JOURNAL OF APPROXIMATION THEORY 73, 93–105 (1993)

Optimal Approximation of Linear Operators Based on Noisy Data on Functionals

L. PLASKOTA

Institute of Applied Mathematics and Mechanics, University of Warsaw, 02-097 Warsaw, Poland

Communicated by T. J. Rivlin

Received February 6, 1991; accepted in revised form December 31, 1991

For a linear operator $S: F \to G$, where F is a Banach space and G is a Hilbert space, we pose and solve the problem of approximating elements $g = Sf$, $f \in F$, based on noisy values of n linear functionals at f. The noise is assumed to be Gaussian with correlation matrix $D = \operatorname{diag}\{\sigma_1^2, ..., \sigma_n^2\}$. The a priori measure μ on F is also Gaussian. We show how to choose the functionals from a ball to minimize the expected error of approximation. The error of the optimal approximation is given in terms of n, σ_i's, and the eigenvalues of the correlation operator of the a priori distribution $\nu = \mu S^{-1}$ on G.

1. INTRODUCTION

In the paper we consider the following approximation problem. Let S be a linear and continuous operator acting from a Banach space F to a Hilbert space G, both separable and over the real field. We wish to approximate elements Sf, for $f \in F$, based on noisy data about f. More specifically, an *information operator* (or *information*) $N: F \to \mathbb{R}^n$ is given by

$$N = [L_1, L_2, ..., L_n], \tag{1.1}$$

where the L_i's are linear and continuous functionals, $L_i \in F^*$, $1 \leqslant i \leqslant n$. For a (unknown) $f \in F$ we observe a random vector $z = [z_1, ..., z_n] \in \mathbb{R}^n$, which has n-dimensional normal distribution with mean Nf and correlation matrix $D = \operatorname{diag}\{\sigma_1^2, \sigma_2^2, ..., \sigma_n^2\}$. We assume that all σ_i's are known. The vector $\sigma = [\sigma_1, ..., \sigma_n]$ will be called a *precision vector*. An approximation to Sf is constructed based on z, i.e., $Sf \sim \phi(z)$, where $\phi: \mathbb{R}^n \to G$ is some transformation. The error of approximation is defined as the root of the expected squared norm of the error with the expectation taken over all f as well as z, i.e.,

$$e(\phi, N, \sigma) = \sqrt{\int_F \int_{\mathbb{R}^n} \|Sf - \phi(z)\|^2 \, \pi(dz \mid f) \, \mu(df)}, \tag{1.2}$$

0021-9045/93 $5.00

where $\pi(\cdot|f)$ is Gaussian on $\mathbb{R}^n$ with mean Nf and correlation matrix D, and μ is an a priori distribution on F. In this paper we assume that μ is the zero mean Gaussian measure with positive definite and symmetric correlation operator $C_\mu: F^* \to F$.

For an information operator N and precision vector σ, let

$$r(N, \sigma) = \inf_{\phi} e(\phi, N, \sigma).$$

It is well known that the infimum in $r(N, \sigma)$ is achieved for $\phi^*(z) = Sm(z)$, where $m(z)$ is the mean of the conditional (a posteriori) measure on F, after z has been observed. For given n and $\sigma = (\sigma_1, ..., \sigma_n)$, we want to minimize $r(N, \sigma)$ over a class of information operators N. More precisely, our aim is to find

$$r(\sigma) = \inf_{N} r(N, \sigma), \tag{1.3}$$

where the infimum is taken with respect to all $N = [L_1, ..., L_n]$ with the L_i's bounded in the μ-norm by $\|L_i\|_\mu = \sqrt{L_i(C_\mu L_i)} \leqslant 1$, $1 \leqslant i \leqslant n$. We also want to know *optimal information* N^*, for which $r(\sigma)$ is achieved, if it exists.

The above and related *optimal design* problems have been studied from different viewpoints in approximation theory, statistics, numerical analysis, and information-based complexity. In most cases, however, the authors assume exact information, see, e.g., Micchelli and Wahba [3], Novak [5], Papageorgiou and Wasilkowski [6], Ritter [8], Traub *et al.* [9], and Woźniakowski [11]. Although in practice, as a rule, observations use *noisy* information (for instance the function values are always contaminated by observational or round-off errors), optimal design results in this direction are very limited.

Estimation of functions from noisy data is a topic of many statistical works, see, e.g., Wahba [10] and references cited there. Some special results may be found in Traub [9] *et al.* Along with random noise, deterministic noise is studied. One of the first results on this subject was given by Micchelli and Rivlin [4]. From recent papers we mention Kacewicz and Plaskota [1].

In the present paper we generalize results of Plaskota [7] where the case $D = \sigma^2 I$ (I-identity matrix) is considered.

We now briefly outline the contents of this paper. In Section 2 we give two auxiliary lemmas. The main result is placed in Section 3, where solution of the optimal design problem is provided via solution of another, much simpler, minimization problem. In Section 4 we show how to construct optimal information, and give an explicit formula for its error. The error is given in terms of n, the σ_i's, and the eigenvalues of the correlation operator of the a priori measure $\nu = \mu S^{-1}$ on G. We also note that adaptive choice of information functionals does not reduce the minimal error of approximation.

2. Auxiliary Lemmas

In this section we give two lemmas which will be used in the main result. For $L_1, L_2 \in F^*$, let

$$\langle L_1, L_2 \rangle_\mu = L_1(C_\mu L_2). \tag{2.1}$$

Then the adjoint space F^* with the inner product (2.1) is a pre-Hilbert space (which is, however, not always complete). For information $N = [L_1, ..., L_n]$, we denote by M_N the Gram matrix $M_N = \{\langle L_i, L_j \rangle_\mu\}_{i,j=1}^n$. Proceeding as in Plaskota [7] for the case $D = \sigma^2 I$ we can now show the following

Lemma 2.1. *The conditional distribution on F with respect to the observed vector $z \in \mathbb{R}^n$ is Gaussian with mean*

$$m(z) = \sum_{i=1}^{n} y_i (C_\mu L_i),$$

where $y = [y_1, ..., y_n] \in \mathbb{R}^n$ is the solution of the linear system $(D + M_N)\, y = z$. Its correlation operator $C_{\mu,N,\sigma}: F^* \to F$ is independent of z and

$$C_{\mu,N,\sigma}(L) = C_\mu(L) - m(N(C_\mu L)), \qquad \forall L \in F^*.$$

Remark. Although we do not use it later, it is worthwhile to mention that the mean $m(z)$ is a smoothing spline. Optimality properties of splines are well known, see, e.g., Traub *et al.* [9] for the exact information case or Wahba [10] for approximation in a reproducing kernel Hilbert space. In our case we have that $m(z)$ is the minimizer in $C_\mu(F^*)$ of

$$\Gamma(f) = \|f\|_\mu^2 + \sum_{i=1}^{n} \frac{1}{\sigma_i^2} (z_i - L_i f)^2$$

(with the convention $0/0 = 0$), and $\Gamma(m(z)) = \langle y, z \rangle_2$. Here $\|f\|_\mu^2 = (C_\mu^{-1} f)(f)$ and y is as in Lemma 2.1. A proof of this fact can be obtained by straightforward calculation with the use of Lemma 2.1.

The second lemma is as follows

LEMMA 2.2. *Let the nonincreasing sequences* $\beta_1 \geqslant \beta_2 \geqslant \cdots \geqslant \beta_n \geqslant 0$ *and* $\eta_1 \geqslant \eta_2 \geqslant \cdots \geqslant \eta_n \geqslant 0$ *be such that*

$$\sum_{i=r}^{n} \eta_i \leqslant \sum_{i=r}^{n} \beta_i, \qquad 1 \leqslant r \leqslant n,$$

and $\sum_{i=1}^{n} \eta_i = \sum_{i=1}^{n} \beta_i$. *Then there exists a matrix* $W = \{w_{ij}\}_{i,j=1}^{n}$, *for which*

$$\sum_{s=1}^{n} w_{is}^2 = \beta_i, \qquad \text{and} \qquad \sum_{s=1}^{n} w_{si} w_{sj} = \eta_i \delta_{ij},$$

for all $1 \leqslant i, j \leqslant n$ (δ_{ij} *stands for the Kronecker delta*).

Proof. We construct the matrix W using induction on n. For $n = 1$ we have $w_{11} = \sqrt{\eta_1}$. Let $n \geqslant 2$. If $\eta_i = \beta_i$, $1 \leqslant i \leqslant n$, then $W = \mathrm{diag}\{\sqrt{\eta_1}, ..., \sqrt{\eta_n}\}$. Otherwise there is an index s, $1 \leqslant s \leqslant n-1$, that $\eta_s > \beta_s \geqslant \eta_{s+1}$. Set $\bar{\eta} = \eta_s + \eta_{s+1} - \beta_s > 0$, and let $U \in \mathbb{R}^{(n-1)\times(n-1)}$ be the required matrix for the sequences $\beta_1 \geqslant \cdots \geqslant \beta_{s-1} \geqslant \beta_{s+1} \geqslant \cdots \geqslant \beta_n$ and $\eta_1 \geqslant \cdots \geqslant \eta_{s-1} \geqslant \bar{\eta} \geqslant \eta_{s+2} \geqslant \cdots \geqslant \eta_n$. Let $u_i \in \mathbb{R}^{n-1}$ be the columns of U, $1 \leqslant i \leqslant n-1$. Then straightforward calculation gives that the desired matrix is $W = \{w_1, ..., w_n\}$, $w_i \in \mathbb{R}^n$, $1 \leqslant i \leqslant n$, where

$$\begin{aligned} w_i &= (u_i^T, 0)^T, \qquad \text{for} \quad i \neq s, s+1, \\ w_s &= (au_s^T, c)^T, \\ w_{s+1} &= (bu_s^T, d)^T \end{aligned}$$

(the superscript T denotes transposition), and

$$a = \left(\frac{\eta_{s+1}(\eta_s - \beta_s)}{\bar{\eta}(\eta_s - \eta_{s+1})}\right)^{1/2}, \qquad b = (1 - a^2)^{1/2},$$

$$c = \left(\frac{\eta_s(\beta_s - \eta_{s+1})}{(\eta_s - \eta_{s+1})}\right)^{1/2}, \qquad d = -(1 - c^2)^{1/2}. \quad \blacksquare$$

3. MAIN RESULT

In this section we present the main result about the optimal design. We assume without loss of generality that

$$0 = \sigma_1^2 = \cdots = \sigma_{n_0}^2 < \sigma_{n_0+1}^2 \leqslant \cdots \leqslant \sigma_n^2$$

(if all the σ_i's are nonzero then $n_0 = 0$).

Let $\nu = \mu S^{-1}$ be the a priori distribution on the space G, induced by the measure μ and the operator S. Then ν is a zero mean Gaussian with correlation operator

$$C_\nu = SC_\mu S^*: G \to G,$$

where $S^*: G \to F^*$ is the adjoint operator to S, $S^*g = \langle S(\cdot), g\rangle$, $\forall g \in G$. Moreover, C_ν is self-adjoint, nonnegative definite, and has finite trace. Let $\{\xi_i\}_{i=1}^{\dim G} \subset G$ be the complete and orthonormal system of eigenelements of the operator C_ν. Let $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \cdots \geqslant 0$ be the corresponding eigenvalues, $C_\nu \xi_i = \lambda_i \xi_i$. We consider the sequence $\{\lambda_i\}$ to be infinite by setting, if necessary, $\lambda_i = 0$ for $i > \dim G$. Define, for $\lambda_i > 0$, the functionals

$$K_i^* = \lambda_i^{-1/2} S^* \xi_i = \lambda_i^{-1/2} \langle S(\cdot), \xi_i \rangle \tag{3.1}$$

(for $\lambda_i = 0$ we formally set $K_i^* = 0$). Recall that the functionals (3.1) are μ-orthonormal, i.e., $\langle K_i^*, K_j^* \rangle_\mu = \delta_{ij}$. Moreover, they form optimal information in the exact information case, $\sigma = 0$, and $r(0) = \sum_{i=n+1}^{\infty} \lambda_i$ (see Traub *et al.* [9]). From this it immediately follows that in the case $\lambda_{n_0+1} = 0$ we have $r(\sigma) = 0$ and the optimal information is $[K_1^*, ..., K_{n_0}^*, 0, 0, ..., 0]$. To avoid these two obvious cases we will assume that $\sigma \neq 0$ (i.e., $n_0 < n$), and that $\lambda_{n_0+1} > 0$.

It turns out that the following minimization problem plays a crucial role in the optimal design:

PROBLEM (MP). *Minimize*

$$\Omega(\eta_{n_0+1}, ..., \eta_n) = \sum_{i=n_0+1}^{n} \frac{\lambda_i}{1+\eta_i}$$

over all nonnegative $\eta_{n_0+1}, ..., \eta_n$ *satisfying*

$$\sum_{i=r}^{n} \eta_i \leqslant \sum_{i=r}^{n} \frac{1}{\sigma_i^2}, \qquad n_0 + 1 \leqslant r \leqslant n,$$

and $\sum_{i=n_0+1}^{n} \eta_i = \sum_{i=n_0+1}^{n} \sigma_i^{-2}$.

Note that for the solution η^* of (MP) we have $\eta_{n_0+1}^* \geqslant \eta_{n_0+2}^* \geqslant \cdots \geqslant \eta_n^*$, which is easy to see. We are now ready to state the main theorem on the optimal design.

THEOREM 3.1. *Let* $\eta^* = (\eta_{n_0+1}^*, ..., \eta_n^*)$ *be the solution of* (MP). *Then*

$$r(\sigma) = \sqrt{\Omega(\eta^*) + \sum_{i=n+1}^{\infty} \lambda_i}.$$

Furthermore, the optimal information is

$$N^* = [K_1^*, \ldots, K_{n_0}^*, L_1^*, \ldots, L_{n-n_0}^*],$$

where

$$L_i^* = |\sigma_{n_0+i}| \sum_{j=1}^{n-n_0} w_{ij}^* K_{n_0+j}^*,$$

and $W^* = \{w_{ij}^*\}_{i,j=1}^{n-n_0}$ *is the matrix from Lemma* 2.2 *applied for*

$$\eta_i = \eta_{n_0+i}^*, \qquad \textit{and} \qquad \beta_i = \frac{1}{\sigma_{n_0+i}^2},$$

$1 \leqslant i \leqslant n - n_0$.

Before providing a proof we make a comment on how to construct optimal functionals (which are in span$\{K_1^*, \ldots, K_n^*\}$). To do this, we first have to solve (MP) and then find the matrix W^*. Solution of (MP) is given in the next section, see Theorem 4.1. The matrix W^* may be found by following the construction from the proof of Lemma 2.2.

Proof of the Theorem. We first provide a proof in the case when all σ_i's are nonzero, $n_0 = 0$. We do it in three steps.

Step 1. We derive a convenenient formula for $r(N, \sigma)$. For an information operator $N = [L_1, \ldots, L_n]$, define the matrix

$$M_{N,\sigma} = D^{-1/2} M_N D^{-1/2} = \{|\sigma_i \sigma_j|^{-1} \langle L_i, L_j \rangle_\mu\}_{i,j=1}^n.$$

It is symmetric and nonnegative definite. Let $M_{N,\sigma} u_i = \eta_i u_i$, $\langle u_i, u_j \rangle_2 = \delta_{ij}$, $1 \leqslant i, j \leqslant n$, and $\eta_1 \geqslant \cdots \geqslant \eta_n \geqslant 0$. Let

$$K_i = \langle D^{-1/2} N(\cdot), u_i \rangle_2 = \sum_{s=1}^n |\sigma_s|^{-1} u_{is} L_s, \tag{3.2}$$

$1 \leqslant i \leqslant n$. These functionals are μ-orthogonal and

$$\begin{aligned} \langle K_i, K_j \rangle_\mu = K_i(C_\mu K_j) &= \sum_{s=1}^n |\sigma_s|^{-1} u_{is} L_s(C_\mu K_j) \\ &= \sum_{s,t=1}^n |\sigma_s \sigma_t|^{-1} u_{is} u_{jt} \langle L_s, L_t \rangle_\mu \\ &= \langle M_{N,\sigma} u_i, u_j \rangle_2 = \eta_i \delta_{ij}. \end{aligned}$$

From (3.2) and Lemma 2.1 we have

$$\begin{aligned} m(D^{1/2}u_j) &= \sum_{i=1}^{n} ((D+M_N)^{-1} D^{1/2}u_j)_i \, C_\mu L_i \\ &= \sum_{i=1}^{n} (D^{-1/2}(I+M_{N,\sigma})^{-1} u_j)_i \, C_\mu L_i \\ &= (1+\eta_j)^{-1} \sum_{i=1}^{n} |\sigma_i|^{-1} u_{ji} C_\mu L_i = (1+\eta_j)^{-1} C_\mu K_j. \end{aligned}$$

Furthermore, for any $L \in F^*$

$$D^{-1/2}N(C_\mu L) = \sum_{j=1}^{n} \langle D^{-1/2}N(C_\mu L), u_j\rangle_2 u_j = \sum_{j=1}^{n} \langle L, K_j\rangle_\mu u_j.$$

Since, in addition, the mean $m(z)$ is linear with respect to z, we get

$$m(N(C_\mu L)) = \sum_{j=1}^{n} \langle L, K_j\rangle m(D^{1/2}u_j) = \sum_{j=1}^{n} (1+\eta_j)^{-1} \langle L, K_j\rangle_\mu \, C_\mu K_j.$$

Hence, the correlation operator $C_{\mu,N,\sigma}\colon F^* \to F$ of the conditional measure (see Lemma 2.1) can be rewritten as

$$C_{\mu,N,\sigma}(N) = C_\mu(L) - \sum_{j=1}^{n} \frac{1}{1+\eta_j} \langle L, K_j\rangle_\mu \, C_\mu K_j, \qquad \forall L \in F^*. \tag{3.4}$$

Using (3.4) and the well known fact that $(r(N,\sigma))^2 = \mathrm{trace}(C_{\nu,N,\sigma})$, where $C_{\nu,N,\sigma} = SC_{\mu,N,\sigma}S^*\colon G \to G$ is the correlation operator of the a posteriori measure on G, we obtain that

$$\begin{aligned} C_{\nu,N,\sigma}(g) &= C_\nu g - \sum_{j=1}^{n} \frac{1}{1+\eta_j} \langle S^* g, K_j\rangle_\mu \, S(C_\mu K_j) \\ &= C_\nu g - \sum_{j=1}^{n} \frac{1}{1+\eta_j} \langle S(C_\mu K_j), g\rangle \, S(C_\mu K_j), \qquad \forall g \in G, \end{aligned}$$

and the desired formula for $r(N,\sigma)$ is

$$\begin{aligned} (r(N,\sigma))^2 &= \sum_{i=1}^{\dim G} \langle C_{\nu,N,\sigma}\xi_i, \xi_i\rangle \\ &= \mathrm{trace}(C_\nu) - \sum_{j=1}^{n} \frac{1}{1+\eta_j} \|S(C_\mu K_j)\|^2. \end{aligned} \tag{3.5}$$

Step 2. Observe that the operator S^* can be treated as adjoint to $SC_\mu: F^* \to G$, with respect to the μ-inner product in F^*. Hence, for any μ-orthonormal functionals $\bar{K}_i$, $1 \leqslant i \leqslant n$, we have

$$\sum_{j=1}^{n} \|S(C_\mu \bar{K}_j)\|^2 = \sum_{j=1}^{n} \langle S^* S C_\mu \bar{K}_j, \bar{K}_j \rangle_\mu$$
$$\leqslant \sum_{j=1}^{n} \langle S C_\mu S^* \xi_j, \xi_j \rangle = \sum_{j=1}^{n} \lambda_j.$$

Moreover, from (3.3) and $\eta_1 \geqslant \cdots \geqslant \eta_n \geqslant 0$ it follows that

$$\sum_{j=1}^{n} \frac{1}{1+\eta_j} \|S(C_\mu K_j)\|^2 \leqslant \sum_{j=1}^{n} \frac{\eta_j}{1+\eta_j} \lambda_j.$$

From this and (3.5) we obtain

$$(r(N, \sigma))^2 \geqslant \sum_{j=1}^{\infty} \lambda_j - \sum_{j=1}^{n} \frac{\eta_j}{1+\eta_j} \lambda_j$$
$$= \Omega(\eta_1, ..., \eta_n) + \sum_{j=n+1}^{\infty} \lambda_j.$$

In addition, for all $1 \leqslant r \leqslant n$ we have

$$\sum_{i=r}^{n} \eta_i \leqslant \sum_{i=r}^{n} \langle M_{N,\sigma} e_i, e_i \rangle_2 = \sum_{i=r}^{n} \frac{1}{\sigma_i^2}$$

(where e_i stands for the ith versor), and $\sum_{i=1}^{n} \eta_i = \sum_{i=1}^{n} \sigma_i^{-2}$. Thus

$$r(\sigma) \geqslant \sqrt{\Omega(\eta^*) + \sum_{i=n+1}^{\infty} \lambda_i}. \tag{3.6}$$

Step 3. We now show that $r(N^*, \sigma)$ is equal to the right side of the inequality (3.6). Indeed, the matrix $M_{N^*,\sigma} = (W^*)(W^*)^T$, the (orthogonal) columns w_i^* of W^* are the eigenvectors of $M_{N^*,\sigma}$, and $M_{N^*,\sigma} w_i^* = \eta_i^* w_i^*$, $1 \leqslant i \leqslant n$. Furthermore, according to (3.2), the corresponding functionals K_i are

$$K_i = \sum_{s=1}^{n} |\sigma_s|^{-1} w_{si}^* L_s^* = \sum_{s=1}^{n} |\sigma_s|^{-1} w_{si}^* |\sigma_s| \sum_{t=1}^{n} w_{st}^* K_t^*$$
$$= \sum_{t=1}^{n} \langle w_i^*, w_t^* \rangle_2 K_t^* = \eta_i^* K_i^*.$$

From this, the formula (3.5), and from the fact that $\|SC_\mu K_j^*\|^2 = \lambda_j$, $j \geq 1$, it follows that

$$r(N^*, \sigma) = \sqrt{\sum_{j=1}^{n} \frac{\lambda_j}{1+\eta_j^*} + \sum_{j=n+1}^{\infty} \lambda_j},$$

which completes the third step and the proof of the case $n_0 = 0$.

Suppose now that $n_0 \geq 1$. Then $N = [N_0, N_1]$ and $\sigma = (\sigma^{(0)}, \sigma^{(1)})$, where $N_0 = [L_1, ..., L_{n_0}]$, $N_1 = [L_{n_0+1}, ..., L_n]$, and $\sigma^{(0)} = (\sigma_1, ..., \sigma_{n_0})$, $\sigma^{(1)} = (\sigma_{n_0+1}, ..., \sigma_n)$. The a posteriori Gaussian measure on F with respect to information N_0 (which is obtained exactly, $\sigma^{(0)} = 0$) has the correlation operator $C_{\mu, N_0} = C_\mu(I - P_{N_0})$, where $P_{N_0}: F^* \to F^*$ is the μ-orthogonal projection onto span$\{L_1, ..., L_{n_0}\}$. For the dominating eigenvalues $\bar{\lambda}_i$ of $SC_{\mu, N_0}S^*$, which is the correlation operator of the a posteriori measure on G with respect to N_0, we have $\bar{\lambda}_i \geq \lambda_{n_0+i}$, $\forall i \geq 1$. Moreover, if $N_0 = N_0^* = [K_1^*, ..., K_{n_0}^*]$ then $\bar{\lambda}_i = \lambda_{n_0+i}$, and the corresponding eigenelements are $\bar{\xi}_i = \xi_{n_0+i}$, $\forall i \geq 1$. Hence, we obtain the desired result by reducing our problem to that of finding optimal N_1, where the precision is $\sigma^{(1)}$ and the a priori distribution on F is Gaussian with correlation operator $C_\mu(I - P_{N_0^*})$. ∎

4. Explicit Formulas for η^* and $r(\sigma)$

In this section we provide an explicit formula for the solution $\eta^* = (\eta_{n_0+1}^*, ..., \eta_n^*)$ of the minimization problem (MP) as well as for the optimal error $r(\sigma)$.

For $n_0 \leq q < r \leq n$, define the following auxiliary minimization problem

Problem (P(q, r)). *Minimize*

$$\Omega_{qr}(\eta_{q+1}, ..., \eta_r) = \sum_{j=q+1}^{r} \frac{\lambda_j}{1+\eta_j}$$

over all nonnegative $\eta_{q+1}, ..., \eta_r$ *satisfying*

$$\sum_{j=q+1}^{r} \eta_j = \sum_{j=q+1}^{r} \frac{1}{\sigma_j^2}.$$

The solution $\eta^+ = (\eta_{q+1}^+, ..., \eta_r^+)$ of (P(q, r)) is as follows. Let $k = k(q, r)$ be the largest integer satisfying $q + 1 \leq k \leq r$ and

$$\frac{\sum_{j=q+1}^{k} \lambda_j^{1/2}}{\sum_{j=q+1}^{r} \sigma_j^{-2} + (k-q)} \leq \lambda_k^{1/2}. \tag{4.1a}$$

Then

$$\eta_i^+ = \frac{\sum_{j=q+1}^{r} \sigma_j^{-2} + (k-q)}{\sum_{j=q+1}^{k} \lambda_j^{1/2}} \cdot \lambda_i^{1/2} - 1, \qquad \text{for} \quad q+1 \leqslant i \leqslant k, \tag{4.1b}$$

and $\eta_i^+ = 0$, for $k+1 \leqslant i \leqslant r$. Furthermore,

$$\Omega_{qr}(\eta^+) = \frac{(\sum_{j=q+1}^{k} \lambda_j^{1/2})^2}{\sum_{j=q+1}^{r} \sigma_j^{-2} + (k-q)} + \sum_{j=k+1}^{r} \lambda_j. \tag{4.1c}$$

We say that the solution η^+ of Problem $(\mathrm{P}(q, r))$ is *acceptable* iff

$$\sum_{j=s}^{r} \eta_j^+ \leqslant \sum_{j=s}^{r} \frac{1}{\sigma_j^2}, \qquad \text{for all} \quad q+1 \leqslant s \leqslant r. \tag{4.2}$$

Let the number p, $0 \leqslant p < n$, and the sequence $0 \leqslant n_0 < n_1 < \cdots < n_p < n_{p+1} = n$ be defined (uniquely) by the condition

$$n_i = \min\{s \geqslant n_0 \colon \text{solution of } (\mathrm{P}(s, n_{i+1})) \text{ is acceptable}\}, \tag{4.3}$$

for all $0 \leqslant i \leqslant p$.

Using the formulas (4.1a)–(4.1b) and the condition (4.3) we can construct the solution of the original minimization problem (MP). Indeed, this follows from the following

THEOREM 4.1. *Let p and the sequence $n_0 < n_1 < \cdots < n_{p+1} = n$ be defined by* (4.3). *Then the optimal η^* is*

$$\eta^* = (\eta^{(0)}, \eta^{(1)}, ..., \eta^{(p)}),$$

where $\eta^{(i)} = (\eta^*_{n_i+1}, ..., \eta^*_{n_{i+1}})$ is the solution of $(\mathrm{P}(n_i, n_{i+1}))$, $0 \leqslant i \leqslant p$.

Proof. Let $t = \max\{n_0 + 1 \leqslant i \leqslant n \colon \eta_i^* > 0\}$. For $n_0 + 1 \leqslant i \leqslant t$, the function

$$\psi_i(\tau) = \Omega(\eta^*_{n_0+1}, ..., \eta^*_{i-2}, \eta^*_{i-1} + \eta^*_i - \tau, \tau, \eta^*_{i+1}, ..., \eta^*_n)$$

is continuous, convex, and attains the minimum at τ_0 such that $\lambda_{i-1}(1 + \eta^*_{i-1} + \eta^*_i - \tau_0)^{-2} = \lambda_i(1 + \tau_0)^{-2}$. From this and from the definition of (MP) it easily follows that

$$\frac{\lambda_{i-1}}{(1+\eta^*_{i-1})^2} \leqslant \frac{\lambda_i}{(1+\eta^*_i)^2}. \tag{4.4a}$$

Moreover, if $\lambda_{i-1}(1+\eta^*_{i-1})^{-2} < \lambda_i(1+\eta^*_i)^{-2}$ then $\sum_{j=i}^{n} \eta_j^* = \sum_{j=i}^{n} \sigma^{-2}$. If $t < n$ then, using the same argument with $i = t+1$, we find that

$$\frac{\lambda_t}{(1+\eta^*_t)^2} \geqslant \lambda_{t+1}. \tag{4.4b}$$

Let $m_1 < \cdots < m_s$ be the sequence of all indices i, $n_0 < i < t$, for which $\lambda_i(1+\eta_i^*)^{-2} < \lambda_{i+1}(1+\eta_{i+1}^*)^{-2}$. Set $m_0 = n_0$ and $m_{s+1} = n$. From (4.4a) it follows that $\sum_{j=m_i+1}^{m_{i+1}} \eta_j^* = \sum_{j=m_i+1}^{m_{i+1}} \sigma_j^{-2}$, $0 \leqslant i \leqslant s$. This and (4.4b) yield that the numbers $\eta_{m_i+1}^*, ..., \eta_{m_{i+1}}^*$ are the solution of $(\mathrm{P}(m_i, m_{i+1}))$, for all $0 \leqslant i \leqslant s$. The complete the proof, it is now enough to show that the sequences $\{m_i\}_{i=0}^{s+1}$ and $\{n_i\}_{i=0}^{p+1}$ are the same, i.e., $\{m_i\}_{i=0}^{q+1}$ satisfies (4.3). Indeed, suppose to the contrary that for some i there is j_0, $0 \leqslant j_0 < m_i$, such that the solution $\eta_{j_0+1}^+, ..., \eta_{m_{i+1}}^+$ of $\mathrm{P}(j_0, m_{i+1}))$ is acceptable. Then

$$\sum_{j=m_i+1}^{m_{i+1}} \eta_j^+ \leqslant \sum_{j=m_i+1}^{m_{i+1}} \frac{1}{\sigma_j^2} = \sum_{j=m_i+1}^{m_{i+1}} \eta_j^*.$$

From this and the formulas (4.1a)–(4.1b) we get that $\eta_j^+ \leqslant \eta_j^*$, for all $m_i + 1 \leqslant j \leqslant m_{i+1}$. Similarly, for $j_0 \leqslant j \leqslant m_i$ we have

$$\frac{\lambda_j}{(1+\eta_j^*)^2} < \frac{\lambda_{m_i+1}}{(1+\eta_{m_i+1}^*)^2} \leqslant \frac{\lambda_{m_i+1}}{(1+\eta_{m_i+1}^+)^2} = \frac{\lambda_j}{(1+\eta_j^+)^2},$$

and consequently $\eta_j^+ < \eta_j^*$. Hence,

$$\sum_{j=j_0+1}^{m_{i+1}} \frac{1}{\sigma_j^2} = \sum_{j=j_0+1}^{m_{i+1}} \eta_j^+ < \sum_{j=j_0+1}^{m_{i+1}} \eta_j^*,$$

which is a contradiction. ∎

Knowing optimal η^* we can write an explicit formula for $r(\sigma)$. Observe that for $i < p$ the number $k = k(n_{i-1}, n_i)$, defined by (4.1a), is equal to n_i. From this, Theorem 4.1, and from the formula (4.1c) we obtain

COROLLARY 4.1. *Let p and $\{n_i\}_{i=0}^{p+1}$ be defined by* (4.3), *and let $k = k(n_p, n)$ be given by* (4.1a). *Then*

$$r(\sigma) = \sqrt{\sum_{i=0}^{p-1} \frac{\left(\sum_{j=n_i+1}^{n_{i+1}} \lambda_j^{1/2}\right)^2}{\sum_{j=n_i+1}^{n_{i+1}} \sigma_j^{-2} + (n_{i+1} - n_i)} + \frac{\left(\sum_{j=n_p+1}^{k} \lambda_j^{1/2}\right)^2}{\sum_{j=n_p+1}^{n} \sigma_j^{-2} + (k - n_p)} + \sum_{j=k+1}^{\infty} \lambda_j}.$$

As we see, the formula for the optimal error $r(\sigma)$, given in terms of the eigenvalues λ_i and precisions σ_i, $1 \leqslant i \leqslant n$, is rather complicated. Let, for simplicity, all σ_i's be nonzero. Then we have the following bounds on $r(\sigma)$:

$$\sqrt{\frac{\left(\sum_{i=1}^{k} \lambda_i^{1/2}\right)^2}{\sum_{i=1}^{n} \sigma_i^{-2} + k} + \sum_{i=k+1}^{\infty} \lambda_i} \leqslant r(\sigma) \leqslant \sqrt{\sum_{i=1}^{n} \frac{\lambda_i}{s + \sigma_i^{-2}} + \sum_{i=n+1}^{\infty} \lambda_i}, \tag{4.5}$$

where k is the largest integer satisfying $1 \leqslant k \leqslant n$ and

$$\frac{\sum_{j=1}^{k} \lambda_j^{1/2}}{\sum_{j=1}^{n} \sigma_j^{-2} + k} \leqslant \lambda_k^{1/2}.$$

Observe that the lower bound in (4.5) depends on the λ_i's and the sum $\sum_{j=1}^{n} \sigma_j^{-2}$. It is achieved if σ belongs to the region

$$\sum_{i=s}^{n} \eta_i^{**} \leqslant \sum_{i=s}^{n} \frac{1}{\sigma_i^2}, \qquad 1 \leqslant s \leqslant n, \tag{4.6}$$

where $\eta^{**} = (\eta_1^{**}, ..., \eta_n^{**})$ is the solution (4.1b) of the problem (P(0, n)). For instance, (4.6) always holds for $\sigma_1 = \cdots = \sigma_n$ (this case is known from Plaskota [7]). On the other hand, the upper bound in (4.5) is achieved if for all $0 \leqslant q < r \leqslant n$ the solution η^+ of (P(q, r)) satisfies

$$\sum_{j=s}^{r} \eta_j^{+} \geqslant \sum_{j=s}^{r} \frac{1}{\sigma_j^2}, \qquad q+1 \leqslant s \leqslant r.$$

This always holds, for instance, for $\lambda_1 = \cdots = \lambda_n$. In such cases the optimal information is $N^* = [K_1^*, ..., K_n^*]$. Note also that if there are equalities in (4.6) for all s then both bounds in (4.5) are the same.

In this paper we have considered only *nonadaptive* information operators. It turns out that the minimal error of approximation coming from the use of n *adaptive* information functionals with precision $\sigma = (\sigma_1, ..., \sigma_n)$ is not less than $r(\sigma)$. This may be shown in a standard way; see Kadane *et al.* [2] and Plaskota [7].

References

1. B. Z. Kacewicz and L. Plaskota, On the minimal cost of approximating linear problems based on information with deterministic noise, *Numer. Funct. Anal. Optim.* **11** (1990), 511–525.
2. J. B. Kadane, G. W. Wasilkowski, and H. Woźniakowski, On adaption with noisy information, *J. Complexity* **4** (1988), 257–276.
3. C. A. Micchelli and G. Wahba, Design problems for optimal surface interpolation, *in* "Approximation Theory and Its Applications" (Z. Ziegler, Ed.), pp. 329–348, Academic Press, Orlando, FL, 1981.
4. C. A. Micchelli and T. J. Rivlin, A survey of optimal recovery, *in* "Estimation in Approximation Theory," pp. 1–54, Plenum, New York, 1977.
5. E. Novak, Deterministic and stochastic error bounds in numerical analysis, *in* "Lecture Notes in Math.," Vol. 1349, Springer-Verlag, New York/Berlin, 1988.
6. A. Papageorgiou and G. W. Wasilkowski, Average complexity of multivariate problems, *J. Complexity* **5** (1990), 1–23.

7. L. PLASKOTA, On average case complexity of linear problems with noisy information, *J. Complexity* **6** (1990), 1–23.
8. K. RITTER, Approximation and optimization on the Wiener space, *J. Complexity* **6** (1990), 337–364.
9. J. F. TRAUB, G. W. WASILKOWSKI, AND H. WOŹNIAKOWSKI, "Information-Based Complexity," Academic Press, Orlando, FL, 1988.
10. G. WAHBA, Spline models for observational data, *in* "CBMS-NSF Regional Conference Series in Appl. Math." Vol. 59, SIAM, Philadelphia, 1990.
11. H. WOŹNIAKOWSKI, Average case complexity of multivariate integration, *Bull. Amer. Math. Soc.* **24** (1991), 185–194.