*Acta Cryst.* (1973). A**29**, 1

# Recursive Bayes Deconvolution in Physical Experiments

M. MENDES AND C. DE POLIGNAC

*Institut Laue–Langevin, Cédex* 156, 38 *Grenoble-Gare, France*

(*Received* 15 *December* 1971; *accepted* 26 *May* 1972)

In the analysis of measured spectra in physical experiments, smearing effects due to instrument resolution are present. Correcting for these effects can be critical owing to statistical disturbances and limited *a priori* information. A recursive algorithm has been produced from the Bayes estimation theory. Computational aspects are discussed in different simulated examples.

## 1. Introduction

In the analysis of measured physical spectra, smearing effects due to the instrumental resolution are generally present. This situation can be represented by the following mathematical model:

$$y(l)=\int_{-\infty}^{+\infty} f(l')s[g(l,l')]\,\mathrm{d}l'+v(l) \qquad (1)$$

where $y(l)$ is the measured value of the spectrum at the point $l$, $f(l)$ is the resolution function of the instrument, $s(l)$ is the true unknown and $v(l)$ is the statistical noise of the measurement. The function $g(l, l')$ takes different forms. For example, for the correction of the slit-height effect in slit-smeared small-angle scattering $g(l, l')=(l^2+l'^2)^{1/2}$. In the special case of convolution $g(l, l')$ is equal to $l-l'$, so that:

$$y(l)=\int_{-\infty}^{+\infty} f(l')\,s(l-l')\,\mathrm{d}l'+v(l) \qquad (2)$$

$$=\int_{-\infty}^{+\infty} f(l-l')\,s(l')\,\mathrm{d}l'+v(l)\,.$$

From a set of measurements $y$, we wish to calculate the true values of $s(l)$. This can be very difficult because:

(1) Non-negligible statistical noise $v(l)$ is always present.
(2) Only (relatively) few measurements are made.
(3) Often the specification of a physically reasonable model for the function $s(l)$ is not possible.
(4) The resolution function $f(l)$ is not exactly known, but usually obtained from measurements with statistical disturbances.

A first approach using a least-squares approximation is possible, with the advantage that it requires practically no knowledge of the statistical properties of disturbances (Points 1, 4).

General assumptions about the true spectrum can be made, for instance, using expansions by orthogonal functions (Point 3). This method has been widely used in control theory (Eykhoff, 1963) and applied in the treatment of smearing problems (Hossfeld, 1968).

These solutions can have severe oscillations with no physical significance, because of the few measurements (Point 2) and the errors involved (Points 1,3). Several authors have suggested different solutions. Weise (1968) added terms to the error function of the least-squares in order to damp these oscillations; Green (1969) used Fourier transforms and excluded the higher frequencies; Grosswendt (1971) smoothed the data using a simple averaging technique.

In terms of statistical estimation theory, these methods have no meaning, so that a distortion of the signal results. Wiener (1943) has given forms of optimal filters for smoothing data. Tournarie (1958) and Porteus (1962) have applied these results for deconvolution; unlike in least-squares techniques, the *a priori* information about the spectrum has to be introduced in a statistical way. This seems to be the best way of avoiding oscillations for small numbers of measurements. Nevertheless, the classical treatment with Fourier transforms, used by these authors, is very cumbersome and a solution for the general case [equation (1)] is not easily obtained. Kalman & Bucy (1961) have developed a Wiener filter for the sequential treatment of data directly in the domain of measurement. These equations are a special application of Bayes estimation methods and have, to our knowledge, not yet been applied in physical problems.

In § 2, some general ideas of estimation theory are summarized. In § 3, a general algorithm is developed, based on the Kalman filter. In § 4, computational aspects are discussed and some examples analysed.

## 2. Statistical estimation theory

Some results of statistical estimation are systematically used in our paper and the main points are summarized here. For a simple and detailed explanation see, for example, Sage & Melsa (1971).

### 2.1 *Bayes estimation procedure*

Consider the following chain of measurements:

$$\mathbf{y}(k)=f(\mathbf{m}, \mathbf{v}(k), k) \quad k=1, \ldots n$$

where $\mathbf{y}(k)$ is the $p\times 1$ vector of measurements at the

point $k$, $\mathbf{m}$ the unknown $r \times 1$ parameter vector to be estimated and $\mathbf{v}(k)$ represents the statistical disturbances. Generally several assumptions can be made:

(1) The statistical disturbances have a density $p(\mathbf{v})$;
(2) The unknown parameter vector $\mathbf{m}$ belongs to a known range $\mathbf{m} \in \mathscr{M}$. This can be described by an *a priori* density $p_0(\mathbf{m})$. Even in the case of deterministic parameters, it is convenient to introduce *a priori* information in a statistical way;
(3) In order to carry out an estimation, an error function $R(\mathbf{m}, \mathbf{m}_M)$ shall be introduced, where $\mathbf{m}_M$ is the parameter vector of a mathematical model. This function allows comparison of different estimations.

The Bayes estimator $\hat{\mathbf{m}}_B$ of $\mathbf{m}$ is obtained from

$$\underset{\mathbf{m}_M}{\text{Min}} [\mathscr{R}(\mathbf{m}_M)] \to \hat{\mathbf{m}}_B \tag{4}$$

where $\mathscr{R}(\mathbf{m}_M)$ is the risk function

$$\mathscr{R}(\mathbf{m}_M) := \int_{\mathscr{M}} \int_{\mathscr{J}} R(\mathbf{m}, \mathbf{m}_M)\, p(\mathbf{m}, y)\, .\, \mathrm{d}\mathbf{m}\, .\, \mathrm{d}\mathbf{y} \tag{5}$$

where $\mathbf{y}$ is the $(n \times p) \times 1$ vector of all $\mathbf{y}(k)$

$$p(\mathbf{m}, \mathbf{y}) = p_0(\mathbf{y})\, .\, p(\mathbf{m}/\mathbf{y})\, . \tag{6}$$

This results in

$$\underset{\mathbf{m}_M}{\text{Min}} [\mathscr{R}(\mathbf{m}_M)] \to \left\{ \int_{\mathscr{J}} p_0(\mathbf{y}) \left[ \underset{\mathbf{m}_M}{\text{Min}} \int_{\mathscr{M}} R(\mathbf{m}, \mathbf{m}_M) p(\mathbf{m}/\mathbf{y}) \mathrm{d}\mathbf{m} \right] \mathrm{d}\mathbf{y} \right\} \tag{7}$$

and for different error functions $R(\mathbf{m}, \mathbf{m}_M)$, different Bayes solutions can be obtained. Only two important cases will be considered here.

Case 1: *minimum variance Bayes estimation*

Let

$$R(\mathbf{m}, \mathbf{m}_M) := \|\mathbf{m} - \mathbf{m}_M\|^2_{\mathbf{Q}}. \tag{8}$$

Then:

$$\hat{\mathbf{m}}_B = \varepsilon\{\mathbf{m}/\mathbf{y}\} := \int_{\mathscr{M}} \mathbf{m} p(\mathbf{m}/\mathbf{y}) \mathrm{d}\mathbf{m} \tag{9}$$

which means the estimator is equal to the mean value of the *a posteriori* density $p(\mathbf{m}/\mathbf{y})$, and is independent of $\mathbf{Q}$.

From equation (7), it follows that

$$\text{Min } R(\mathbf{m}, \mathbf{m}_M) = R(\mathbf{m}, \hat{\mathbf{m}}_B) = \varepsilon_{\mathbf{y}}\{\text{Trace } (\mathbf{M}\, .\, \mathbf{Q})\} \tag{10}$$

where $\mathbf{M}$ is the covariance matrix of the estimator $\hat{\mathbf{m}}_B$.

Case 2: *maximum a posteriori Bayes estimation*

Suppose that the function $R(\mathbf{m}, \mathbf{m}_M)$ takes the form of Fig. 1. It is easily demonstrated that equation (7) leads to:

$$\underset{\mathbf{m}_M}{\text{Min}} [\mathscr{R}(\mathbf{m}_M)] \to \underset{\mathbf{m}}{\text{Max}} [p(\mathbf{m}/\mathbf{y})] \to \hat{\mathbf{m}}_{\text{MAP}}\, . \tag{11}$$

When it is possible to calculate the *a posteriori* density $p(\mathbf{m}/\mathbf{y})$ (density of $\mathbf{m}$ after measuring $\mathbf{y}$), $\hat{\mathbf{m}}_{\text{MAP}}$ is easier to find than the minimum variance estimator. Applying the Bayes rule:

$$p(\mathbf{m}/\mathbf{y}) = \frac{p(\mathbf{y}/\mathbf{m})\, .\, p_0(\mathbf{m})}{p_0(\mathbf{y})}, \tag{12}$$

it follows from equation (11) that

$$\underset{\mathbf{m}_M}{\text{Min}} [\mathscr{R}(\mathbf{m}_M)] \to \underset{\mathbf{m}}{\text{Max}} [\ln p(\mathbf{y}/\mathbf{m}) + \ln p_0(\mathbf{m})] \tag{13}$$

because the logarithm ln is a monotonic function. In the case where no *a priori* information exists, equation (13) reduces to the well known method of maximum likehood.

2.2 *Linear measurement scheme*

Consider now the linear system of equations:

$$\mathbf{y} = \mathbf{H}\, .\, \mathbf{m} + \mathbf{v} \tag{14}$$

where $\mathbf{H}$ is a $(n \times p) \times r$ matrix. Moreover let $\mathbf{v}$ and $\mathbf{m}$ have the *a priori* Gaussian densities $N(\mathbf{0}, \mathbf{V})$, and $N[\bar{\mathbf{m}}(0), \mathbf{M}(0)]$. In this case, the minimum variance estimator and the maximum *a posteriori* estimator are identical. In order to obtain the maximum *a posteriori* estimator $\hat{\mathbf{m}}_{\text{MAP}}$, $p(\mathbf{y}/\mathbf{m})$ must be first calculated. It results from equation (14) that

$$p(\mathbf{y}/\mathbf{m}) \equiv p(\mathbf{v} = \mathbf{y} - \mathbf{H}\, .\, \mathbf{m}) = (2\pi)^{-n/2}\, .\, [\det(\mathbf{V})]^{-1/2} \times \exp\{-0{\cdot}5 \|\mathbf{y} - \mathbf{H}\, .\, \mathbf{m}\|^2 \mathbf{V}^{-1}\}\, . \tag{15}$$

After some simple manipulation, the Gaussian density $N(\bar{\mathbf{m}}, \mathbf{M})$ results for $p(\mathbf{m}/\mathbf{y})$ with

$$\hat{\mathbf{m}}_{\text{MAP}} = \bar{\mathbf{m}} = \mathbf{M}\, .\, [\mathbf{H}^T\, .\, \mathbf{V}^{-1}\, .\, \mathbf{y} + \mathbf{M}^{-1}(0) \bar{\mathbf{m}}(0)] \tag{16a}$$

$$\mathbf{M} = [\mathbf{M}^{-1}(0) + \mathbf{H}^T \mathbf{V}^{-1} \mathbf{H}]^{-1}\, . \tag{16b}$$

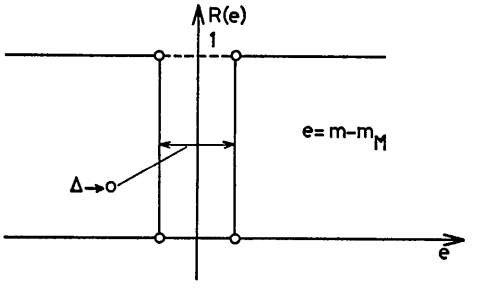



Fig. 1. Error function for the maximum *a posteriori* Bayes estimation.

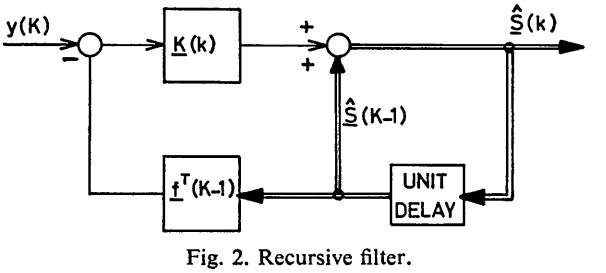



Fig. 2. Recursive filter.

In the case where no *a priori* information is known about $\mathbf{m}$, $M(0) \to \infty$ and equation (16) yields

$$\hat{\mathbf{m}}_M = \mathbf{M}\mathbf{H}^T\mathbf{V}^{-1}\mathbf{y}$$
$$\mathbf{M} = (\mathbf{H}^T\mathbf{V}^{-1}\mathbf{H})^{-1} . \quad (17)$$

This is the Markov solution, which can also be derived from the *maximum-likehood* method. A special form of this builds the 'least-squares' solution with $\mathbf{V} \equiv \mathbf{E}$,

$$\hat{\mathbf{m}}_{L.S} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y} . \quad (18)$$

2.3 *Sequential algorithms*

Global methods are normally very cumbersome, especially when a great number of measurements or unknown parameters are considered. A sequential treatment of the measurements can be advisable in these cases.

Suppose that, at the point $k$ ($L$ being the spacing of consecutive measurements), the vector $\mathbf{y}(k)$ is measured and that, with the help of the preceding information $[\mathbf{y}(1), \ldots, \mathbf{y}(k-1)]$, an estimate $\hat{\mathbf{m}}$ $(k-1)$, with a covariance matrix $\mathbf{M}(k-1)$ has been obtained. In order to correct $\hat{\mathbf{m}}(k-1)$ considering the information of the new measurement $\mathbf{y}(k)$, let us introduce the *a posteriori* density:

$$p_k(\mathbf{m}) : = p[\mathbf{m}/\mathbf{Y}(1 \to k)] \quad (19)$$

with

$$\mathbf{Y}(1 \to k) : = [\mathbf{y}(1), \ldots, \mathbf{y}(k)] . \quad (20)$$

Now

$$p[\mathbf{m}, \mathbf{y}(k)/\mathbf{Y}(1 \to k)]$$
$$= p[\mathbf{m}/\mathbf{Y}(1 \to k-1)] \cdot p[\mathbf{y}(k)/\mathbf{Y}(1 \to k-1), \mathbf{m}]$$
$$= p[\mathbf{m}/\mathbf{Y}(1 \to k-1), \mathbf{y}(k)] . p[\mathbf{y}(k)/\mathbf{Y}(1 \to k-1)]$$

which means that

$$p_k(\mathbf{m}) = \frac{p_{k-1}(\mathbf{m}) . p[\mathbf{y}(k)/\mathbf{Y}(1 \to k-1), \mathbf{m}]}{p[\mathbf{y}(k)/\mathbf{Y}(1 \to k-1)] .} \quad (21)$$

This fundamental equation gives the density $p_k(\mathbf{m})$ at the point $k$ as a function of the density $p_{k-1}(\mathbf{m})$ at the point $k-1$ and of the new measurement $\mathbf{y}(k)$. This is the counterpart of equation (12) in the global case.

In general, equation (21) can be very difficult to manage as the characterization of a density requires an infinity of its moments. In the linear Gaussian case alone is the solution simple. The measurement chain can be written:

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{m} + \mathbf{v}(k) \quad (22)$$

for $k = 1, \ldots n$, where $\mathbf{H}(k)$ is a $p \times r$ matrix and

$$p[\mathbf{v}(k)] = N[\mathbf{0}, \mathbf{V}(k)] \text{ and } p_0(\mathbf{m}) = N[\bar{\mathbf{m}}(0), \mathbf{M}(0)] .$$

It can be proved by induction that $p_k(\mathbf{m})$ is also Gaussian. As in equation (15), the densities $p[\mathbf{y}(k)/\mathbf{Y}(1 \to k-1), \mathbf{m}]$ and $p[\mathbf{y}(k)/\mathbf{Y}(1 \to k-1)]$ can be calculated from equation (22); the introduction of

$$p_i(\mathbf{m}) = N[\bar{\mathbf{m}}(i), \mathbf{M}(i)]$$

results, after some manipulations, in:

$$\hat{\mathbf{m}}(k) = \hat{\mathbf{m}}(k)$$
$$= \hat{\mathbf{m}}(k-1) + \mathbf{K}(k)[\mathbf{y}(k) - \mathbf{H}(k) . \hat{\mathbf{m}}(k-1)] \quad (23a)$$

$$\mathbf{K}(k) = \mathbf{M}(k-1)\mathbf{H}^T(k)[\mathbf{V}(k)$$
$$+ \mathbf{H}(k) . \mathbf{M}(k-1) . \mathbf{H}^T(k)]^{-1} \quad (23b)$$

$$\mathbf{M}(k) = [\mathbf{E} - \mathbf{K}(k) . \mathbf{H}(k)] . \mathbf{M}(k-1) . \quad (23c)$$

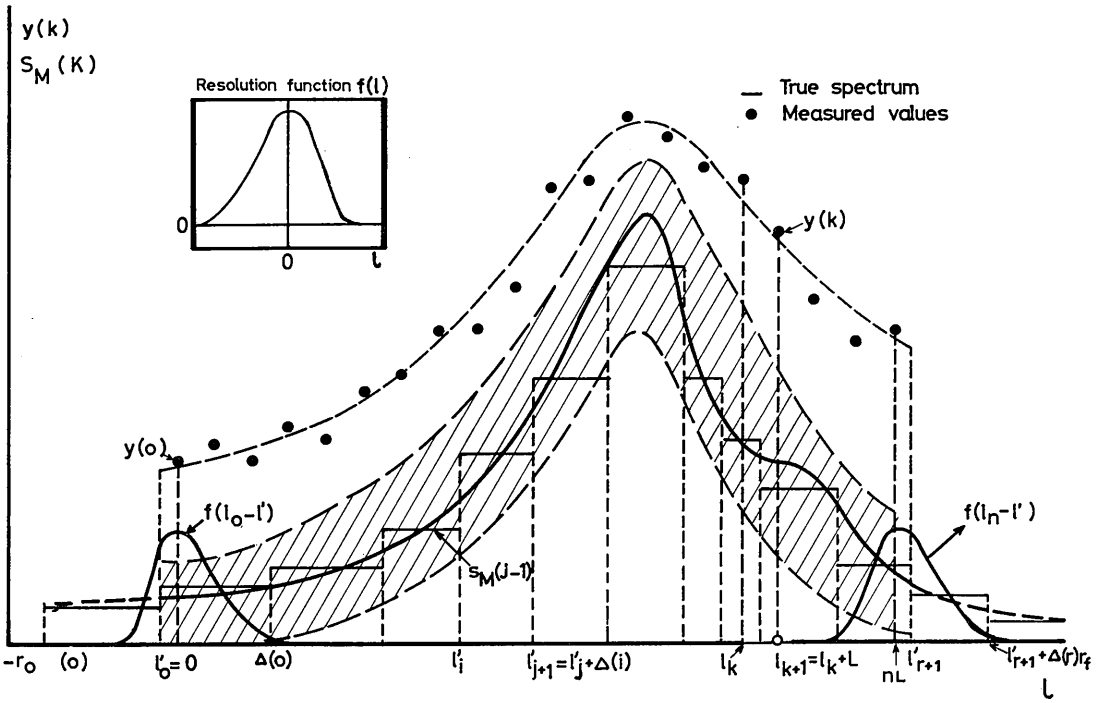



Fig. 3. Deconvolution by staircase approximation.

These equations are a special form of the *Kalman Filter*. For the first measurement $\mathbf{y}(1)$, they start with $\bar{\mathbf{m}}(0)$, $\mathbf{M}(0)$. A simple representation can be found in Fig. 2. It is important to see that, when no *a priori* information about $\mathbf{m}$ is known [$\mathbf{M}(0) \to \infty$], the Markov solution is obtained with $\mathbf{V}(k) \equiv \mathbf{E}$.

### 3. Deconvolution method

Now, based on the preceding considerations, a general method for deconvolution is developed.

3.1 *Discretization of the convolution integral*

If no physical model can be considered, a linear representation of equation (2) can be obtained by using a staircase approximation for $s(l)$.

For mathematical convenience, consider the measured points $y(l_k)$ centred on:

$$l_k := \frac{L}{2} + (k-1)L \tag{24}$$

with $k = 1, \ldots n$. $L$ is the distance between consecutive points (normally but not necessary constant). Considering steps of variable length (Fig. 3) for a better representation of $s(l)$, there results

$$s(l') = s_M(i) = \text{constant} \quad \text{for} \quad l'_i \le l' \le l'_{i+1} \tag{25}$$

with $i = 1, \ldots r$.

Considering $n(i)$ measured points in each interval

$$l'_{i+1} - l'_i := \Delta(i) = n(i) \,.\, L \,.$$
$$l'_0 := 0 \,.$$

The sequence $\{n_i\}$ can be selected to a first approximation from the measured spectrum and in order that $r \ll n$, it is mostly convenient to put $n(i) \ge n_{\min}$. In many cases, $\Delta(i)$ can be constant $\Delta(i) = \Delta$.

As can be seen from Fig. 3, the resolution function $f(l')$ has to be inverted to give $f(-l')$ and translated by $l_k$, to give $f(l_k - l')$ for each measured point $y(l_k)$. To take into account all the measurements, the spectrum $s_M(l')$ must be extrapolated for $l' \le 0$ and $l' \ge l'_{r+1}$. For simplicity, let us introduce $r_0$ steps of length $\Delta(0)$ on the left and $r_f$ steps of length $\Delta(r)$ on the right. The new partition of intervals is:

$$\{l_i^x\} := \{-r_0\Delta(0), \ldots, -\Delta(0), 0, l'_1, \ldots, l'_{r+1}, l'_{r+1} + \Delta(r), \ldots, l'_{r+1} + r_f\Delta(r)\} \,. \tag{27}$$

The corresponding unknown values of $s_M(i)$ are

$$\mathbf{s}_M^T := \{s_M(-r_0), \ldots, s_M(0), s_M(1), \ldots, s_M(r), \ldots, s_M(r + r_f)\} \,. \tag{28}$$

Introducing these relations in equation (2), we obtain

$$y(k) := y(l_k) = \sum_{i=-r_0}^{r+r_f} \left[ s_M(i) \int_{l_i^*}^{l_{i+1}^*} f(l_k - l') \,\mathrm{d}l' \right] + v(l_k) \tag{29}$$

with

$$f_{ki} := \int_{l_i^*}^{l_{i+1}^*} f(l_k - l') \,\mathrm{d}l' \tag{30}$$

$$\mathbf{f}^T(k) := \{f_{ki}\} \,. \tag{31}$$

This results finally in

$$\boxed{y(k) = \mathbf{f}^T(k) \,.\, \mathbf{s}_M + v(k)} \,. \tag{32}$$

If an unknown background noise $V_B(l)$ is present, the measurement model (2) must be augmented by this term. In the case of linear noise

$$V_B(l) = Al + B \,. \tag{33}$$

The unknown parameters $A$ and $B$ may be estimated by transforming equation (32). With

$$\mathbf{f}^{T*}(\mathbf{k}) = \left[ f^T(k) \;\; \frac{L}{2} + (k-1)L \;\; 1 \right]$$
$$\mathbf{s}_M^{T*} = [\mathbf{s}_M^T \;\; A \;\; B] \tag{34}$$

which results in

$$y(k) = \mathbf{f}^{T*}(k)\mathbf{s}_M^* + v(k) \quad k = 1, \ldots n \tag{35}$$
$$= y_M(k) + v(k)$$

with

$$y_M(k) := \mathbf{f}^{T*}(k) \,.\, \mathbf{s}_M^* \,. \tag{36}$$

In the sequential algorithm (23), this form of the chain of measurements is immediately used with $\mathbf{H}(k)$ replaced by $\mathbf{f}^{T*}(k)$. In the classical global method (16), the measurements are gathered in the matrix form.

$$\mathbf{y} = \mathbf{F}^* \,.\, \mathbf{s}_M^* + \mathbf{v} \tag{37}$$

with

$$\mathbf{F}^{T*} := [\mathbf{f}^*(1) \;\; \mathbf{f}^*(2) \;\; \ldots \;\; \mathbf{f}^*(n)] \,. \tag{38}$$

3.2 *Statistical characteristics of unknowns and disturbances*

*Measurement noise*

The set of measurements $\mathbf{y}$ is obtained by counting the number of neutrons that arrive in a certain time in each detector. In most cases neighbouring measurements are uncorrelated and the probability of counting $y(k)$ particles is

$$p[y(k)] = [y_M(k)]^{y(k)} \,.\, \frac{\exp\,[-y_M(k)]}{[y(k)]!} \tag{39}$$

which means that $y(k)$ is Poisson distributed with a mean value $y_M(k)$ given by equation (39). In order to utilize the developed algorithms a Gaussian approximation must be considered. For $y(k) > 200$ it is well known that:

$$p[y(k)] \simeq N[y_M(k), y_M(k)] \,. \tag{40}$$

Because $y_M(k)$ is unknown, $y(k)$ is used for the mean value and variance. Then

$$p(\mathbf{v}) = N(\mathbf{O}, \mathbf{V}) \tag{41}$$

with

$$\mathbf{V} = \mathrm{DIAG}[y(1), \ldots y(n)] . \tag{42}$$

*A priori knowledge about s*

The sequential algorithms start with $\bar{\mathbf{s}}(0)$, $\mathbf{S}(0)$, when it is supposed that the *a priori* density of $\mathbf{s}(0)$ is Gaussian:

$$p[\mathbf{s}(0)] = N[\bar{\mathbf{s}}(0), \mathbf{S}(0)] . \tag{43}$$

There is a simple way of determining this density. Suppose (Fig. 5) that at least a range is known where the true spectrum must be located. For each interval $[l_i^*, l_{i+1}^*]$ a mean value $\bar{s}_i(0)$ is considered with a variance $\sigma_i^2(0) \simeq 2 . b(i)$ where $b(i)$ is the estimated width in the domain. In this way, a diagonal matrix $\mathbf{S}(0)$ can be constructed.

*Resolution function errors*

In most physical problems, the resolution function is not given analytically. In neutron scattering, for example, one measures it repeatedly with the same experimental instrument. Each detector then gives the integral of the resolution function in its range. In this paper, we suppose that the measurements have lasted long enough to eliminate strong oscillations. This can be done because the resolution function is measured once for all. A special consideration of propagation of errors is then not necessary and will not be considered here.

3.3 *Possibilities of on-line implementation*

Some of the components of the filter equations (23) such as the matrix $\mathbf{F}$ or its lines $\mathbf{f}^T(k)$ can be calculated at the beginning of the experiment and stored in an auxiliary memory of the computer. Otherwise suboptimal algorithms will be discussed, which can considerably reduce the time of computation, so that on-line implementation becomes possible.

(1) For approximately constant variances of measured points (for example with relatively high background noise), the covariance matrix $\mathbf{V}$ will be equal to $\sigma^2 \mathbf{E}$ so that the time-consuming recursion, [equation (23*c*)] of $\mathbf{S}(k)$ is independent of the measurements and can be treated off-line, once for all. The gain filter $\mathbf{K}(k)$ [equation (23*b*)] will also be calculated and stored in the auxiliary memory in this preliminary phase.

(2) When this kind of approximation is not allowed [very different values of $y(k)$] an on-line calculation can also be considered, when the matrices $\mathbf{S}(k)$ are constrained to be diagonal. From the recursion [equation (23*c*)], it follows for the diagonal elements that

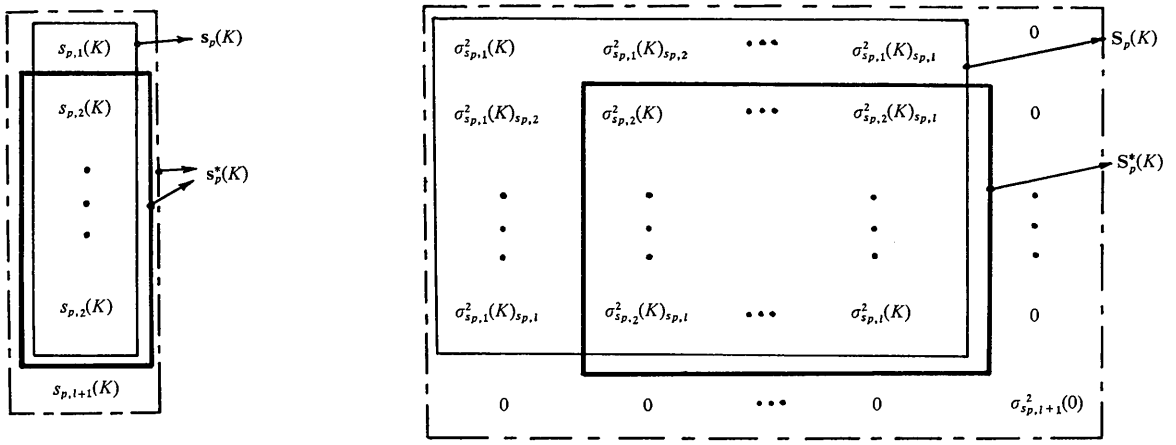
$$\sigma_{s_j}^2(k) = [1 - \alpha^{-1} f_j^{*4}(k) . \sigma_{s_j}^2(k-1)] \sigma_{s_j}^2(k-1) \tag{44}$$

with

$$\alpha = y(k) + \sum_{\nu=1}^{r} f_\nu^{*2}(k) . \sigma_{s_\nu}^2(k-1) \tag{45}$$

for $j = 1, \ldots r$.

(3) Because normally the resolution function has a limited width, only a finite number of estimators $\hat{s}_j$ are




——— case (*b*)

— - — case (*c*)

Fig. 4. Construction of intermediate values for the sub-optimal deconvolution method.

correlated; this means that in order to use the information of each measurement $y(k)$, only a vector $\hat{\mathbf{s}}_p(k)$, part of the whole vector $\hat{\mathbf{s}}(k)$, has to be considered.

From the mechanism of convolution, which corresponds to a continuous translation from $l_k$ to $l_{k+1}$ of the inverted resolution function, the following different situations can occur: (*a*) the vectors $\hat{\mathbf{s}}_p(k)$ and $\hat{\mathbf{s}}_p(k+1)$ have the same components; (*b*) the first component of $\hat{\mathbf{s}}_p(k)$ is deleted; (*c*) a new component has to be added at the end of $\hat{\mathbf{s}}_p(k)$; (*d*) the two operations (*b*) and (*c*) are performed simultaneously. In cases (*b*), (*c*) and (*d*) an intermediate vector $\hat{\mathbf{s}}_p^*(k)$ results, which will be utilized as starting value in the recursion, equation (23*a*). Simultaneously, a starting covariance matrix $\mathbf{S}_p^*(k)$ has to be constructed from $\mathbf{S}_p(k)$ (Fig. 4): in (*b*) the first line and a column of $\mathbf{S}_p(\mathbf{k})$ are deleted, in (*c*), a new last column and line are added, with elements equal to 0 except the diagonal element which is set to $\sigma_{s_p}^2, {}^{(0)}_{l+2}$ (the *a priori* variance value). Finally, the case (*d*) is a combination of (*b*) and (*c*). The following scheme results from these considerations:

| stage $k$ | stage $(k+1)$ |
|---|---|
| $\hat{\mathbf{s}}_p(k)$; $\mathbf{S}_p(k) \rightarrow \hat{\mathbf{s}}_p^*(k)$; (Fig. 4) | $\mathbf{S}_p^*(k) \rightarrow \hat{\mathbf{s}}_p(k+1)$; $\mathbf{S}_p(k+1)$ equation (23). |

## 4. Computational aspects

### 4.1 *Structure of programs*

A general program of deconvolution has been derived. As can be seen on the flow-chart (Fig. 5), this program has the following parts:

(1) The raw data (measured points and resolution function) are read. Because the measured data represent integrals of the corresponding function in the range of each detector, some corrections may be needed. For example the $y(l)$ must be divided by $L$. To avoid this, the same scale is used for $f(l)$ and $s(l)$, for example $L=1$.

(2) The number of points $\{n_i\}$ per interval is read, so that the computation of all the $l_i'$ and then $l_i^*$ can be carried out.

(3) The initialization of $\bar{\mathbf{s}}(0)$ and $\mathbf{S}(0)$ is worked out by means of the method discussed in equation (3).

(4) Each point $y(k)$ must be treated with the help of recursive equations (23); a raw vector $\mathbf{f}^T{}^*(k)$ must be calculated. The computation of $f_{kj}^*$ is easily done because the raw data already give integrals of the resolution function; one centres $f(-l)$ on the point $L/2+(k-1)\,L$ and sums the ordinates of $f(-l)$ in each of the $r_0+r+r_f$ intervals $[l_j^*, l_{j+1}^*]$.

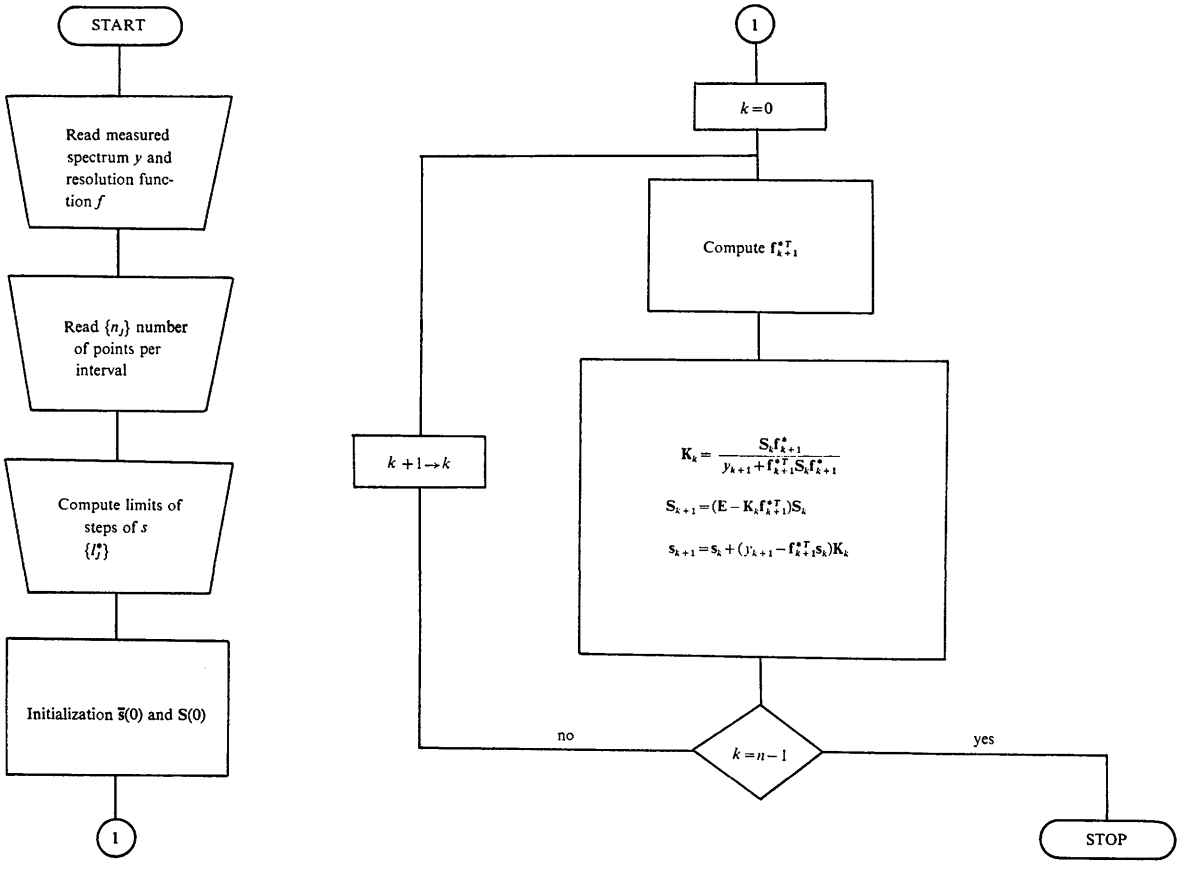



Fig. 5. Flow-chart of programs.

4.2 *Simulation examples*

In order to test the efficiency of the algorithm, raw data have been simulated with the help of a convolution procedure. For simplicity, the sum of three Gaussian curves has been convoluted with a Gaussian resolution function. A random generator of Gaussian distributed numbers has been utilized [equation (40)].

The program written in FORTRAN IV(G) has run under the time-sharing system CP-CMS, on the computer 360/67 IBM of Grenoble University in a virtual machine of 256K bytes. The indicated execution times, including a plot of the results, are the equivalent CPU time on a standard computer 360/65 IBM. The size of the program designed to treat up to 300 measured points with the estimation of a maximum of 60 parameters is about 70 k bytes.

*Case* 1 (*Fig.* 6)

Here from the measured data, a fairly good idea of the line shape can be made; this means that there is enough *a priori* information about the unknown spectrum to give good results. Nonetheless, the third line could only just be detected.

The estimation of 23 values of the spectrum and of the constant noise, from 200 measured points in the range [0,80], with a resolution function given by 75 points, takes about 120 sec.

*Case* 2 (*Fig.* 7)

Three very sharp lines have been convoluted with a broad resolution function, so that no good *a priori* assumption can be made from the measured data. The initialization will be represented by the hachured zone.

With 200 measured points, a choice of intervals may be the following: the rightmost and leftmost intervals of the unknown spectrum contain 50 measured points; the others are chosen, to a first approximation, as the same with 5 points. From the results, it is suspected that three lines are present (curve 1). Another deconvolution with this new information confirms this assumption (curve 2).

The estimation of 24 values of the spectrum and of the constant noise, from 200 measured points in the range [0,60] with a resolution function given by 101 points takes about 80 sec.

*Remarks*

The algorithm works well. Nevertheless, care must be taken for the initialization of some of its parameters:

(1) The choice of the *a priori* density $N[\bar{s}(0), S(0)]$ can be critical. When, for example, a very optimistic (but false) hypothesis is made, *i.e.* small variances combined with false mean values, the solutions can be very biased. In Fig. 7 it is then better to consider large values for the *a priori* variances.

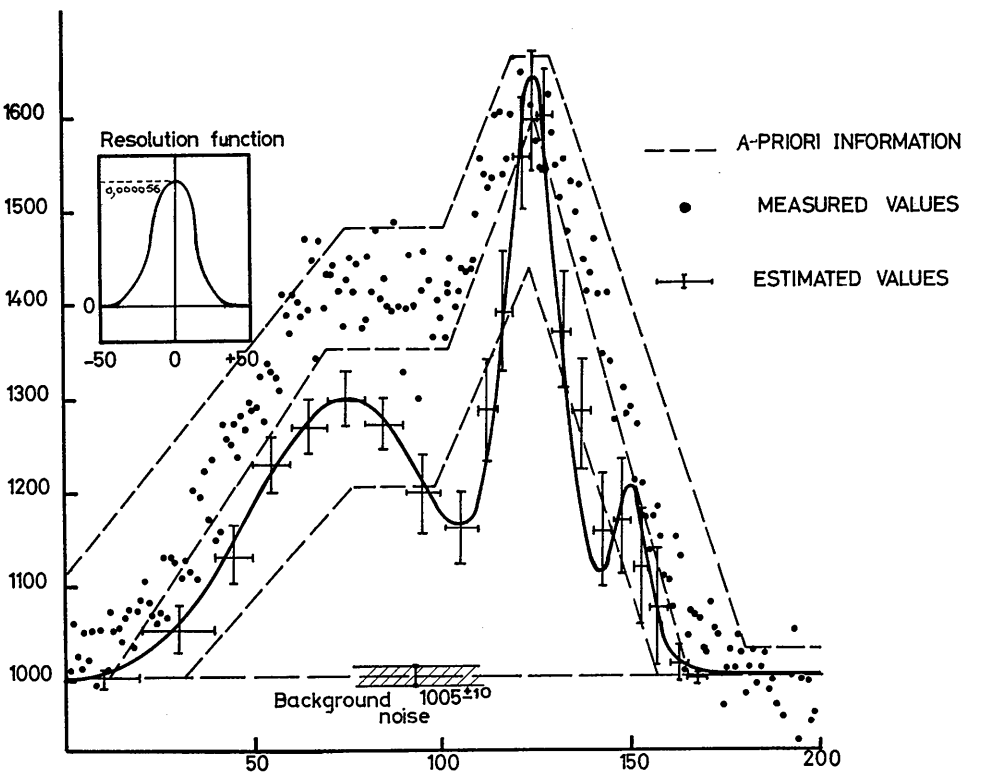



Fig. 6. Simulation example.

(2) For some choices of intervals, the staircase model can give a bad approximation; this can be corrected by a better choice of interval lengths, after a first deconvolution.

With a time-sharing system such as CP.CMS the modification of the parameters can be made easily. It seems to us that this kind of algorithm can ideally be realized with the help of interaction display techniques.

## 5. Applications

The deconvolution method can be applied in several fields of physical experiments:
- analysis of $\gamma$-ray spectra;
- determination of phonon line shapes;
- deconvolution of neutron time-of-flight data;
- correction of slit effects in small-angle scattering;
- analysis of quasi-elastic scattering data.

Fig. 8 shows the results of the deconvolution method applied to neutron time-of-flight data from an experiment on liquid sulphur. The resolution function has the typical asymmetric shape of a Be-filtered neutron spectrum obtained from a chopper. The measured spectrum is peaked at zero energy transfer ($\lambda_1 \simeq 4$ Å) and is expected to be the convolution of the resolution function with the following shape:

$$s(\lambda) = \frac{c}{\lambda^4} \exp g(\lambda) \,.\, S(\lambda) \qquad (46)$$

where $S(\lambda)$ is symmetrical and $g(\lambda)$ known.

From Fig. 8, it results that 4·06 [Å] $\leq \lambda_1 \leq$ 4·11 [Å] and a second peak at 3·58 [Å] $\leq \lambda_2 \leq$ 3·74 [Å] must be present. Because of the symmetry of $S(\lambda)$, a third peak on the left should also be observed. This could not be convincingly confirmed because of the relatively small amplitude of the peak, and because of bad statistics.

By taking the model [equation (46)] directly into account in the deconvolution, better results should be obtained. This can be done in the following way:

$$y_M(\lambda) = \int_{-\infty}^{\infty} f(\lambda-\lambda') s(\lambda') \mathrm{d}\lambda' = \int_{-\infty}^{\infty} f(\lambda-\lambda') \,.\, \frac{c}{\lambda'^4}$$

$$\exp g(\lambda') S(\lambda') \mathrm{d}\lambda' = \int_{-\infty}^{\infty} f^*(\lambda, \lambda') S(\lambda') \mathrm{d}\lambda' \qquad (47)$$

with

$$f^*(\lambda, \lambda') := f(\lambda-\lambda') \,.\, \frac{c}{\lambda'^4} \exp g(\lambda') \,.$$

This problem takes the general form of equation (1).

## 6. Conclusion

An algorithm for deconvolution of neutron scattering data has been developed with the help of Bayes estimation. The equations obtained turn out to be a special case of the Kalman filter, and because of their recursive character, can be applied in on-line situations. The *a priori* information has been utilized in a simple way and allows for a solution even when a small number of measurements is not sufficient for the least-squares application.

The staircase approximation has the advantage of simplicity. Nevertheless, when the statistics are bad, a certain number of measured points per interval may be needed, which contradicts the convenience of short in-

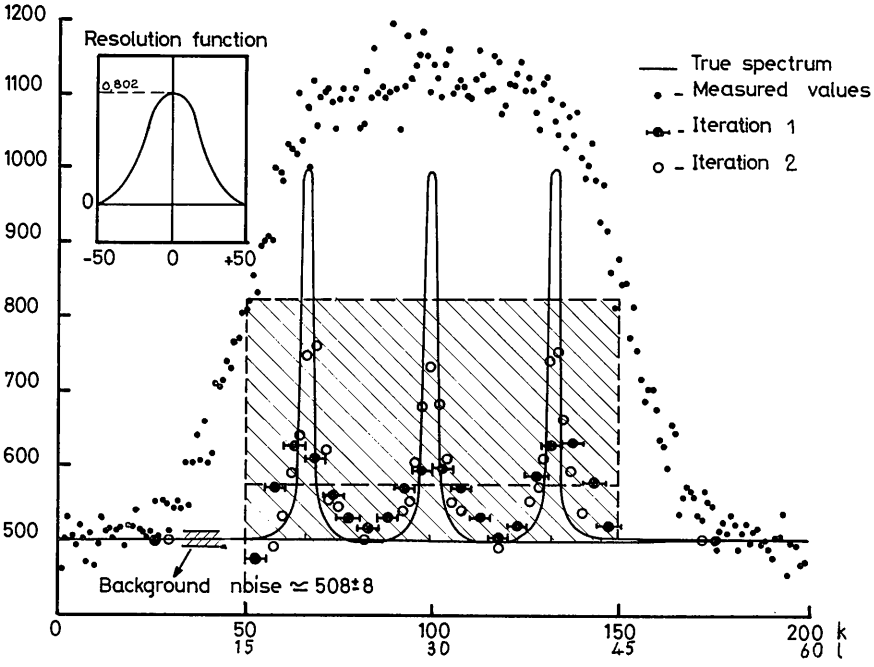



Fig. 7. Simulation example 2.

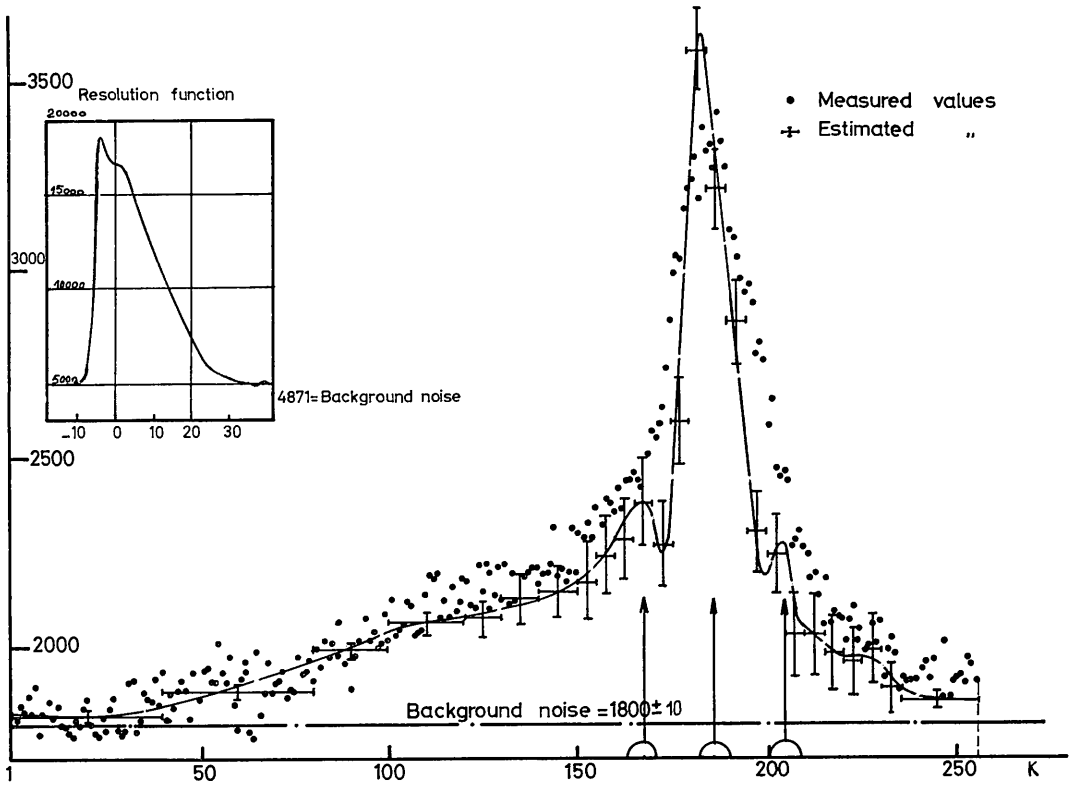



Fig. 8. Deconvolution of neutron time-of-flight data.

tervals for good representation; in this case, other models (straight lines, parabolas, polynomials) can be more suitable. Moreover, the choice of the *a priori* densities is often decisive and special care must be taken so that, when possible, systematic methods are preferable.

A treatment of equation (1) will be considered in a forthcoming publication.

The authors are very grateful to Professor Maier-Leibnitz for encouragement in this work and to several physicists of the Institute Laue–Langevin for interesting discussions and suggestions.

## Notation

$p(\mathbf{x})$: probability density of random vector $\mathbf{x}$.
$p_0(\mathbf{x})$: *a priori* probability density of random vector $\mathbf{x}$.
$p(\mathbf{x}, \mathbf{y})$: joint density of random vectors $\mathbf{x}$ and $\mathbf{y}$.
$p(\mathbf{x}/\mathbf{y})$: conditional density of random vector $\mathbf{x}$ given $\mathbf{y}$.
$\bar{\mathbf{x}} = \varepsilon\{\mathbf{x}\} = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x}$: expectation (mean value) of random vector $\mathbf{x}$.
$\bar{\mathbf{X}} = \varepsilon\{(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T\}$: covariance matrix of random vector $\mathbf{x}$.
$N(\bar{\mathbf{x}}, \bar{\mathbf{X}})$: Gaussian (normal) density of $\mathbf{x}$ with mean value $\bar{\mathbf{x}}$ and covariance matrix $\bar{\mathbf{X}}$.
$\varepsilon\{R\} = \int R p(\mathbf{y}) d\mathbf{y}$: expectation of $R$.
$\varepsilon\{\mathbf{x}/\mathbf{y}\} = \int \mathbf{x} p(\mathbf{x}/\mathbf{y}) d\mathbf{x}$: conditional expectation of $\mathbf{x}$ given $\mathbf{y}$.
$\mathbf{y}(k)$: $p \times 1$ measurement vector at the point $k$.
$\mathbf{y}$: $(p \times n) \times 1$ total measurement vector.
$\mathbf{Y}(1 \to k)$: $p \times k$ matrix of the first $k$ measurement vectors $\mathbf{y}(i)$.
$\mathbf{H}(k)$: partial $p \times r$ matrix in the measurement chain.
$\mathbf{H}$: total $(n \times p) \times r$ matrix in the measurement chain.
$\hat{\mathbf{m}}(y)$: $r \times 1$ estimate vector of the unknown $\mathbf{m}$.
$\bar{\mathbf{m}}(0)$: *a priori* mean value of $\hat{\mathbf{m}}$.
$\mathbf{M}(0)$: *a priori* covariance matrix of $\hat{\mathbf{m}}$.
$p_k(\mathbf{m})$ [equation (19)]: *a posteriori* density of vector $\mathbf{m}$ given the measurement matrix $\mathbf{Y}(1 \to \mathbf{k})$.
$\hat{\mathbf{m}}(k) = \bar{\mathbf{m}}(k)$: estimate at the point $k$ given $\mathbf{Y}(1 \to k)$.
$\mathbf{M}(k)$: covariance of estimate vector at the point $k$.
$\mathbf{v}$: vector of statistical disturbances with mean value $\mathbf{0}$.

## References

EYKHOFF, P. (1963). *IEEE Trans. Autom. Control*, AC**8**, 347–357.
GREEN, D. W. (1969). *Nucl. Instrum. Methods*, **76**, 349–356.
GROSSWENDT, B. (1971). *Nucl. Instrum. Methods*, **93**, 461–472.
HOSSFELD, F. (1968). *Acta Cryst.* A**24**, 643–650.
KALMAN, R. & BUCY, R. S. (1961). *Trans. ASME, Ser. D, J. Basic Eng.* **83**, 95–108.
PORTEUS, J. O. (1962). *J. Appl. Phys.* **33**, 700–707.
SAGE, A. P. & MELSA, J. L. (1971). *Estimation Theory with Applications & Communications and Control*. New York: McGraw-Hill.
TOURNARIE, M. (1958). *Bull. Soc. Franç. Minér. Crist.* **81**, 278–286.
WEISE, K. (1968). *Nucl. Instrum. Methods*, **65**, 189–194.
WIENER, N. (1943). *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*. New York: Wiley.