle for choosing the value of α are given in Fig. 7. It can be seen from these figures that, except for the estimate by the spectral-window method (Fig. 7a), the estimates $\hat{S}(x)$ are fairly close to the estimates $\hat{S}(x)$ for the optimum value of α.

The residual principle can be given the following interpretation. Except for the dual method, $\bar{Y}_n(\hat{S})=(G\tilde{G}^{-1}Y)_n\to Y_n$ for the methods as $\alpha\to 0$. Thus, fixing α on the basis of the condition $\Phi(\hat{S})=N$ is actually tantamount to the requirement that on the average the deviation from zero of $Y_n-\bar{Y}_n(\hat{S})$ as a result of the regularization be equal to the measurement error Y_n.

In the previously mentioned paper of Ref. 11, in which the solution was sought in the form of an expansion in B splines, a specific method of choosing the regularization parameter was proposed. In our approach, this method can be interpreted as the choice of the parameter on the basis of the condition

$$\sum_l f(\lambda_l,\alpha)\leqslant N_0,$$

where N_0 is the number of statistically significant coefficients in the spectral representation of the estimate.

Use of *a priori* information

It is sometimes known that a measured differential cross section is close to a known function $S_0(x)$. In this case, it is natural to seek an estimate in the form $\hat{S}(x)=S_0(x)+Z(x)$, where $Z(x)$ is an unknown function. Then the connection between the measurements of Y_n and $Z(x)$ can be expressed in the form

$$Y_n=Y_n^0+\int dx A_n(x)Z(x)+\varepsilon_n,$$

where $Y_n^0=\int dx A_n(x)S_0(x)$, and, more succinctly,

$$\tilde{Y}_n=\int dx A_n(x)Z(x)+\varepsilon_n, \qquad (62)$$

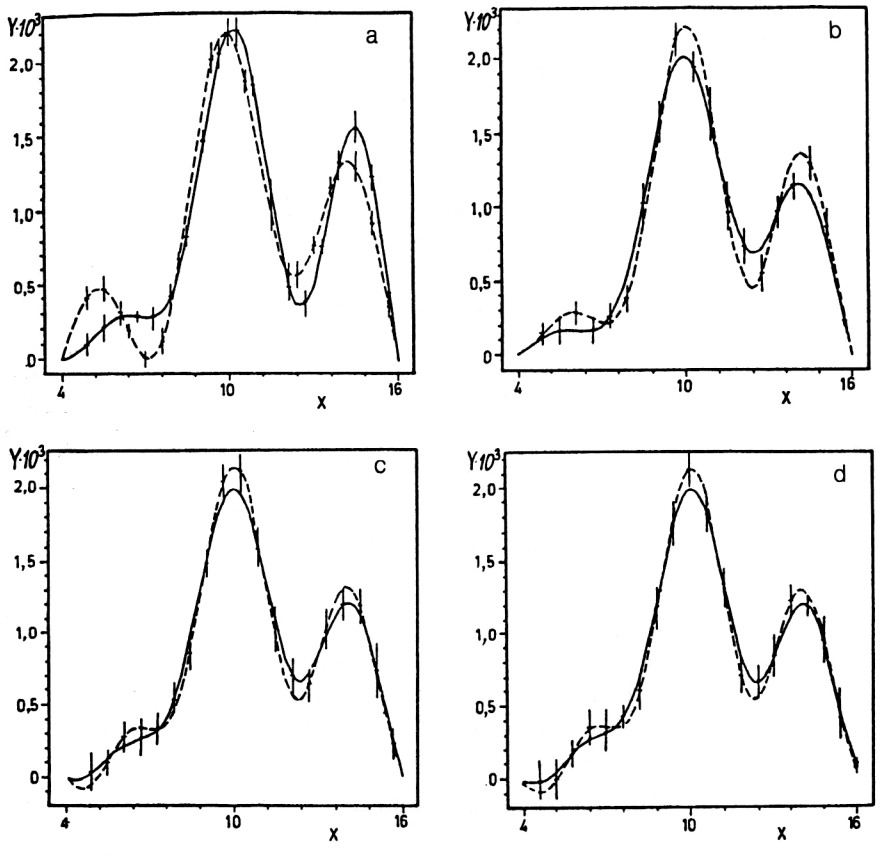

FIG. 7. Comparison of estimates obtained from the condition $\Phi(\hat{S})=N$ (continuous curve) with estimates that minimize $\overline{D_2}$ (broken curve). As an illustration we give at some points as the errors the values of $\Sigma(x,x)^{1/2}$. a) Spectral-window method, $\mathrm{Sp}\,A_R=6.00$, $\mathrm{Sp}\,\Sigma=2.56\cdot 10^5$; b) iteration method, $\mathrm{Sp}\,A_R=5.01$, $\mathrm{Sp}\,\Sigma=6.07\cdot 10^4$; c) Tikhonov's method, $\mathrm{Sp}\,A_R=5.46$, $\mathrm{Sp}\,\Sigma=9.40\cdot 10^4$; d) dual method, $\mathrm{Sp}\,A_R=5.43$, $\mathrm{Sp}\,\Sigma=1.02\cdot 10^5$.

where $\tilde{Y}_n=Y_n-Y_n^0$. The expression (62) is formally identical to (3). Therefore, using for $Z(x)$ the expression (14) with Y_n replaced by $\tilde{Y}_n$, we obtain for $\hat{S}(x)$ the result

$$\hat{S}^+(x)=S_0(x)+\sum_{n,m}A_n(x)G_{nm}^{-1}(Y_m-Y_m^0). \tag{63}$$

Averaging the expression (63), we obtain

$$\overline{\hat{S}^+}(x)=S_0(x)-\sum_{n,m}A_n(x)G_{nm}^{-1}\int dx'A_m(x)S_0(x)$$
$$+\sum_{n,m}A_n(x)G_{nm}^{-1}\int dx'A_m(x)S(x). \tag{64}$$

Recalling that the expression (15) is the projector onto the "measured" space, we see that $\hat{S}^+(x)$ is equal to $S_0(x)$ from which its projection onto the "measured" space has been removed and the projection of the true function $S(x)$ is inserted. As before, the noise component $\hat{S}^+(x)$ is determined by the expression (17) for $\Sigma^+(x,x')$. It is obvious that, as before, the one or other regularization method reduces to the replacement of the matrix G by the corresponding matrix $\tilde{G}$. Figure 8 gives estimates of the cross section (96) for $Q^2=50\ \mathrm{GeV}^2/c^2$ without and with the use of $S_0(x)=c/x$.

The use of other general *a priori* information $S(x)>0$ will be considered in the following section.

In conclusion, we emphasize the following:

1) when histograms are used as experimental information, the expressions $A_R(x,x')$ in the dual method and Tikhonov's method, and also $\overline{D^2}$, $\Sigma(x,x')$ in all methods become approximate on account of the need to use $K_{nm}\approx\delta_{nm}Y_n$;

2) all the methods that we have considered to obtain estimates reduce to regularization of the least-squares estimate (pseudoinversion) by replacement of the matrix G by the corresponding $\tilde{G}(\alpha)$;

3) in general, we do not completely achieve our aim, which is that of obtaining a model- and detector-independent estimate; the residual instrumental function is not equal to the δ function, it depends on the instrumental functions $A_n(x)$, which are determined by the properties of the detector, and, in addition, the functions $A_n(x)$ may be model-dependent.

Use of linear estimates $\hat{S}(x)$ in parametric analysis

When there exists a parametric model for $S(x)$: $S(x)=f(x|\mathbf{a})$, where $\mathbf{a}$ is a vector of parameters, there arises a problem of estimation and verification of the hypothesis that $f(x|\mathbf{a})$ does not contradict the experimental data Y_n.

This is often done in the following way. One first solves the inverse problem and finds the estimate of $S(x)$. One then goes over from this estimate to M independent statistics Z_m:

$$\hat{Z}_m=\int dx\varphi_m(x)\hat{S}(x),$$

where $\varphi_m(x)$ is a set of linearly independent functions, and $M\leqslant N$. For example, $\varphi_m(x)$ may be the characteristic func-

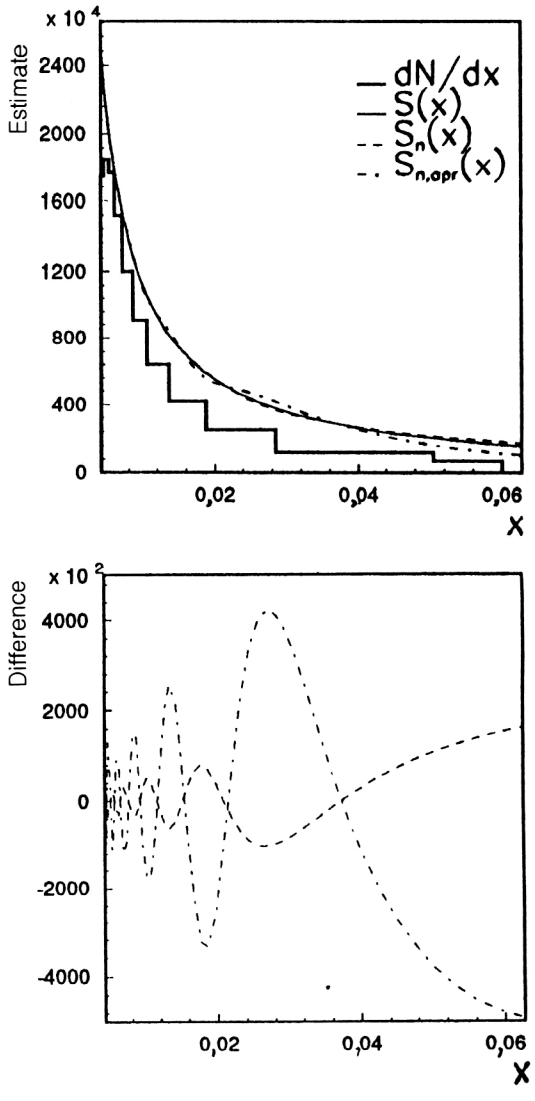

FIG. 8. Comparison of estimates of cross sections obtained using *a priori* information [broken curve, $S_0(x)=c/x$] and without it (chain curve). The corresponding differences of the cross sections and estimates are given in the lower part of the figure.

tion of the mth bin of the histogram or $\varphi_m(x)=\delta(x-x_m)$, where x_m are the coordinates of the center of the mth bin. The estimate of the parameters $\hat{a}$ is sought from the minimum of the functional

$$\Phi=\sum_{n,m}[\hat{Z}_n-\langle\varphi_n|f(x|\mathbf{a})\rangle]D(Z)^{-1}_{nm}\times[\hat{Z}_m-\langle\varphi_m|f(x|\mathbf{a})\rangle],$$

where $D(Z)$ is the error matrix of the statistics $\hat{Z}_m$. For definiteness, we shall assume that the estimate $\hat{S}(x)$ has the form

$$\hat{S}(x)=\sum_{n,m}A_n(x)\tilde{G}^{-1}_{nm}Y_m.$$

Substituting it in the expression for $\bar{\hat{Z}}_m$, we obtain

$$\bar{\hat{Z}}_m=\sum_{n,l}\langle\varphi_m|A_n\rangle\tilde{G}^{-1}_{nl}\langle A_l|S\rangle.$$

It can be seen from this that $\bar{\hat{Z}}_m$ differs from $Z_m=\langle\varphi_m|S\rangle$ even if $\varphi_m=A_m$ but $\tilde{G}^{-1}\neq G^{-1}$ (i.e., any of the regularization methods has been used). In general, this may lead to bias of the estimate of **a** even in the limit $N\to\infty$.

It seems to us that the most consistent approach in parametric analysis is to give up the use of $\hat{S}(x)$ and return to the direct measurements Y_n. This means that the estimate of the parameter vector **a** must be sought by the method of least squares (χ^2 method) from the minimum of the functional

$$\Phi=\sum_{n,m}(Y_n-\langle A_n|f(x|\mathbf{a})\rangle)K^{-1}_{nm}\times(Y_n-\langle A_m|f(x|\mathbf{a})\rangle).$$

2. NONLINEAR ESTIMATES

Nonlinear estimates are nonlinear functions of the measurements, and this leads to considerable difficulties in obtaining $\Sigma(x,x')$ and to the absence of $A_R(x,x')$. Therefore, in our opinion their use is justified when one uses the *a priori* information $S(x)>0$ when without it the linear estimates are unacceptably bad.

Correction-factor method

This method is very widely used at the present time, although there do not exist publications that justify it. Therefore, we shall rely on our own paper of Ref. 15.

There exist two forms of the correction-factor method. We begin with the one most widely used—the bin-in-bin correction method (see, for example, Ref. 16), which proceeds as follows: the Monte Carlo method is used to generate events whose distribution with respect to the variable x is described by the function $S_0(x)$; from it the histogram S^0_n with respect to the variable x at the entrance to the detector is constructed; the histogram $Y^{(0)}_n$ is then constructed after analysis of the response of the detector to these events; as corrected data one takes the histogram

$$\hat{S}^{(1)}_n=\frac{S^{(0)}_n}{Y^{(0)}_n}Y_n\equiv\frac{S^{(0)}_n}{\int dxA_n(x)S_0(x)}Y_n. \tag{65}$$

Figure 9 gives the histogram $\hat{S}^{(1)}_n$ and the histogram S_n obtained from $S(x)$ [$S(x)$ is taken from the example of the previous section, and $S_0(x)$ was obtained from $S(x)$ by a 10% change of the parameters of $S(x)$]:

$$S_n=\int dx\chi_n(x)S(x),$$

$$\chi_n(x)=\begin{cases}1/(x_{n+1}-x_n), & x\in[x_n,x_{n+1}]\\ 0, & x\notin[x_n,x_{n+1}].\end{cases}$$

It can be seen from Fig. 9 that $\hat{S}^{(1)}_n$ remained close to their measurements Y_n. The explanation for this can be seen in the following considerations.

We write the relation (3) in the form

$$Y_n=\int dxA_n(x)S_0(x)\frac{S(x)}{S_0(x)}+\varepsilon_n. \tag{66}$$

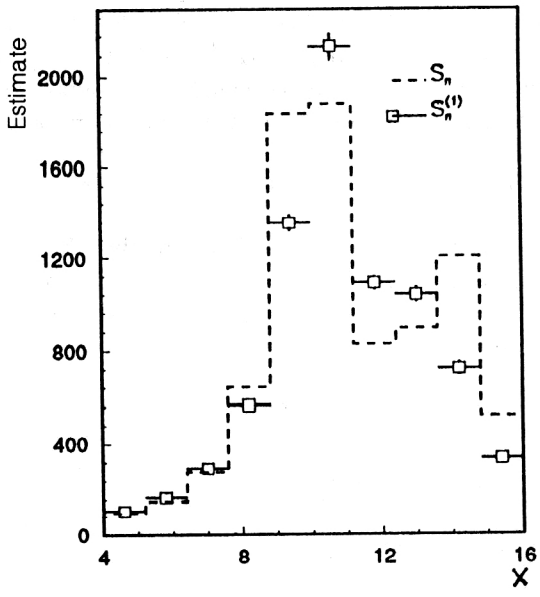

FIG. 9. The broken histogram is S_n obtained from the function $S(x)$; $\hat{S}_n^{(1)}$ is the estimate of S_n in the correction-factor method after the first iteration.

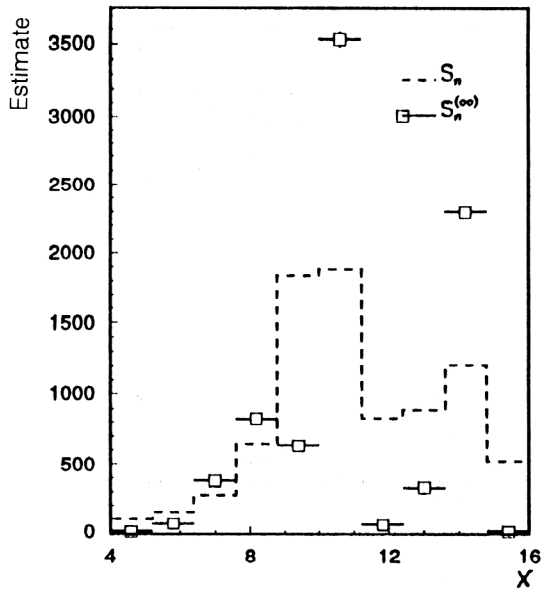

FIG. 10. The broken histogram is S_n obtained from the function $S(x)$; $\hat{S}_n^{(\infty)}$ is the estimate of S_n by the correction-factor method on stabilization of the solution.

If we have fulfillment of at least one of the two conditions (a) the functions $A_n(x)$ are fairly narrow and concentrated near the points $\bar{x}_n=(x_{n+1}+x_n)/2$ (good resolution, narrow bin) or (b) the function $S_0(x)$ is fairly close to the function $S(x)$, so that $S(x)/S_0(x)\approx$const in the neighborhood of $\bar{x}_n$, then we can write down the approximate equation

$$Y_n\approx\frac{S(\bar{x}_n)}{S_0(\bar{x}_n)}\int dxA_n(x)S_0(x)+\varepsilon_n$$

or

$$S(\bar{x}_n)\approx\frac{S_0(\bar{x}_n)}{\int dxA_n(x)S_0(x)}Y_n. \tag{67}$$

Multiplying (67) by $x_{n+1}-x_n$, we obtain an approximate analog of the expression (65). It can be seen from these considerations that a bad resolution and appreciable deviation of $S_0(x)$ from $S(x)$ have the consequence that $\hat{S}_n^{(1)}$ have not departed appreciably from the original measurements Y_n.

Another more complicated form of the correction-factor method takes the relation (65) as the first step of an iteration process

$$S_n^{(i)}=\frac{S_n^{(i-1)}}{\int dxA_n(x)S^{(i-1)}(x)}Y_n, \tag{68}$$

where $S^{(i-1)}(x)=\Sigma_n S_n^{(i-1)}\chi_n(x)$. Figure 10 shows $\hat{S}_n^{\infty}$ obtained at the end of the iteration process on the basis of the condition $\max|(S_n^{(i)}-S_n^{(i-1)})/S_n^{(i)}|<0.01$ (number of iterations 85). It can be seen from the figure that the instability of the solution of the problem with respect to the fluctuations of Y_n, which we mentioned in discussing the least-squares method, is also manifested in this approach. This can be understood from the following considerations.

We write the denominator of the relation (68) in the form

$$\int dxA_n(x)\sum_m S_m^{(i-1)}\chi_m(x)=\sum_m A_{nm}S_m^{(i-1)},$$

where $A_{nm}=\langle A_n|\chi_m\rangle$. The condition for stabilization of the iteration process means that

$$\frac{Y_n}{\Sigma_m A_{nm}S_m^{(i-1)}}\approx 1,$$

i.e., on stabilization, we have the approximate equation[6)]

$$\sum_m A_{nm}S_m^{(i-1)}\approx Y_n. \tag{69}$$

However, the matrix A_{nm}, like the Gram matrix G_{nm}, may have small eigenvalues, and this leads to an unacceptably large noise component of $\hat{S}_n^{(\infty)}$.

The picture is radically changed if Y_n is replaced by $\bar{Y}_n$, i.e., in the absence of fluctuations of Y_n (Fig. 11). The histogram $\hat{S}_n^{+}=\int dx\hat{S}^{+}(x)\chi_n(x)$ obtained from the least-squares solution is much closer to S_n. This reflects the fact that $\hat{S}_n^{+}$ was sought in the form of a linear combination of the functions $A_n(x)$, which form a basis of the "measured" space, while $\hat{S}^{(\infty)}(x)$ was sought in the form of a linear combination of $\chi_n(x)$. The difference of $\hat{S}_n^{(\infty)}$ and $\hat{S}_n^{+}$ from S_n characterizes the minimum systematic error of the least-squares method and correction-factor method.

Speaking generally, the correction-factor method is thus a regularization method whose parameter is the number of iterations. If the number of iterations is chosen to minimize $D^2=\Sigma_n(S_n^{(i)}-S_n)^2$, then we obtain the $\hat{S}_n^{(i)}$ shown in Fig. 12 (four iterations). We also give there $\hat{S}_n^{T}=\int dx\hat{S}^{T}(x)\chi_n(x)$, where $\hat{S}^{T}(x)$ is the estimate in Tikhonov's method for a minimum of $\overline{D^2}=\int dx[S^{T}(x)-S(x)]^2$. In this case, we obtain sensible es-

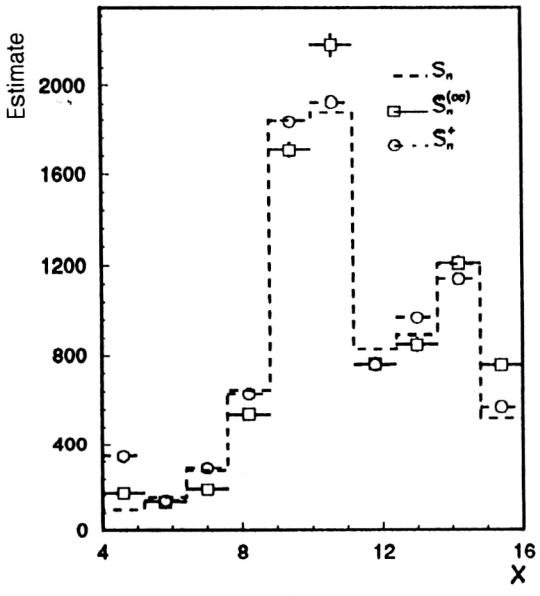

FIG. 11. Comparison of estimates by the correction-factor and least-squares methods in the absence of errors in the measurements. The broken histogram is S_n obtained from the function $S(x)$; $\hat{S}_n^{(\infty)}$ is the correction-factor estimate on stabilization of the solution; $\hat{S}_n^+$ is the least-squares estimate.

timates $\hat{S}_n^{(i)}$, but both estimates $\hat{S}_n^{(i)}$ and $\hat{S}_n^T$ acquire an additional bias on account of the regularization. Because of the nonlinear procedure for obtaining the estimates, we can neither calculate their statistical errors nor give a characteristic of the systematic error in analytic form. We merely note that in the correction-factor method one gives as statistical errors the errors obtained in the first iteration,

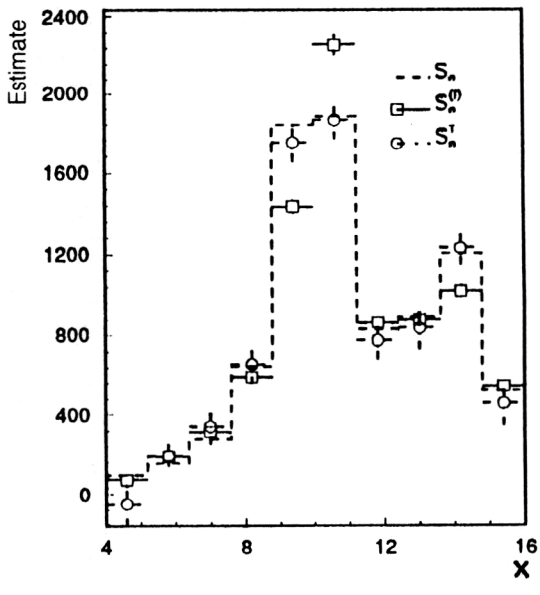

FIG. 12. Comparison of the estimates by the correction-factor method and Tikhonov's method from the condition of a minimum of $\overline{D^2}$ in the presence of errors in the measurements. The broken histogram is S_n obtained from the function $S(x)$; $\hat{S}_n^{(i)}$ is the estimate in the correction-factor method; $\hat{S}_n^T$ is the estimate in Tikhonov's method.

which we show in the figures. As can be seen from comparison of Figs. 9 and 10, these will be clear underestimates, since with increasing number of iterations they increase. Note that the values of $\hat{S}_n^{(i)}$ become correlated from the second iteration. At the same time, for $\hat{S}_n^T$ we can calculate the error matrix Σ and the residual instrumental function $A_{Rn}(x)$:

$$\Sigma = A^T \tilde{G}^{-1} K \tilde{G}^{-1} A,$$

$$A_{Rn}(x) = \sum_{kl} A_{nk}^T \tilde{G}_{kl}^{-1} A_l(x) \neq \chi_n(x).$$

In the analysis of real data, D^2 is, naturally, unknown, and to choose the optimum value of the regularization parameter—the number of iterations—it is necessary to use other criteria, not a minimum of D^2. In the previous section, we considered the problem of the choice of the optimum value of the regularization parameter for linear estimates on the basis of an analysis of Sp Σ, Sp A_R, χ^2. In the case of the correction-factor method, χ^2 remains available for analysis. In accordance with the residual principle,[14] the regularization parameter is determined from the condition $\chi^2 = N$. For Fig. 12, $\chi^2 = 22.9$. It can be seen from it that for the optimum number of iterations the value of χ^2 can deviate appreciably from N.

In the NA4 experiment (see, for example, Ref. 18), an iteration process based on the relation (67) was used to estimate the structure function $F_2(x,Q^2)$. However, in contrast to the previous treatment [the relation (68)], it was written in the form

$$S^{(i)}(\bar{x}_n) = \frac{S^{(i-1)}(\bar{x}_n)}{\int dx A_n(x) S^{(i-1)}(x|\mathbf{a})} Y_n, \tag{70}$$

where $S^{(i-1)}(x|\mathbf{a})$ is a parametric function whose parameters are determined from the fit of $S^{(i-1)}(x|\mathbf{a})$ to $S^{(i-1)}(\bar{x}_n)$. In this case, the stabilization condition leads to the approximate equation

$$\int dx A_n(x) S^{(i-1)}(x|\mathbf{a}) \approx Y_n. \tag{71}$$

In Fig. 13, we give $\hat{S}^{(\infty)}(\bar{x}_n)$ for such an approach (20 iterations).

Note, however, that in place of the parametric function we have used the spline approximation $\hat{S}^{(i-1)}(\bar{x}_n)$. It can be seen from Fig. 13 that in this approach the noise component $\hat{S}^{(\infty)}(\bar{x}_n)$ is unacceptably large. However, comparison of it with Fig. 10 shows that the noise component is smaller than in the step approximation of $\hat{S}_n^{(\infty)}$. It appears that the algorithm (70) is more stable to the fluctuations of Y_n. If in place of Y_n we use $\bar{Y}_n$, the result is better than in Fig. 10. However, the systematic error is still quite large.

Figure 14 gives the estimate $\hat{S}^{(i)}(\bar{x}_n)$ $(i=3,\ \chi^2=6.5)$ obtained from the condition of a minimum of $D^2 = \Sigma_n (S(\bar{x}_n) - S^{(i)}(\bar{x}_n))^2$ and the values of $\hat{S}^T(\bar{x}_n)$ obtained by Tikhonov's method.

Thus, for the solution by the correction-factor method one can calculate the noise characteristic after one iteration, but the solution contains a systematic error, which depends on the function $S_0(x)$ (model dependence).

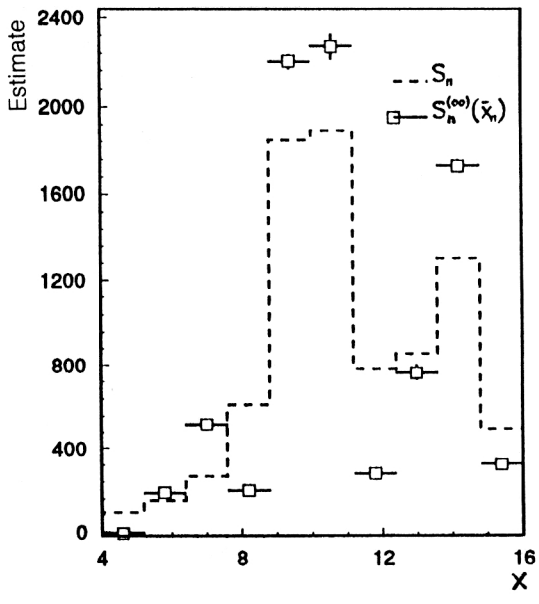

FIG. 13. The broken histogram is S_n obtained from the function $S(x)$; $\hat{S}_n^{(\infty)}(\bar{x}_n)$ is the estimate in the correction-factor method [Eq. (70)] under the condition of stabilization.

With regard to the solution of the correction-factor method in accordance with the iteration scheme, we can draw the following conclusions:

the correction-factor method reduces to the solutions of the systems of equations (69) and (71) in the absence of noise in Y_n;

the exit from the iteration process using the stabilization criterion can lead to an unacceptably large statistical error;

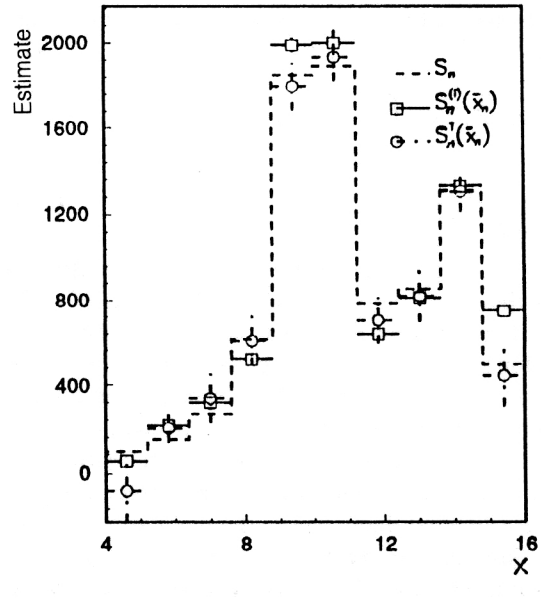

FIG. 14. Comparison of estimates in the correction-factor method [Eq. (70)] and in Tikhonov's method. The broken histogram is S_n obtained from the function $S(x)$; $\hat{S}_n^{(i)}(\bar{x}_n)$ is the estimate in the correction-factor method, and $\hat{S}_n^T(\bar{x}_n)$ is the estimate in Tikhonov's method.

by choosing a smaller than optimal number of iterations, one can make the statistical error acceptable;

it is difficult to interpret the estimate based on the correction-factor method, since, the procedure being nonlinear, it is impossible to calculate the characteristics of the noise and the systematics in the estimate;

the estimate by Tikhonov's method was found to be better, and for it one can calculate the characteristics of the noise and the systematics.

Examples of the use of other nonlinear iteration schemes can be found in Refs. 9, 19, and 20. However, it appears to us that, without the use of *a priori* information, such nonlinear estimates have only one advantage, the simple computational algorithm, but at the same time one loses the possibility of a statistical interpretation of the estimates.

It seems to us better to use nonlinear estimates that are obtained in conjunction with *a priori* information on the positivity of $S(x)$.

Estimates satisfying the property $\hat{S}(x)>0$

In Ref. 21, the functional (35) of Tikhonov's method was algebrized for $k=0$. This was done by using some numerical quadrature formula, and the integrals were replaced by sums and the function $S(x)$ by the vector S of its values at the integration point $S_i=S(x_i)$, $i=1,...,I$. Then

$$\Phi=\sum_{n,m}\left(Y_n-\sum_{i=1}^{I}A_{ni}S_i\right)W_{nm}\times\left(Y_m-\sum_{j=1}^{I}A_{mj}S_j\right)+\alpha\sum_i S_i^2. \tag{72}$$

As an estimate $\hat{S}_i$, one takes the vector S_i that minimizes (72) under the restrictions $S_i>0$. Numerically, the problem is solved by the iteration method of Ref. 22. In Ref. 23, an estimate $\hat{S}_i>0$ is sought from the functional (72), in which $\Sigma_i S_i^2$ is replaced by $\Sigma_i(S_{i+2}-2S_{i+1}+S_i)^2$, which corresponds to Tikhonov's method for $k=2$. The details of this approach can be found in Refs. 24 and 25. Unfortunately, the authors analyze neither the systematic nor the statistical errors that arise in such a method of estimating $S(x)$. All that is obvious is the fact that the transition from $S(x)$ to $S(x_i)$ with interpolation between the points x_i during the integration already leads to a systematic error. The desire to reduce this error leads to a large value of I and, therefore, to a high dimension of the problem. Note that the dimension of the problems of obtaining linear estimates is always equal to the number of measurements Y_n.

A more advanced method in the sense of the interpretation of the estimate is the approach proposed in Ref. 3. In it, the estimate $S(x)$ is sought in the form $S(x)=\exp\{Z(x)\}$, and this automatically ensures that it is positive. The problem is then also algebrized by replacing $Z(x)$ by $Z_i=Z(x_i)$. In the calculation of the integrals $\int dx A_n(x)S(x)$, a linear interpolation $Z(x)$ between the points x_i was made. To estimate Z_i, Tikhonov's method for $k=2$ was used. The functional, which is nonlinear in this case, was minimized by means of the program FUMILI.[26]

Besides the minimization, the program estimates the error matrix of the parameters Z_i, and this, in its turn, makes it possible to calculate an approximate matrix of estimates of $S(x_i)$. Moreover, the authors obtained an expression A_R for the "quasiresidual" instrumental matrix for Z_i, the deviation of which from unity characterizes the systematic errors in the estimates of Z_i:

$$(A_R)_{kj}=\sum_{i,m} Z_{ki}^{-1}\frac{\partial f_n}{\partial Z_i}W_n\frac{\partial f_n}{\partial x_j}, \tag{73}$$

where

$$Z_{ki}=\sum_n \frac{\partial f_n}{\partial Z_k}W_n\frac{\partial f_n}{\partial Z_i}+\alpha\sum_l \frac{\partial g_l}{\partial x_k}\frac{\partial g_l}{\partial x_i},$$

$$f_n=\int dx A_n(x)\exp\{Z(x)\},$$

$$g_l(Z)=Z_{l+2}-2Z_{l+1}+Z_l,$$

in which $g_l(Z)$ is the finite-difference second derivative. The error corridor of the function $S(x)$ is given approximately by

$$S^{\pm}(x_i)=\exp\{\hat{Z}_i\pm(\Sigma_{ii}^Z)^{1/2}\},$$

where

$$\Sigma_{km}^Z=\sum_{i,j,n} Z_{ki}^{-1}\frac{\partial f_n}{\partial Z_i}W_n\frac{\partial f_n}{\partial Z_j}Z_{jm}^{-1}.$$

The numerical examples given in Ref. 3 show that the expressions for A_R and Σ^Z calculated in the last iteration are fairly accurate. In the same paper a study was made of the systematic and statistical errors of functionals of $S(x)$:

$$D=\int dx D(x)S(x).$$

The program written on the basis of this approach has been used for many years at the Institute of High Energy Physics to estimate the spectra of scattered neutrons by means of Bonner spheres. In this case, $A_n(x)$, $n=1,\ldots,N=5$–7, are the response functions of N independent detectors. The small number of measurements and the very "broad" functions $A_n(x)$ make the use of the information $S(x)>0$ simply essential here. This is illustrated in Fig. 15, which gives the estimate $\hat{S}(x)$ by Tikhonov's method with and without the use of this information.

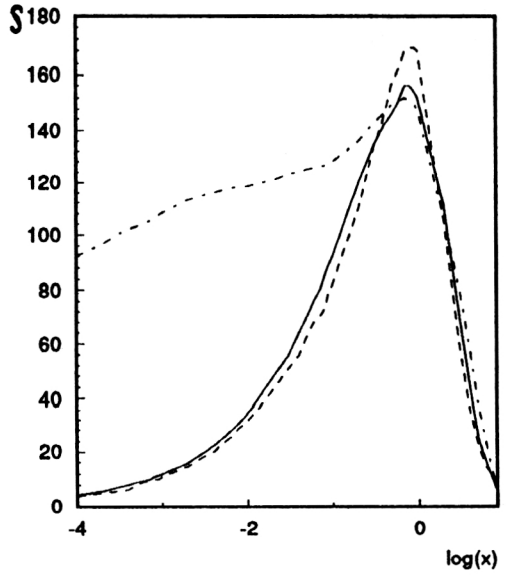

FIG. 15. Recovery of the spectrum of ^{252}Cf. The continuous curve is the true spectrum, the chain curve is obtained by the linear method of estimation, and the broken curve by the nonlinear method.

Maximum-entropy principle

This principle for obtaining an estimate of $S(x)$ enjoys quite wide popularity (Refs. 27–31 and 40). The fact is that this principle ensures not only positivity of the estimate $\hat{S}(x)$ but also has an information statistics justification. Whereas, for example, Tikhonov's method proposes that among all functions $S(x)=Cp(x)$ that satisfy the given experimental restrictions one choose the solution with the minimum norm $\|S\|$, this principle proposes that one should choose the one that has maximum entropy H, which is equal to

$$H(p)=-\int dx \ln p(x)p(x). \tag{74}$$

Classical examples of the use of the maximum-entropy principle demonstrate that the choice of (74) as a measure of the uncertainty of $p(x)$ is sensible.

1. The distribution that maximizes the entropy on the interval $[a,b]$ is the uniform distribution $p(x)=1/(b-a)$.

2. On the axis $(-\infty,\infty)$ the function that maximizes (74) and has $\bar{x}=\mu$, $\overline{(x-\mu)^2}=\sigma^2$ is the Gaussian distribution with these parameters μ and σ.

In Ref. 27, the case when nothing is known about the distribution ε_n was considered. In our exposition, we shall follow Ref. 28.

To solve our problem, it is natural to seek the estimate of $p(x)$ on the basis of the condition of an extremum of the functional

$$\Phi=\sum_n w_n\left[Y_n-C\int dx A_n(x)p(x)\right]^2-\alpha H(p)$$
$$+\lambda\left[\int dx p(x)-1\right], \tag{75}$$

where $C=1/\int dx S(x)$, and λ is the Lagrangian multiplier of the term that takes into account the fact that $p(x)$ is normalized. The function $\hat{p}(x)$ that extremizes the functional (75) for given α must, first, be compatible with the experimental data Y_i and satisfy the normalization condition and, second, must have maximum entropy, i.e., must be the least informative estimate. Then one reduces the risk of including in $\hat{p}(x)$ information that is not needed to describe the experimental data. This is the main attraction of using the maximum-entropy principle for regularization.

In addition, the use of $H(p)$ as a stabilizer leads, as we shall see, to positivity of $\hat{p}(x)$, i.e., to automatic allowance for the only accessible *a priori* information of a general nature.

We first consider the case when C is known, i.e., $E(x)=1$ and $1/C=\Sigma_n Y_n$. Variation of the functional (75) with respect to $p(x)$ gives the following equation for $p(x)$:

$$\frac{\delta\Phi}{\delta p(x)}=\alpha\ln p(x)+\alpha-2\sum_n w_n\left[Y_n-C\times\int dx' A_n(x')p(x')\right]CA_n(x)+\lambda=0. \tag{76}$$

Introducing the notation

$$a_n=\frac{2Cw_n}{\alpha}\left[Y_n-C\int dx A_n(x)p(x)\right],\qquad \gamma=-1-\lambda/\alpha, \tag{77}$$

we can write a formal solution of Eq. (76) in the form

$$\hat{p}(x)=\exp\left\{\gamma+\sum_i a_i A_i(x)\right\}. \tag{78}$$

It can be seen from the expression (78) that positivity of the estimate $\hat{p}(x)$ is ensured by its functional form.

Substitution of (78) in the equation

$$\frac{\partial\Phi}{\partial\lambda}=\int dx p(x)-1=0 \tag{79}$$

enables us to eliminate γ and write $\hat{p}(x)$ in the form

$$\hat{p}(x|a)=\exp\left\{\sum_n a_n A_n(x)\right\}/Z(a), \tag{80}$$

where

$$Z(a)=\int dx\exp\left\{\sum_n a_n A_n(x)\right\}. \tag{81}$$

The equations for the parameters a_n are obtained by substituting (80) in (77):

$$F_n\equiv Y_n-CC_n(a)-\frac{\alpha}{2Cw_n}a_n=0, \tag{82}$$

where

$$C_n(a)=\int dx A_n(x)\hat{p}(x|a).$$

The system (82) is nonlinear with respect to the parameters a_n. We shall describe our method of solution below.

Now suppose that $E(x)\leqslant 1$, i.e., that the parameter C is unknown. Differentiation of the functional (75) with respect to C gives an additional equation, and, combining it with the system (82), we arrive at a system of $N+1$ equations for the parameters a_n, C:

$$F_n\equiv Y_n-CC_n(a)-\frac{\alpha}{2Cw_n}a_n=0,\quad n=1,\ldots,N,$$
$$F_{N+1}\equiv\sum_n w_n C_n(a)[Y_n-CC_n(a)]=0. \tag{83}$$

Figure 16 shows the result of a numerical experiment on the recovery of $S(x)=Cp(x)$ from the first section.

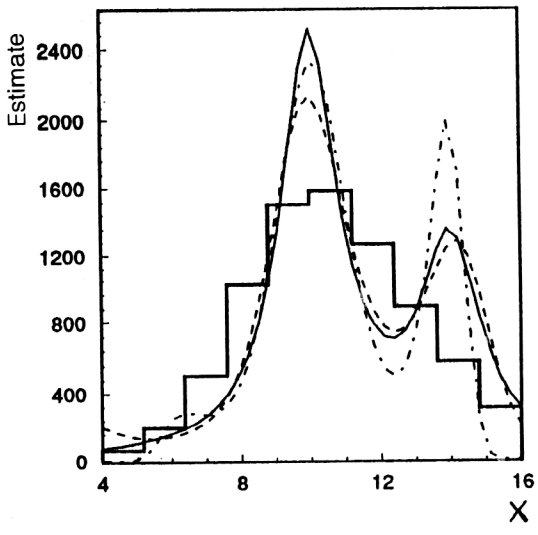

FIG. 16. Reconstruction of $S(x)=Cp(x)$ using the maximum-entropy principle for the example of the first section. The broken curve is for known C, and the chain curve is for C as a parameter.

It should be noted that the inclusion of C among the unknown parameters makes the convergence of the iteration process poorer. This is a consequence of the fact that the last equation of the system (83) for $\alpha=0$ can be obtained from the first N equations. Moreover, for unknown C the range of values of α in which a significant improvement of the resolution can be achieved is appreciably smaller than the region for known C.[28] The solution of the systems (82) and (83) was determined by minimizing $\Sigma_n(F_n)^2$ by means of the program FUMILI[26] and was achieved for known C after ~ 10 iterations but for known C after ~ 25.

It can be seen from the expression (80) that, generally speaking, the estimate $\hat{p}(x)$ is biased and the difference $p(x)-\bar{\hat{p}}(x)$ is determined by the possibility of describing $\ln p(x)$ by a linear combination of $A_n(x)$. In addition, the terms $\alpha a_n/2Cw_n$ in the system (83) lead to an additional bias.

We now consider the noise component of the estimate $\hat{p}(x)$. It is determined by the error matrix of the parameters a, which can be obtained approximately as follows. Let $\tilde{a}_j$ be the solution of the system (83) for $\varepsilon_n=0$, and $a_j=\tilde{a}_j+\Delta a_j$ be the solution for $\varepsilon_n\neq 0$, i.e., for $Y_n=\bar{Y}_n+\varepsilon_n$. We write the system (83) in the form

$$F_i(\tilde{a}_j+\Delta a_j, Y_n=\bar{Y}_n+\varepsilon_n)=0,\qquad j,i=1,\ldots,N+1,\quad n=1,\ldots,N, \tag{84}$$

and expand F_i with respect to Δa_j and ε_n, restricting ourselves to the terms linear in Δa_j, ε_n ($a_{N+1}\equiv C$). We have

$$F_i(\tilde{a}_j,\bar{Y}_n)+\sum_j\frac{\partial F_i}{\partial a_j}\Delta a_j+\sum_n\frac{\partial F_i}{\partial Y_n}\varepsilon_n=0.$$

By definition, $F_i(\tilde{a}_j,\bar{Y}_n)=0$, and we can therefore write

$$\Delta a=-B^{-1}D\varepsilon,\quad B_{ij}=\frac{\partial F_i}{\partial a_j},\quad D_{in}=\frac{\partial F_i}{\partial Y_n},$$

$$B_{ij}=-CC_{ij}-\frac{\alpha}{2Cw_i}\delta_{ij},\quad i,j=1,...,N,$$

$$B_{N+1,j}=\sum_i w_iY_iC_{ij}-2C\sum_i w_iC_iC_{ij},$$

$$B_{i,N+1}=-C_i+\frac{\alpha}{2C^2w_i}C_i,\quad B_{N+j,N+1}=-\sum_i w_iC_i^2, \tag{85}$$

$$D_{ij}=\delta_{ij},\quad i,j=1,...,N,\quad D_{N+1,i}=w_iC_i,$$

$$C_{ij}=\frac{\partial C_i(a)}{\partial a_j}.$$

From (85) we find that the error matrix a_j is

$$\Sigma^a=B^{-1}D\sigma^2D^TB^{T-1},\quad \sigma^2_{nm}=\overline{\varepsilon_n\varepsilon_m}=Y_n\delta_{nm}. \tag{86}$$

The matrix Σ^a_{ij}, $i,j=1,...,N$ is the error matrix of the parameters a_i, $i=1,...,N$, in terms of which we obtain approximately the error operator $\hat{p}(x)$:

$$\Sigma^p(x,x')=\sum_{i,j=1}^{N}\frac{\partial\hat{p}(x|\hat{a})}{\partial a_i}\sum_{ij}^{a}\frac{\partial\hat{p}(x'|\hat{a})}{\partial a_j}. \tag{87}$$

There exists a class of problems in which it is known that the unknown distribution $p(x)$ must be close to some given $p_0(x)$. In this case, it is natural to use the Kullback–Leibler distance (Refs. 32 and 33).[7)]

$$\Omega(p)\equiv I(p,p_0)=\int dxp(x)\ln[p(x)/p_0(x)]\geqslant 0,$$

which reaches a minimum when $p(x)=p_0(x)$. When $p_0(x)$ is a uniform distribution, $p_0(x)=1/L$, where L is the length of the interval of measurement of x in the experiment, then

$$I(p,p_0)=\int dxp(x)\ln p(x)-\int dxp(x)(1/L)$$
$$=-H(p)-1/L,$$

i.e., the maximum-entropy principle is equivalent to minimization of $I(p,p_0)$ when $p_0(x)$ is a uniform distribution. Thus, the principle of a minimum of the Kullback–Leibler distance generalizes the maximum-entropy principle for a given *a priori* distribution p_0, becoming identical to it when p_0 is a uniform distribution.

Thus, we shall seek an estimate of $p(x)$ from the condition of an extremum of the functional

$$\Phi=\sum_n w_n\left[Y_n-C\int dxA_n(x)p(x)\right]^2$$
$$+\alpha I(p,p_0)+\lambda\left[\int dxp(x)-1\right].$$

Varying the functional Φ with respect to $p(x)$ and using the normalization condition, we can obtain

$$\hat{p}(x|a)=p_0(x)\exp\left[\sum_n a_nA_n(x)\right]/Z(a), \tag{88}$$

where

$$Z(p)=\int dxp_0(x)\exp\left[\sum_n a_nA_n(x)\right],$$

$$a_n=\frac{2Cw_n}{\alpha}\left[Y_n-C\int dxp(x)A_n(x)\right].$$

Substitution of the expression for $\hat{p}(x|a)$ in the last equation leads to the system of equations

$$F_n\equiv Y_n-CC_n(a)-\frac{\alpha}{2Cw_n}a_n=0, \tag{89}$$

where

$$C_n(a)=\int dxA_n(x)\hat{p}(x|a).$$

The system (89) is formally identical to the system (82), like the equation for $\partial F/\partial C$. The only difference is the appearance of the function $p_0(x)$ in the integrand in the definition of $C_n(a)$:

$$C_n(a)=\int dxA_n(x)p_0(x)\exp\left[\sum_n a_nA_n(x)\right]/Z(a).$$

It should be noted that a distribution of the form (88) is obtained when $I(p,p_0)$ is minimized subject to the condition that

$$\int dxA_n(x)p(x)=Y_n,$$

and it is called the conjugate distribution.[34]

In Ref. 31, the estimate $S(x)$ is sought from the minimum of the functional (75) for known C, and the function $p(x)$ is approximated by a step function; however, the systematic and statistical errors of the method are not analyzed.

In Ref. 29, it is proposed to use the maximum-entropy principle to correct an experimental distribution with respect to multiplicity $0(m)$ equal to

$$0(m)=\sum_{n=1}^{N}P_{mn}T(n),\quad m=1,...,M, \tag{90}$$

where $T(n)$ is the studied distribution, and P_{mn} is a matrix that describes the distortions of the detector. However, instead of minimization of the discrete analog of the functional (75) it is proposed in the paper to maximize the entropy of the distribution $T(n)$ under the condition that the moments

$$\langle m^q\rangle=\sum_{n=1}^{M}m^q0(m)=\sum_{n=1}^{N}\left(\sum_{m=1}^{M}m^qP_{mn}\right)T(n) \tag{91}$$

are equal to their experimental values. The problem is solved by the method of Lagrangian multipliers. The number of employed constraints is increased until χ^2 reaches the value M, and the variable criterion of the series is $M/2$. The decision not to use all the experimental information $0(m)$ but only some moments is justified by the author by the simplicity of the problem compared with the χ^2 minimization. However, this approach also leads to the problem of solving nonlinear equations. There is no analysis of

the systematic and statistical errors. The numerical experiments made by the author gave satisfactory results.

In Ref. 30, the estimate of the distribution is sought from the condition of a maximum of the entropy of this distribution under the condition that χ^2 be equal to the number of experimental data. The problem is solved by the method of Lagrangian multipliers, and this actually corresponds to choosing the regularization parameter α in the functional (75) in accordance with the residual principle.[14] This does not guarantee an optimum value of α or, therefore, an optimal estimate $\hat{S}(x)$.

3. ESTIMATES OF STRUCTURE FUNCTIONS USING DATA ON DEEP INELASTIC LEPTON–NUCLEON SCATTERING

One of the main problems of experiments on deep inelastic lepton–nucleon scattering is the model-independent extraction of the nucleon structure functions. These are used to obtain the parton distributions in the framework of the QCD parton model.

Muon and electron scattering in the neutral-current channel

Let Y_n be the numbers of events in the bins of the histogram of the measured cross section $d^2\sigma/dxdQ^2$ for given $s=(p_{\mu(e)}+p_N)^2$. For brevity, we write $d^2\sigma/dxdQ^2 \equiv S(x,Q^2|s)$; $E(x,Q^2)$ is the probability of detecting an event with given x, Q^2 (the detector acceptance); $K(x',Q^{2'}|x,Q^2,s)$ is the probability of obtaining as a result of analysis of an event the measured values x', $Q^{2'}$ instead of the true x, Q^2 (it is the function that describes the resolution of the experiment and also the radiative emission of a lepton); ε_n is the fluctuation of the number of events Y_n; $\bar{Y}_n$ is the expected number of events. Then Y_n can be written in the form

$$Y_n=\bar{Y}_n+\varepsilon_n=L\int_{S_n}dx'dQ^{2'}\int dxdQ^2K(x',Q^{2'}|x,Q^2,s)\times E(x,Q^2)S(x,Q^2)|s)+\varepsilon_n, \tag{92}$$

where L is a number determined either by the number of leptons directed onto the target in the experiment and the number of nucleons in the target (extracted-beam experiment) or by the luminosity of the collider and the time of statistics collection (colliding-beam experiment), and S_n is the area of the nth bin of the experimental histogram with respect to x, Q^2. For what follows, it is convenient to write the expression (92) in the form

$$Y_n=\int dxdQ^2A_n(x,Q^2,s)S(x,Q^2|s)+\varepsilon_n, \tag{93}$$

where

$$A_n(x,Q^2,s)=L\int_{S_n}dx'dQ^{2'}K(x',Q^{2'}|x,Q^2,s)E(x,Q^2).$$

If we make no model assumptions about the structure functions, then

$$S^{\pm}_{NC}(x,Q^2|s)=f_L(y,Q^2)F_L(x,Q^2)+f_2(y,Q^2)F_2(x,Q^2)\mp f_3(y,Q^2)F_3(x,Q^2), \tag{94}$$

where

$$y=Q^2/(s-M^2)x,\quad f_L(y,Q^2)=-4\pi\alpha^2y^2/xQ^4,$$
$$f_2(y,Q^2)=4\pi\alpha^2(1-y+y^2/2)/xQ^4,$$
$$f_3(y,Q^2)=4\pi\alpha^2(y-y^2/2)/Q^4.$$

The problem simplifies if we assume $F_L=0$ (the Callan–Gross relation holds). The cross sections take the form

$$S^{\pm}_{NC}(x,Q^2|s)=f_2(y,Q^2)F_2(x,Q^2)\mp f_3(y,Q^2)F_3(x,Q^2). \tag{95}$$

Another approximation $S^{\pm}_{NC}(x,Q^2|s)$ is obtained by ignoring the parity-violating terms:

$$S^{\pm}_{NC}(x,Q^2|s)=f_L(y,Q^2)F_L(x,Q^2)+f_2(y,Q^2)F_2(x,Q^2). \tag{96}$$

Finally, a very simple expression for $S^{\pm}_{NC}(x,Q^2|s)$ is obtained by using both approximations:

$$S^{\pm}_{NC}(x,Q^2|s)=f_2(y,Q^2)F_2(x,Q^2). \tag{97}$$

At the present time, the experimental resolution and radiative corrections are taken into account using the correction-factor method. However, as we have seen, this is not the best approach to the problem. We now describe the approach presented in Ref. 35.

Estimation of the structure functions using estimates of the cross sections

We now base the estimation of the structure functions on equality of the cross sections in one of the models (94), (95), (96), (97) to their estimates in accordance with one of the methods of the first section.

Model (97). In this case, the model contains one unknown structure function $F_2(x,Q^2)$, and therefore its estimate is

$$\hat{F}_2(x,Q^2)=\hat{S}^{\pm}_{NC}/f_2(x,Q^2,s)=\sum_{n,m}A_n(x,Q^2,s)\tilde{G}^{-1}_{nm}Y^{\pm}_m/f_2(x,Q^2,s). \tag{98}$$

Using the fact that in the framework of this model

$$\bar{Y}^{\pm}_n=\int dxdQ^2A_n(x,Q^2,s)f_2(x,Q^2,s)F_2(x,Q^2),$$

we find that

$$\bar{\hat{F}}^2(x,Q^2)=\int dx'dQ^{2'}A_R(x,Q^2,x',Q^{2'},s)F_2(x',Q^{2'}),$$

$$A_R(x,Q^2,x',Q^{2'},s)$$

$$= \sum_{n,m} f_2^{-1}(x,Q^2,s)A_n(x,Q^2,s)$$

$$\times \tilde{G}_{nm}^{-1}A_m(x',Q^{2'},s)f_2(x',Q^{2'},s),$$

$$\mathrm{Sp}\, A_R = \mathrm{Sp}\, G\tilde{G}^{-1}. \tag{99}$$

The noise component of the estimate (98) is characterized by the error operator

$$\Sigma(x,Q^2,x',Q^{2'},s)$$

$$= \sum_{n,m,k,l} f_2^{-1}(x,Q^2,s)A_n(x,Q^2,s)\tilde{G}_{nm}^{-1}K_{mk}^{\pm}$$

$$\times \tilde{G}_{kl}^{-1}A_l(x',Q^{2'},s)f_2^{-1}(x',Q^{2'},s),$$

where $K^{\pm}$ is the error matrix of the measurements $Y_n^{\pm}$, and as a whole[8)]

$$\overline{\|F_{2\varepsilon}\|^2} = \mathrm{Sp}\, \tilde{G}^{-1}K\tilde{G}^{-1}\langle f_2^{-1}A | f_2^{-1}A\rangle,$$

$$F_{2\varepsilon} = \sum_{n,m} A_n(x,Q^2,s)\tilde{G}_{nm}^{-1}\varepsilon_m^{\pm}/f_2(x,Q^2,s),$$

where the matrix $\langle f_2^{-1}A | f_2^{-1}A\rangle$ is $\langle f_2^{-1}A | f_2^{-1}A\rangle_{nm} = \int dx dQ^2 f_2^{-2}A_nA_m$.

Model (96). In this case, the cross section contains two unknown structure functions $F_L(x,Q^2)$ and $F_2(x,Q^2)$, and it is required to measure, for example, the cross section $S_{NC}^{-}(x,Q^2|s)$ for two values of s: s_1 and s_2. Then after estimates of $S_{NC}^{-}(x,Q^2|s_i)$ have been obtained, we have the system of equations

$$\left.\begin{aligned} &f_L(x,Q^2,s_1)\hat{F}_L(x,Q^2) + f_2(x,Q^2,s_1)\hat{F}_2(x,Q^2) \\ &\quad = \hat{S}_{NC}^{-}(x,Q^2|s_1) \\ &f_L(x,Q^2,s_2)\hat{F}_L(x,Q^2) + f_2(x,Q^2,s_2)\hat{F}_2(x,Q^2) \\ &\quad = \hat{S}_{NC}^{-}(x,Q^2|s_2) \end{aligned}\right\}. \tag{100}$$

We introduce the notation

$$C(x,Q^2) = \begin{pmatrix} f_L(x,Q^2,s_1) & f_2(x,Q^2,s_1) \\ f_L(x,Q^2,s_2) & f_2(x,Q^2,s_2) \end{pmatrix}.$$

Then the estimate of $F_2(x,Q^2)$, for example, can be written in the form

$$\hat{F}_2(x,Q^2) = \sum_j C_{2j}^{-1}\hat{S}_{NC}^{-}(x,Q^2|s_j)$$

$$= \sum_{j,n,m} C_{2j}^{-1}A_n(x,Q^2,s_j)\tilde{G}_{nm}^{-1}(s_j)Y_m^j, \tag{101}$$

where Y_m^j are the measurements for $s=s_j$. From the expression (101) and the fact that in the framework of this model the $\bar{Y}_n^j$ satisfy

$$\bar{Y}_n^j = \int dx dQ^2 A_n(x,Q^2,s_j)[f_1F_L + f_2F_2],$$

it follows that

$$\bar{\hat{F}}_2(x,Q^2) = \int dx' dQ^{2'} A_{R2L}F_L + \int dx' dQ^{2'} A_{R22}F_2, \tag{102}$$

where

$$A_{R2L}(x,Q^2,x',Q^{2'}) = \sum_{j,n,m} C_{2j}^{-1}A_n(x,Q^2,s_j)\tilde{G}_{nm}^{-1}(s_j)$$

$$\times A_m(x',Q^{2'},s_j)f_L(x',Q^{2'},s_j),$$

$$A_{R22}(x,Q^2,x',Q^{2'}) = \sum_{j,n,m} C_{2j}^{-1}A_n(x,Q^2,s_j)\tilde{G}_{nm}^{-1}(s_j)$$

$$\times A_m(x',Q^{2'},s_j)f_2(x',Q^{2'},s_j).$$

It follows from the expression (102) that, in contrast to the model (97), the bias of the estimate $\hat{F}_2(x,Q^2)$ depends not only on the function $F_2(x,Q^2)$ itself but also on the function $F_L(x,Q^2)$.

Similarly, for $\hat{F}_L(x,Q^2)$ we readily find that

$$\bar{\hat{F}}_L(x,Q^2) = \int dx' dQ^{2'} A_{RLL}F_L + \int dx' dQ^{2'} A_{RL2}F_2, \tag{103}$$

where

$$A_{RLL}(x,Q^2,x',Q^{2'}) = \sum_{j,n,m} C_{Lj}^{-1}A_n(x,Q^2,s_j)\tilde{G}_{nm}^{-1}(s_j)$$

$$\times A_m(x',Q^{2'},s_j)f_L(x',Q^{2'},s_j),$$

$$A_{RL2}(x,Q^2,x',Q^{2'}) = \sum_{j,n,m} C_{Lj}^{-1}A_n(x,Q^2,s_j)\tilde{G}_{nm}^{-1}(s_j)$$

$$\times A_m(x',Q^{2'},s_j)f_2(x',Q^{2'},s_j).$$

The noise characteristics $\hat{F}_{i\varepsilon}(x,Q^2)$ of the estimates $\hat{F}_i(x,Q^2)$ are determined by the error operator

$$\Sigma_i(x,Q^2,x',Q^{2'},s) = \sum_{j,n,m,l,k} C_{ij}^{-1}A_n(x,Q^2,s_j)\tilde{G}_{nm}^{-1}(s_j)$$

$$\times K_{ml}^j\tilde{G}_{lk}^{-1}A_k(x',Q^{2'},s_j)C_{ij}^{-1}, \tag{104}$$

where K^j is the error matrix of the measurements Y_n^j for $s=s_j$, and we have used the fact that the errors of the measurements for different s_j are independent.

Model (95). In this case, the estimates of the structure functions F_2 and F_3 can be sought either from measurements of the two cross sections $S_{NC}^{\pm}$ for one value of s or from measurements of one cross section for two values of s: s_1 and s_2. We consider first the first possibility. Equating the cross sections to their estimates, we obtain the system of equations

$$\left.\begin{aligned} &f_2(x,Q^2,s)\hat{F}_2(x,Q^2) + f_3(x,Q^2,s)\hat{F}_3(x,Q^2) \\ &\quad = \hat{S}_{NC}^{-}(x,Q^2|s) \\ &f_2(x,Q^2,s)\hat{F}_2(x,Q^2) - f_3(x,Q^2,s)\hat{F}_3(x,Q^2) \\ &\quad = \hat{S}_{NC}^{+}(x,Q^2|s) \end{aligned}\right\}. \tag{105}$$

From the system (105), we obtain[9)]

$$\left.\begin{aligned}
&\hat{F}_2(x,Q^2)=\tfrac{1}{2}(\hat{S}^-_{NC}+\hat{S}^+_{NC})\\
&f_2=\tfrac{1}{2}\sum_n A_n\tilde{G}^{-1}_{nm}(Y^-_m+Y^+_m)/f_2\\
&\hat{F}_3(x,Q^2)=\tfrac{1}{2}(\hat{S}^-_{NC}-\hat{S}^+_{NC})\\
&f_3=\tfrac{1}{2}\sum_n A_n\tilde{G}^{-1}_{nm}(Y^-_m+Y^+_m)/f_3
\end{aligned}\right\}. \tag{106}$$

In the framework of this model

$$\bar{Y}^{\pm}_n=\int dxdQ^2A_n(x,Q^2,s)[\,f_2F_2\mp f_3F_3]. \tag{107}$$

Using this expression, we find that

$$\bar{\bar{F}}_i(x,Q^2)=\int dx'dQ^{2\prime}A_{Ri}(x,Q^2,x',Q^{2\prime},s)F_i(x',Q^{2\prime}),$$

$$A_{Ri}(x,Q^2,x',Q^{2\prime},s)=\sum_{n,m} f_i^{-1}(x,Q^2,s)A_n(x,Q^2s)\times\tilde{G}^{-1}_{nm}A_m(x',Q^{2\prime},s)f_i(x',Q^{2\prime},s),$$

$$\mathrm{Sp}\,A_R=\mathrm{Sp}\;G\tilde{G}^{-1}. \tag{108}$$

The noise component of the estimate $\hat{F}_i(x,Q^2)$ is characterized by the error operator

$$\Sigma_i(x,Q^2,x',Q^{2\prime},s)=\sum_{n,m,k,l}\tfrac{1}{4}f_i^{-1}(x,Q^2,s)A_n(x,Q^2,s)\tilde{G}^{-1}_{nm}\times K^i_{mk}\tilde{G}^{-1}_{kl}A_l(x',Q^{2\prime},s)f_i^{-1}(x',Q^{2\prime},s), \tag{109}$$

where K^2 is the error matrix of the measurements $Y^-_n+Y^+_n$, K^3 is the error matrix of the measurements $Y^-_n+Y^+_n$, and as a whole

$$\overline{\|F_{ie}\|^2}=\mathrm{Sp}\;\tilde{G}^{-1}K^i\tilde{G}^{-1}\langle f_i^{-1}A\,|\,f_i^{-1}A\rangle.$$

Model (94). Suppose that we are given measurements $\hat{S}^{\pm}_{NC}(x,Q^2|s)$ for $s=s_1$, s_2. Then there is a possibility of estimating all three structure functions $F_L(x,Q^2)$, $F_2(x,Q^2)$, $F_3(x,Q^2)$ in accordance with the following scheme. From the system of equations for $s_1>s_2$,

$$\left.\begin{aligned}
&f_L(x,Q^2,s_1)\hat{F}_L(x,Q^2)+f_2(x,Q^2,s_1)\hat{F}_2(x,Q^2)\\
&\quad+f_3(x,Q^2,s_1)\hat{F}_3(x,Q^2)=\hat{S}^-_{NC}(x,Q^2|s_1),\\
&f_L(x,Q^2,s_1)\hat{F}_L(x,Q^2)+f_2(x,Q^2,s_1)\hat{F}_2(x,Q^2)\\
&\quad-f_3(x,Q^2,s_1)\hat{F}_3(x,Q^2)=\hat{S}^+_{NC}(x,Q^2|s_1)
\end{aligned}\right\} \tag{110}$$

we have

$$\hat{F}_3(x,Q^2)=\tfrac{1}{2}[\hat{S}^-_{NC}(x,Q^2|s_1)-\hat{S}^+_{NC}(x,Q^2|s_1)]/f_3(x,Q^2,s_1), \tag{111}$$

$$f_L(x,Q^2,s_1)\hat{F}_L(x,Q^2)+f_2(x,Q^2,s_1)\hat{F}_2(x,Q^2)=\tfrac{1}{2}[\hat{S}^-_{NC}(x,Q^2|s_1)+\hat{S}^+_{NC}(x,Q^2|s_1)].$$

In this case,

$$\bar{Y}^{\pm}_n=\int dxQ^2A_n(x,Q^2,s_1)[\,f_LF_L+f_2F_2\mp f_3F_3]. \tag{112}$$

Using the last relation, we calculate $\bar{\bar{F}}_3(x,Q^2)$:

$$\begin{aligned}
\bar{\bar{F}}_3(x,Q^2)&=\tfrac{1}{2}f_3^{-1}(x,Q^2,s_1)\sum_{n,m}A_n(x,Q^2,s_1)\tilde{G}^{-1}_{nm}(s_1)[\,Y^-_m-Y^+_m]\\
&=f_3^{-1}(x,Q^2,s_1)\sum_{n,m}A_n(x,Q^2,s_1)\tilde{G}^{-1}_{nm}(s_1)\\
&\quad\times\int dx'dQ^{2\prime}A_mf_3F_3\\
&=\int dx'dQ^{2\prime}A_{R3}(x,Q^2,x',Q^{2\prime})F_3(x',Q^{2\prime}),
\end{aligned} \tag{113}$$

where

$$A_{R3}(x,Q^2,x',Q^{2\prime})=\sum_{n,m}f_3^{-1}(x,Q^2,s_1)A_n(x,Q^2,s_1)\times\tilde{G}^{-1}_{nm}(s_1)A_m(x',Q^{2\prime},s_1)f_3(x',Q^{2\prime},s_1).$$

For the error operator of the estimate $\hat{F}_3(x,Q^2)$ the following expression is valid:

$$\Sigma_3(x,Q^2,x',Q^{2\prime},s_1)=\sum_{n,m,k,l}\tfrac{1}{4}f_3^{-1}(x,Q^2,s_1)\tilde{G}^{-1}_{nm}(s_1)(K^-+K^+)_{mk}\times\tilde{G}^{-1}_{kl}(s_1)A_l(x',Q^{2\prime},s_1)f_3^{-1}(x',Q^{2\prime},s_1). \tag{114}$$

To obtain estimates of $F_L(x,Q^2)$ and $F_2(x,Q^2)$, we use the half-sum of the cross sections $\hat{S}^{\pm}_{NC}(x,Q^2|s)$ for $s=s_2$. Together with Eq. (110), we have

$$\left.\begin{aligned}
&f_L(x,Q^2,s_1)\hat{F}_L(x,Q^2)+f_2(x,Q^2,s_1)\hat{F}_2(x,Q^2)\\
&\quad=\tfrac{1}{2}[\hat{S}^-_{NC}(x,Q^2|s_1)+\hat{S}^+_{NC}(x,Q^2|s_1)],\\
&f_L(x,Q^2,s_2)\hat{F}_L(x,Q^2)+f_2(x,Q^2,s_1)\hat{F}_2(x,Q^2)\\
&\quad=\tfrac{1}{2}[\hat{S}^-_{NC}(x,Q^2|s_2)+\hat{S}^{\pm}_{NC}(x,Q^2|s_2)]
\end{aligned}\right\}. \tag{115}$$

Introducing the matrix $C(x,Q^2)$, as we did in the solution of the system (100), we obtain for $\hat{F}_i(x,Q^2)$ the expression

$$\begin{aligned}
\hat{F}_i(x,Q^2)&=\tfrac{1}{2}\sum_j C^{-1}_{ij}[\hat{S}^-_{NC}(x,Q^2|s_j)+\hat{S}^+_{NC}(x,Q^2|s_j)]\\
&=\tfrac{1}{2}\sum_j C^{-1}_{ij}\sum_{n,m}A_n(x,Q^2,s_j)\tilde{G}^{-1}_{nm}(s_j)\\
&\quad\times[\,Y^-_m(s_j)+Y^+_m(s_j)].
\end{aligned} \tag{116}$$

Using the expression (112), we obtain $\bar{\bar{F}}_i(x,Q^2)$:

$$\bar{\bar{F}}_i(x,Q^2)=\tfrac{1}{2}\sum_j C^{-1}_{ij}\sum_{n,m,k}A_n(x,Q^2,s_j)\tilde{G}^{-1}_{nm}(s_j)\times\int dx'dQ^{2\prime}A_m(x',Q^{2\prime},s_j)\sum_k f_k\times(x',Q^{2\prime},s_j)F_k(x',Q^{2\prime}),\quad k=L,2.$$

Thus,

$$\bar{\hat{F}}_i(x,Q^2)=\sum_k \int dx'dQ^{2'}A_{Rik}(x,Q^2,x',Q^{2'})F_k(x',Q^{2'}),$$

$$A_{Rik}(x,Q^2,x',Q^{2'})=\sum_j C_{ij}^{-1}\sum_{n,m} A_n(x,Q^2,s_j)\tilde{G}_{nm}^{-1}(s_j)$$
$$\times A_m(x',Q^{2'},s_j)f_k(x',Q^{2'},s_j). \quad (117)$$

For the error operator of the noise component

$$\hat{F}_{i\varepsilon}(x,Q^2)=\tfrac{1}{2}\sum_j C_{ij}^{-1}\sum_{n,m} A_n(x,Q^2,s_j)\tilde{G}_{nm}^{-1}(s_j)$$
$$\times[\varepsilon_m^-(s_j)+\varepsilon_m^+(s_j)]$$

we obtain the expression

$$\Sigma_i(x,Q^2,x',Q^{2'})=\tfrac{1}{4}\sum_j C_{ij}^{-1}\sum_{n,m,l,k} A_n(x,Q^2,s_j)\tilde{G}_{nm}^{-1}(s_j)$$
$$\times[K^-(s_j)+K^+(s_j)]$$
$$\times\tilde{G}_{lk}^{-1}(s_j)A_k(x',Q^{2'},s_j)C_{ij}^{-1}. \quad (118)$$

As we showed, $\bar{\hat{S}}(x,Q^2|s)\neq S(x,Q^2|s)$. Therefore, our use of the equation $S(x,Q^2|s)=\hat{S}(x,Q^2|s)$ is not entirely correct and leads to a systematic error in the estimate of the structure functions $F_i(x,Q^2)$. Therefore, it is natural to estimate the structure functions directly from the experimental data. Of course, these estimates will also be biased. However, *a priori*, without making numerical experiments, it is difficult to give preference to one of these approaches.

Estimate of the structure functions F_L, F_2, F_3 from experimental data

Model (97). Suppose that we are given measurements of, for example, the cross section $S_{NC}^{-}(x,Q^2|s)$. In the framework of this very simple model, their expected values can be written in the form

$$\bar{Y}_n=\int dxdQ^2A_n^2(x,Q^2,s)F_2(x,Q^2),$$

where

$$A_n^2(x,Q^2,s)=A_n(x,Q^2,s)f_2(x,Q^2,s).$$

Therefore, the problem of estimating $F_2(x,Q^2)$ is formally identical to the problem of estimating the cross section $S_{NC}^{-}(x,Q^2|s)$ itself. Thus, we write down directly the expression for the estimate,

$$\hat{F}^2(x,Q^2)=\sum_{n,m} A_n^2(x,Q^2,s)\tilde{G}_{nm}^{-1}Y_m,$$

where $\tilde{G}$ is the regularized Gram matrix of the functions $A_n^2(x,Q^2,s)$, and its characteristics

$$A_R(x,Q^2,x',Q^{2'})=\sum_{n,m} A_n^2(x,Q^2,s)\tilde{G}_{nm}^{-1}A_m^2(x',Q^{2'},s),$$

$$\mathrm{Sp}\,A_R=\mathrm{Sp}\,G\tilde{G}^{-1},$$

$$\Sigma(x,Q^2,x',Q^2)=\sum_{n,m,l,k} A_n^2(x,Q^2,s)\tilde{G}_{nm}^{-1}K_{ml}\tilde{G}_{lk}^{-1}A_k^2(x',Q^{2'},s),$$

$$\overline{\|F_{2\varepsilon}\|}=\mathrm{Sp}\,G\tilde{G}^{-1}K\tilde{G}^{-1}.$$

Model (96). Suppose, as before, that we are given measurements $Y_n(s_1)$ and $Y_n(s_2)$, $n=1...,N$, of, for example, the cross section $S_{NC}^{-}(x,Q^2|s)$ for two values of s. We write $\bar{Y}_n(s_i)$ in the form

$$\bar{Y}_n^i\equiv\bar{Y}_n(s_i)=\int dxdQ^2A_n^L(x,Q^2,s_i)F_L(x,Q^2)$$
$$+\int dxdQ^2A_n^2(x,Q^2,s_i)F_2(x,Q^2), \quad (119)$$

where

$$A_n^L(x,Q^{2'},s_i)=A_n(x,Q^2,s_i)f_L(x,Q^2,s_i),$$

$$A_n^2(x,Q^{2'},s_i)=A_n(x,Q^2,s_i)f_2(x,Q^2,s_i).$$

We introduce a single vector of the measurements

$$\langle Y|=(Y_1...Y_NY_{N+1}...Y_{2N}),$$

its weight matrix W, its covariance matrix K, the vector functions

$$\langle A^L(x,Q^2)|=(A_1^L...A_N^LA_{N+1}^L...A_{2N}^L),$$

$$\langle A^2(x,Q^2)|=(A_1^2...A_N^2A_{N+1}^2...A_{2N}^2),$$

and the vector of structure functions $\langle F(x,Q^2)|=(F_L,F_2)$. Choosing different forms of the stabilizer in the functional of Tikhonov's method, we can obtain different estimates $|\hat{F}(x,Q^2)\rangle$. We choose the simplest one:

$$\Phi=(Y-\bar{Y})^TW(Y-\bar{Y})+\alpha\int dxdQ^2F_L^2+\alpha\int dxdQ^2F_2$$
$$\equiv(Y-\bar{Y})^TW(Y-\bar{Y})+\alpha\|F\|^2. \quad (120)$$

We consider the estimate $|\hat{F}(x,Q^2)\rangle$ in accordance with the functional (120). Variation of (120) with respect to $F_L(x,Q^2)$ and $F_2(x,Q^2)$ leads to the system of equations

$$\frac{\delta\Phi}{\delta F_L}=F_L-\langle A^L(x,Q^2)|W|Y-\bar{Y}\rangle=0,$$
$$\frac{\delta\Phi}{\delta F_2}=F_2-\langle A^2(x,Q^2)|W|Y-\bar{Y}\rangle=0. \quad (121)$$

It follows from it that for $\hat{F}_L(x,Q^2)$ and $\hat{F}_2(x,Q^2)$

$$\left|\begin{matrix}\hat{F}_L(x,Q^2)\\ \hat{F}_2(x,Q^2)\end{matrix}\right\rangle=\frac{1}{\alpha}\sum_n\Delta_n\left|\begin{matrix}A_n^L(x,Q^2)\\ A_n^2(x,Q^2)\end{matrix}\right\rangle, \quad (122)$$

where

$$|\Delta\rangle=W|Y-\bar{Y}\rangle. \quad (123)$$

Substitution of the expressions (122) in (123) with the use of (119) leads to the following system of equations for Δ_n:

$$[\alpha K+\langle A^L|A^L\rangle+\langle A^2|A^2\rangle]\,|\Delta\rangle=|Y\rangle.$$

The elements of the matrices $\langle A^L|A^L\rangle$ and $\langle A^2|A^2\rangle$ have the form

$$\langle A^L|A^L\rangle_{nm}=\int dxdQ^2A_n^LA_m^L,$$

$$\langle A^2|A^2\rangle_{nm}=\int dxdQ^2A_n^2A_m^2.$$

In what follows, we shall for brevity use the notation

$$R(\alpha)=\alpha K+\langle A^L|A^L\rangle+\langle A^2|A^2\rangle.$$

Thus, we arrive at the following estimates for $F_L(x,Q^2)$ and $F_2(x,Q^2)$:

$$\left|\begin{matrix}\hat F_L(x,Q^2)\\ \hat F_2(x,Q^2)\end{matrix}\right\rangle=\sum_{n,m}\left|\begin{matrix}A_N^L(x,Q^2)\\ A_n^2(x,Q^2)\end{matrix}\right\rangle R_n^{-1}(\alpha)Y_m. \tag{124}$$

Using the expression (119) for $\bar Y_m$, we can write $\hat F_L$ and $\hat F_2$ in the form

$$\left|\begin{matrix}\bar{\hat F}_L\\ \bar{\hat F}_2\end{matrix}\right\rangle=\int dx'dQ^{2'}A_R\left|\begin{matrix}F_L\\ F_2\end{matrix}\right\rangle$$

$$=\int dx'dQ^{2'}\begin{pmatrix}A_{RLL} & A_{RL2}\\ A_{R2L} & A_{R22}\end{pmatrix}\left|\begin{matrix}F_L\\ F_2\end{matrix}\right\rangle, \tag{125}$$

where

$$A_{RLL}=\sum_{n,m}A_n^L(x,Q^2)R_{nm}^{-1}(\alpha)A_m^L(x',Q^{2'}),$$

$$A_{RL2}=\sum_{n,m}A_n^L(x,Q^2)R_{nm}^{-1}(\alpha)A_m^2(x',Q^{2'}),$$

$$A_{R2L}=\sum_{n,m}A_n^2(x,Q^2)R_{nm}^{-1}(\alpha)A_m^L(x',Q^{2'}),$$

$$A_{R22}=\sum_{n,m}A_n^2(x,Q^2)R_{nm}^{-1}(\alpha)A_m^2(x',Q^{2'}).$$

It follows from (124) that the error operator of the function vector $|\hat F\rangle$ is

$$\Sigma(x,Q^2|x',Q^{2'})=\begin{pmatrix}\Sigma_{LL} & \Sigma_{L2}\\ \Sigma_{2L} & \Sigma_{22}\end{pmatrix}, \tag{126}$$

where

$$\Sigma_{LL}=\sum_{n,m,l,k}A_n^L(x,Q^2)R_{nm}^{-1}(\alpha)K_{ml}R_{lk}^{-1}(\alpha)A_k^L(x',Q^{2'}),$$

$$\Sigma_{L2}=\Sigma_{2L}^T$$

$$=\sum_{n,m,l,k}A_n^L(x,Q^2)R_{nm}^{-1}(\alpha)K_{ml}R_{lk}^{-1}(\alpha)A_k^2(x',Q^{2'}),$$

$$\Sigma_{22}=\sum_{n,m,l,k}A_n^2(x,Q^2)R_{nm}^{-1}(\alpha)K_{ml}R_{lk}^{-1}(\alpha)A_k^2(x',Q^{2'}).$$

As a whole, the estimate $|\hat F(x,Q^2)\rangle$ is characterized by

$$\mathrm{Sp}\,A_R=\mathrm{Sp}\,R(0)R^{-1}(\alpha),$$

$$\overline{\|F_\varepsilon\|^2}=\mathrm{Sp}\,R(0)R^{-1}(\alpha)KR^{-1}(\alpha).$$

We now consider what in this case are the "measured" and "unmeasured" components of the function vector $|F\rangle$. It follows from the definition (119) for $\bar Y_n$ that for the "unmeasured" component $F^\perp$

$$\int dxdQ^2A_n^LF_L^\perp+\int dxdQ^2A_n^2F_2^\perp=0. \tag{127}$$

The "measured" component $|F^A\rangle$ must be orthogonal to the component $|F^\perp\rangle$ for the chosen scalar product in the space of the function vectors $|F^A\rangle$. If it is chosen in accordance with the stabilizer in the functional (120), we must have

$$\int dxdQ^2F_L^AF_L^\perp+\int dxdQ^2F_2^AF_2^\perp=0.$$

The relation (127) will hold if

$$\left\langle F^A(x,Q^2)\right|=\left\langle\sum_n a_nA_n^L,\sum_n a_nA_n^2\right|. \tag{128}$$

One can show that substitution of $|F^A(x,Q^2)\rangle$ in accordance with (128) in the least-squares functional and its minimization with respect to a_n leads to the estimate

$$\left|\begin{matrix}\hat F_L^+(x,Q^2)\\ \hat F_2^+(x,Q^2)\end{matrix}\right\rangle=\sum_{n,m}\left|\begin{matrix}A_n^L(x,Q^2)\\ A_n^2(x,Q^2)\end{matrix}\right\rangle R_n^{-1}(0)Y_m, \tag{129}$$

which agrees with (124) for $\alpha=0$.

We consider one further method of regularizing $R(0)$—the spectral-window method. We introduce the function vector $|\Psi_l(x,Q^2)\rangle$ in accordance with the expression

$$|\Psi_L\rangle=\frac{1}{\sqrt{\lambda_l}}\sum_n r_n^{(l)}\left|\begin{matrix}A_n^L\\ A_n^2\end{matrix}\right\rangle, \tag{130}$$

where $r_n^{(l)}$ are the eigenvectors of the matrix $R(0)$, and λ_l are the eigenvalues.

It can be shown that the functions $|\Psi_l\rangle$ form an orthogonal basis in the "measured" space for the previously introduced scalar product:

$$\langle\Psi_k|\Psi_l\rangle=\sum_{n,m}\frac{1}{\sqrt{\lambda_l}}\frac{1}{\sqrt{\lambda_k}}r_n^{(l)}r_m^{(k)}[\langle A^L|A^L\rangle+\langle A^2|A^2\rangle]_{nm}$$

$$=\sum_{n,m}\frac{1}{\sqrt{\lambda_l}}\frac{1}{\sqrt{\lambda_k}}r_n^{(l)}R_{nm}(0)r_m^{(k)}$$

$$=\sum_n\frac{1}{\sqrt{\lambda_l}}\frac{1}{\sqrt{\lambda_k}}r_n^{(l)}\lambda_kr_m^{(k)}$$

$$=\delta_{lk}. \tag{131}$$

Replacing $R^{-1}(0)$ in (129) by

$$\tilde R_{nm}^{-1}=\sum_l\frac{f(\lambda_l)}{\lambda_l}r_n^{(l)}r_m^{(l)}, \tag{132}$$

where $f(\lambda)\leqslant 1$ and $f(\lambda)/\lambda\to\text{const}$ as $\lambda\to 0$, we have

$$\left|\begin{matrix}\hat F_L(x,Q^2)\\ \hat F_2(x,Q^2)\end{matrix}\right\rangle=\sum_{l,n}|\Psi_l(x,Q^2)\rangle\frac{f(\lambda_l)}{\sqrt{\lambda_l}}r_n^{(l)}Y_n. \tag{133}$$

The simplest choice of $f(\lambda)$ is the step function: $f(\lambda)=1$ for $\lambda>\lambda_{N_R}$ and $f(\lambda)=0$ for $\lambda<\lambda_{N_R}$. Proceeding as before, we find that in this case

$$A_R(x,Q^2,x',Q^{2'})=\sum_l|\Psi_l(x,Q^2)\rangle f(\lambda_l)\langle\Psi_l(x',Q^{2'})|,$$

$$\Sigma(x,Q^2,x',Q^{2'})=\sum_{l,n,m,k}|\Psi_l(x,Q^2)\rangle\frac{f(\lambda_l)}{\sqrt{\lambda_l}}g_n^{(l)}$$
$$\times K_{nm}g_m^{(k)}\frac{f(\lambda_k)}{\sqrt{\lambda_k}},\Psi_k(x',Q^{2'})|,$$
$$\overline{\|F_\varepsilon\|^2}=\sum_{l,m,n}\frac{1}{\lambda_l}f^2(\lambda_l)g_n^{(l)}K_{nm}g_m^{(l)}.$$

A different orthogonal basis, in which the expressions A_R, $\mathrm{Sp}\,A_R$, Σ, $\overline{\|F_\varepsilon\|^2}$ simplify, is given in Ref. 35.

Model (95). Suppose that we are given measurements $S_{NC}^+(x,Q^2|s)$ and $S_{NC}^-(x,Q^2|s)$ for one value of s. Then estimates of $F_2(x,Q^2)$, $F_3(x,Q^3)$ can be constructed as follows. From the expression (95) we can calculate the structure functions F_2 and F_3 in terms of the cross sections $S_{NC}^+(x,Q^2|s)$ and $S_{NC}^-(x,Q^2|s)$:

$$F_2(x,Q^2)=[S_{NC}^-(x,Q^2|s)+S_{NC}^+(x,Q^2|s)]/2f_2(y,Q^2),$$
$$F_3(x,Q^2)=[S_{NC}^-(x,Q^2|s]-S_{NC}^+(x,Q^2|s))/2f_3(y,Q^2).$$

This fact suggests that it is sensible to go over from the measurements $Y_n^\pm$ to the statistics

$$Y_n^2=\tfrac{1}{2}(Y_n^-+Y_n^+),\quad Y_n^3=\tfrac{1}{2}(Y_n^--Y_n^+).$$

Their mathematical expectations in the framework of this model are

$$\bar{Y}_n^2=\int dxdQ^2A_n^2(x,Q^2,s)F_2(x,Q^2),$$
$$\bar{Y}_n^3=\int dxdQ^2A_n^3(x,Q^2,s)F_3(x,Q^2),$$

where

$$A_n^2(x,Q^2,s)=A_n(x,Q^2,s)f_2(x,Q^2,s),$$
$$A_n^3(x,Q^2,s)=A_n(x,Q^2,s)f_3(x,Q^2,s).$$

Therefore, the "measured" components of $F_2(x,Q^2)$ and $F_3(x,Q^2)$ are linear combinations of $A_n^2(x,Q^2,s)$ and $A_n^3(x,Q^2,s)$, respectively. Thus, we seek estimates of F_2 and F_3 in the form

$$\hat{F}_2(x,Q^2)=\sum_{n=1}^{N}\hat{a}_nA_n^2(x,Q^2|s),$$
$$\hat{F}_3(x,Q^2)=\sum_{n=1}^{N}\hat{b}_nA_n^3(x,Q^2|s). \tag{134}$$

We obtain the estimates $\hat{a}_n$, $\hat{b}_n$ in accordance with the least-squares method by minimizing the functionals

$$\Phi_2=(Y^2-\bar{Y}^2)^TW^2(Y^2-\bar{Y}^2),$$
$$\Phi_3=(Y^3-\bar{Y}^3)^TW^3(Y^3-\bar{Y}^3),$$

and substituting in them the expressions (22) and (23):

$$\hat{a}_n=\sum_m(G_2)_{nm}^{-1}Y_m^2,\quad (G_2)_{nm}=\langle A_n^2|A_m^2\rangle,$$
$$\hat{b}_n=\sum_m(G_3)_{nm}^{-1}Y_m^3,\quad (G_3)_{nm}=\langle A_n^3|A_m^3\rangle.$$

It can be shown that the estimates

$$\hat{F}_2^+(x,Q^2)=\sum_{n,m}A_n^2(x,Q^2,s)(G_2)_{nm}^{-1}Y_m^2,$$
$$\hat{F}_3^+(x,Q^2)=\sum_{n,m}A^3(x,Q^2,s)(G_3)_{nm}^{-1}Y_m^3 \tag{135}$$

are unbiased estimates of the "measured" components of F_2 and F_3 with the residual instrumental functions

$$A_R^{+(2)}(x,Q^2|x',Q^{2'})=\sum_{n,m}A_n^2(x,Q^2,s)$$
$$\times(G_2)_{nm}^{-1}A_m^2(x',Q^{2'},s),$$
$$A_R^{+(3)}(x,Q^2|x',Q^{2'})=\sum_{n,m}A_n^3(x,Q^2,s)$$
$$\times(G_3)_{nm}^{-1}A_m^3(x',Q^{2'},s)$$

and correlation operators

$$\Sigma^{+(2)}(x,Q^2|x',Q^{2'})=\sum_{n,m,l,k}A_n^2(x,Q^2,s)$$
$$\times(G_2)_{nm}^{-1}K_{ml}^2(G_2)_{lk}^{-1}A_k^2(x',Q^{2'},s),$$
$$\Sigma^{+(3)}(x,Q^2|x',Q^{2'})=\sum_{n,m,l,k}A_n^3(x,Q^2,s)$$
$$\times(G_3)_{nm}^{-1}K_{ml}^3(G_3)_{lk}^{-1}A_k^3(x',Q^{2'},s).$$

Overall, it is convenient to characterize $\hat{F}^+(x,Q^2)$ by the quantities

$$\mathrm{Sp}\,A_R^+=\int dxdQ^2A_R^+(x,Q^2|x,Q^2)=N,$$
$$\overline{\|F_\varepsilon^+\|^2}=\mathrm{Sp}\,KG^{-1},\quad F_\varepsilon^+(x,Q^2)=\sum_{n,m}A_n(x,Q^2)G^{-1}\varepsilon_m. \tag{136}$$

Model (94). Suppose that we are given measurements of $S_{NC}^\pm(x,Q^2|s)$ for $s=s_1$ and $S_{NC}^-(x,Q^2|s)$ for $s=s_2<s_1$: $Y^\pm(s_i)$. The mathematical expectations of these quantities can be expressed in the form

$$\bar{Y}_n^\pm(s_i)=\int dxdQ^2[A_n^L(x,Q^2,s)F_L(x,Q^2)$$
$$+A_n^2(x,Q^2,s_i)F_2(x,Q^2)$$
$$\mp A_n^3(x,Q^2,s_i)F_3(x,Q^2)], \tag{137}$$

where

$$A_{in}^L\equiv A_n^L(x,Q^2,s_i)=A_n(x,Q^2,s_i)f_L(x,Q^2,s_i),$$
$$A_{in}^2\equiv A_n^2(x,Q^2,s_i)=A_n(x,Q^2,s_i)f_2(x,Q^2,s_i),$$
$$A_{in}^3\equiv A_n^3(x,Q^2,s_i)=A_n(x,Q^2,s_i)f_3(x,Q^2,s_i).$$

We order all the measurements Y_n as follows: $Y_n=Y_n^+(S_1)$ for $n=1,...,N$; $Y_n=Y_n^-(s_1)$ for $n=N+1,...,2N$; $Y_n=Y_n^-(s_2)$ for $n=2N+1,...,3N$. We introduce accordingly function vectors $|A_n\rangle$, which are equal to

$$|A_n\rangle=\begin{vmatrix} A_n^L \\ A_n^2 \\ -A_n^3 \end{vmatrix} \quad n=1,\ldots,N;$$

$$|A_n\rangle=\begin{vmatrix} A_n^L \\ A_n^2 \\ A_n^3 \end{vmatrix} \quad n=N+1,\ldots,2N; \tag{138}$$

$$|A_n\rangle=\begin{vmatrix} A_n^L \\ A_n^2 \\ A_n^3 \end{vmatrix} \quad n=2N+1,\ldots,3N.$$

Using the expressions (137) and (138), we see that the mean value of any measurement Y_n can be expressed in the form of the scalar product

$$\bar{Y}_n=\langle A_n|F\rangle, \quad n=1,\ldots,3N, \tag{139}$$

where $\langle F|=(F_L,F_2,F_3)$. It follows from this that the "unmeasured" component $|F^{\perp}\rangle$ satisfies the conditions

$$\langle A_n|F^{\perp}\rangle=0, \quad n=1,\ldots,3N.$$

Therefore, if $|F^A\rangle$ is sought in the form

$$|F^A\rangle=\sum_n a_n|A_n\rangle, \quad n=1,\ldots,3N, \tag{140}$$

then the condition of orthogonality $\langle F^A|F^{\perp}\rangle$ of the "measured" and "unmeasured" components will be satisfied.

We seek the least-squares estimate of $|F\rangle$ by substituting the expression (140) in the least-squares functional

$$\Phi=\sum_m (Y_m-\bar{Y}_m)W_m(Y_m-\bar{Y}_m)$$

and minimizing with respect to a_n.

We then obtain for $|\hat{F}^+\rangle$ the expression

$$|\hat{F}^+\rangle=\sum_{n,m}|A_m\rangle G_{nm}^{-1}Y_m, \tag{141}$$

where G is the Gram matrix of the function vectors $|A_n\rangle$: $G_{nm}=\langle A_n|A_m\rangle$.

Using (141), we obtain

$$\overline{|\hat{F}^+\rangle}=\sum_{n,m}|A_m\rangle G_{nm}^{-1}\langle A_m|F\rangle.$$

Therefore, A_R in this case is a matrix integral operator

$$A_{Rij}^+(x,Q^2,x',Q^{2'})=\sum_{n,m}|A_n^i(x,Q^2)\rangle G_{nm}^{-1}\langle A^j(x',Q^{2'})|. \tag{142}$$

The error operator of the noise component

$$|\hat{F}_{\varepsilon}^+\rangle=\sum_{n,m}|A_n\rangle G_{nm}^{-1}\varepsilon_m \tag{143}$$

is

$$\Sigma^+(x,Q^2,x',Q^{2'})=\sum|A_n\rangle G_{nm}^{-1}K_{ml}G_{lk}^{-1}\langle A_k|. \tag{144}$$

If some regularization method is used, the expressions (141)–(144) are obtained with the replacement of G by the regularized Gram matrix $\tilde{G}$, and the systematic and statistical errors $|\hat{F}\rangle$ are characterized by

$$\mathrm{Sp}\,A_R=\mathrm{Sp}\,\tilde{G}G^{-1},$$
$$\|\hat{F}_{\varepsilon}\|^2=\mathrm{Sp}\,G\tilde{G}^{-1}K\tilde{G}^{-1}.$$

Neutrino and antineutrino scattering through the charged-current channel

The nucleon structure functions can be estimated in this case by the methods presented for muon and electron scattering. However, one cannot directly use the given expressions for the estimates and their characteristics, since, in contrast to monochromatic beams of muons (or electrons), beams of neutrinos (or antineutrinos) have rather broad distributions with respect to their energies. This significantly changes the expressions. For lack of space, we refer the reader to Ref. 36. In another paper,[37] the reader can find an approach that generalizes the use of B splines.[11] We mention that the use, in the study, of only two-dimensional splines even though the problem depends on the three variables E_ν, x, Q^2 leads to unjustified approximations.

Use of estimates of structure functions to determine Λ in QCD

It is obvious that the arguments given at the end of the first section relating to parametric analysis also remain valid for this problem. The use of the initial experimental data Y_n to determine Λ in QCD from a nonsinglet fit for data from a 15-ft bubble chamber can be found in Ref. 38.

CONCLUSIONS

We thank our coauthors of the original studies used in this review: V. V. Ammosov, E. A. Belogorlov, T. B. Kostkina, A. A. Spiridonov, and Z. U. Usubov; we thank S. N. Sokolov, L. A. Tyapkin, and Yu. P. Pyt'ev for fruitful discussions; and we thank E. V. Zaĭchenko and R. A. Makeeva for tremendous assistance in preparing the text and diagrams.

The results of the numerical experiments and the illustrations (except for Fig. 15) used in the review were obtained by means of the program package SLIP1 (Solution of Linear Inverse Problem in 1-Dimension).[39]

[1)] In what follows, the bar above an expression will denote averaging with respect to the measurement errors, and the symbol ˆ over a function will denote an estimate of the function.

[2)] Note that if the resolution with respect to the variable x depends on other kinematic variables, the instrumental functions may depend on the model of the interaction adopted in the Monte Carlo program.

[3)] This can be used to eliminate $\hat{S}(x)$ from the terms of the expansion for which the coefficients are statistically insignificant, i.e., $\langle\tilde{g}^{(l)}|Y\rangle<3\sigma=3$.

[4)] In the exposition of this approach, we use the equations and notation employed earlier.

[5)] Note that for all methods except the dual method the behavior of $\|A_R^+-A_R\|_2^2$ as a function of α is similar to that of the function $N-\mathrm{Sp}\,\underline{A}_R$, since $\|A_R^+-A_R\|_2^2=N-2\mathrm{Sp}\,G\tilde{G}^{-1}+\mathrm{Sp}\,G\tilde{G}^{-}\to 0$, $\mathrm{Sp}\,A_R=\mathrm{Sp}\,G\tilde{G}^{-1}\to N$ as $\alpha\to 0$; in Fig. 6, we therefore give only $\mathrm{Sp}\,A_R$.

[6)]Note that the scheme of the correction-factor method presented here is Gold's method[17] of solving (69).

[7)]The expression "distance" in this case has a nominal nature, since $I(p,p_0)$ does not satisfy the symmetry condition $\rho(p,p_0)=\rho(p_0,p)$ for a distance. Kullback[32] himself calls $I(p,p_0)$ the mean information for the separation of p and p_0 with respect to ρ.

[8)]Here and in what follows, we use as the characteristic of the noise component its norm $\|F_\varepsilon\|^2$, which is equal to Sp Σ only in the case of the L_2 norm.

[9)]Here and in what follows, we assume for simplicity that $N^-=N^+$, that the binning on measurement of the cross sections $S_{NC}^{\pm}$ is the same ($A_n^+=A_n^-=A_n$), and $\tilde{G}^+=\tilde{G}^-=\tilde{G}$ (the regularization parameters in the estimates $\hat{S}_{NC}^+$ and $\hat{S}_{NC}^-$ are the same).

[1] V. P. Zhigunov, Preprint 81-5 [in Russian], IHEP, Serpukhov (1981); Nucl. Instrum. Methods **216**, 183 (1983).

[2] V. B. Anykeyev, A. A. Spiridonov, and V. P. Zhigunov, Nucl. Instrum. Methods **A303**, 350 (1991).

[3] E. A. Belogorlov and V. P. Zhigunov, Nucl. Instrum. Methods **A235**, 146 (1984).

[4] E. A. Belogorlov and V. P. Zhigunov, in *Third All-Union Conference on Shielding from Ionizing Radiations of Nuclear Technological Facilities, Tbilisi, 1981. Abstracts of Papers* [in Russian] (Tbilisi, 1981), p. 6.

[5] P. P. Zabreĭko *et al.*, *Integral Equations* [in Russian] (Noordhoff, Leyden, 1975) [Russ. original, Nauka, Moscow, 1975].

[6] M. Bertero, INFN/TC-88/2 19 (1988).

[7] Yu. P. Pyt'ev, *Mathematical Methods of Interpreting Experiments* [in Russian] (Vysshaya Shkola, Moscow, 1989).

[8] V. B. Anykeyev, A. A. Spiridonov, and V. P. Zhigunov, Preprint 89-15, IHEP, Serpukhov (1989).

[9] G. I. Marchuk, *Methods of Numerical Mathematics* (Springer-Verlag, Berlin, 1975); S. F. Gilyazov, *Methods of Solution of Linear Improperly Posed Problems* [in Russian] (Moscow State University, Moscow, 1979).

[10] A. N. Tikhonov and V. Ya. Arsenin, *Solutions of Ill-Posed Problems* (Halsted, New York, 1977) [Russ. original, 2nd ed., Nauka, Moscow, 1979].

[11] V. Blobel, in *Proc. of the 1984 CERN School of Computing, Alguablava, Spain*, 85-09 (CERN, 1985), p. 88.

[12] W. R. Burrus and V. V. Verbinsky, Nucl. Instrum. Methods **67**, 181 (1969).

[13] G. Backus and F. Gilbert, Philos. Trans. R. Soc. **766**, 123 (1970).

[14] V. A. Morozov, *Regular Methods of Solution of Improperly Posed Problems* [in Russian] (Nauka, Moscow, 1987) [English translation of earlier edition, *Methods for Solving Incorrectly Posed Problems* (Springer-Verlag, Berlin, 1984)].

[15] V. B. Anykeyev, A. A. Spiridonov, and V. P. Zhigunov, Nucl. Instrum. Methods **A322**, 280 (1992).

[16] P. Aarnio *et al.*, CERN-EP/90-19 (1990).

[17] R. Gold, ANL-6984 (1964).

[18] A. C. Benvenutti *et al.*, CERN-EP/89-07 (1989).

[19] A. Kondor, Nucl. Instrum. Methods **A216**, 183 (1983).

[20] H. N. Mülthei and B. Schorr, Nucl. Instrum. Methods **A257**, 371 (1984).

[21] D. A. Chombless and J. A. Broadway, Nucl. Instrum. Methods **179**, 563 (1981).

[22] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice Hall, New York, 1974).

[23] H. J. Behrend *et al.*, Z. Phys. C **44**, 63 (1989).

[24] S. W. Provencher, Comput. Phys. Commun. **27**, 213 (1982).

[25] B. W. Rust and W. R. Burrus, *Mathematical Programming and the Numerical Solution of Linear Equations* (New York, 1972).

[26] S. N. Sokolov and I. N. Silin, Preprint D810 [in Russian], JINR, Dubna (1961); I. N. Silin, FUMILI CERN Program Library, D510 (1971).

[27] L. M. Soroko, Fiz. Elem. Chastits At. Yadra **12**, 754 (1981) [Sov. J. Part. Nucl. **12**, 303 (1981)].

[28] V. P. Zhigunov *et al.*, Nucl. Instrum. Methods **A273**, 362 (1988).

[29] C. Fuglesang, Nucl. Instrum. Methods **A278**, 765 (1989).

[30] D. S. Sivia *et al.*, Nucl. Instrum. Methods **A290**, 492 (1990).

[31] A. Breakstone *et al.*, Z. Phys. C **40**, 207 (1989).

[32] S. Kullback, *Information Theory and Statistics* (Dover, New York, 1968) [Russ. transl., Nauka, Moscow, 1967].

[33] A. A. Borovikov, *Mathematical Statistics* [in Russian] (Nauka, Moscow, 1984); S. Kullback and R. A. Leibler, Ann. Statist. **22**, 79 (1951).

[34] A. Ya. Khinchin, *Mathematical Foundations of Statistical Mechanics* [in Russian] (GITTL, Moscow, 1943).

[35] V. B. Anykeyev and V. P. Zhigunov, Preprint 91-74, IHEP, Protvino (1991).

[36] V. B. Anykeyev and V. P. Zhigunov, Preprint 91-79, IHEP, Protvino (1991).

[37] F. Bergsma *et al.*, Phys. Lett. **123B**, 269 (1983).

[38] V. V. Ammosov, Z. U. Usubov, and V. P. Zhigunov, Nucl. Instrum. Methods **A295**, 224 (1990).

[39] V. B. Anykeyev, A. A. Spiridonov, and V. P. Zhigunov, *Proc. of the CHEP-92* (Annecy, France, 1992).

[40] M. Schmelling, CERN CN/93/3 (1993).

Translated by Julian B. Barbour