Direct Estimation of Linear Functionals from Indirect Noisy Observations

Peter Mathé

Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39,
D-10117 Berlin, Germany
E-mail: mathe@wias-berlin.de

and

Sergei V. Pereverzev

Ukrainian Academy of Sciences, Institute of Mathematics, Tereshenkivska Str. 3,
Kiev 4, Ukraine
E-mail: serg-p@mail.kar.net

Received January 2, 2001; revised March 13, 2001; accepted May 2, 2001;
published online October 22, 2001

The authors study the efficiency of the linear-functional strategy, as introduced
by Anderssen in 1986, for inverse problems with observations blurred by Gaussian
white noise with known intensity $\sigma$. The optimal accuracy is presented and it is
shown how this can be achieved by a linear-functional strategy based on the noisy
observations. This optimal linear-functional strategy is obtained from Tikhonov
regularization of some dual problem. Next, the situation is treated when only a
finite number of noisy observations, given beforehand, is available. Under appro-
priate smoothness assumptions best possible accuracy still can be attained if the
number of observations corresponds to the noise intensity in a proper way. It is also
shown that, at least asymptotically, this number of observations cannot be reduced.

Key Words: inverse problems; optimal data-functional strategy; discretization;
Gaussian white noise; information complexity.

The research of S. V. Pereverzev was supported by a grant of the Deutsche Forschungs-
gemeinschaft (DFG).
Inverse problems can often be represented in the form of ill-posed linear operator equations \( Ax = y \) in some real Hilbert space \( X \). Here \( x \) is the quantity of interest, but only indirect measurements \( y \) are available, often blurred by noise, such that we assume the actual form

\[
y_d = Ax + \delta \xi,
\]

where \( \xi \) denotes some normalized noise and \( \delta > 0 \) represents its intensity.

In many cases one is not interested in completely knowing \( x \), but in some derived quantities of it. As has been pointed out by Anderssen [1], such derived quantities often correspond to bounded linear functionals of the solution. Then the problem is to estimate \( \langle f, x \rangle \), where \( f \) is any given functional, under complete knowledge of \( y_d \).

**Problem 1.A.** Given \( f \), find approximation to \( \langle f, x \rangle \), based on observations \( y_d = Ax + \delta \xi \).

A straightforward approach to this end would be to find some approximate solution of (1) by some means of regularization and then apply the given \( f \) to this. Anderssen [1] referred to this as the solution-functional strategy. It was analyzed by Engl and Neubauer [4] in the case when the noise \( \xi \) is deterministic, subject to \( \| \xi \|_X \leq 1 \), and Goldschlager and Pereverzev [9] in the case of Gaussian white noise.

Another approach is to estimate \( \langle f, x \rangle \) directly from \( y_d \), as proposed by Golberg [7] and Anderssen [1], who referred to this approach as the data-functional strategy or linear-functional strategy. The idea consists in transforming the functional \( f \), originally defined on the solution to a functional \( z \) defined on the data, given by \( A^* z = f \), where \( A^* \) is the adjoint operator of \( A \) in (1). The advantage is obvious, since this transformation allows to precompute the data functional beforehand, and then use this for any given data. For some special inverse problems this approach has been studied extensively, see [1, 3, 5, 8, 12, 16, 22]. But, only regular solution functionals can be treated in this way. More precisely, the solution functional must belong to the range of \( A^* \).

This disadvantage was overcome partly by Louis [13], who proposed the mollifier method, which consists in applying least-square regularization to the defining equation \( A^* z = f \), such that the data functional will be obtained as \( z = \arg \min \{ \| A^* u - f \|_X, u \in X \} \). As we see, in the mollifier method, the solution functional must still be regular, as it must belong to \( X \).

If, for example we knew that the solution \( x \) was a differentiable function...
and we were interested in the value of a certain derivative at some point, then mollification would not be applicable directly.

It is one of the goals of this paper to keep the advantages of the linear-functional strategy, but to extend the range of possible applications. This will be done within the context of Hilbert scales.

There is one more issue to be pointed out. Typically, the indirect observations \( y_\delta \) are not given completely. Instead we can observe only finitely many functionals, say \( \Phi = \{ \phi_1, \ldots, \phi_n \} \) on \( X \), called design, such that we actually have knowledge

\[
y_{\delta, i} := \langle y_\delta, \phi_i \rangle = \langle Ax, \phi_i \rangle + \delta \langle \xi, \phi_i \rangle, \quad i = 1, \ldots, n. \tag{2}
\]

So if we are free to choose the design, then the observation of only one functional suffices to reach the best possible order of accuracy. It will be transparent that the best possible order of accuracy can be achieved by some linear data-functional strategy, namely the data functional \( z \), as described above. This situation will be studied below in Section 3.

We note that (2) can also be regarded as direct observations

\[
y_{\delta, i} = \langle x, f_i \rangle + \delta \xi_i, \quad i = 1, \ldots, n,
\]

with \( f_i := A^* \phi_i \). But for such \( f_i \) the correlation structure of the noise vector \( (\xi_i) \) depends on the operator \( A \) which is not natural unless \( A \) is the identity. We refer to [2] for further discussion of this topic.

In most cases the design is given beforehand, independently of the operator \( A \) and the functional \( f_i \), i.e., as Fourier coefficients or as averages of histogram bins \( [t_{i-1, n}, t_{i, n}) \) with bin limits \( 0 = t_{0, n} < t_{1, n} < \cdots < t_{n, n} = 1 \). In the latter case the design consists of the \( n \) (normalized) characteristic functions of the intervals \( [t_{i-1, n}, t_{i, n}) \). Throughout we shall assume that \( \Phi \) is given as a finite family of orthonormal functions in \( X \).

**Problem 1.B.** Given \( f \) and design \( \Phi = \{ \phi_1, \ldots, \phi_n \} \), find approximation to \( \langle f, x \rangle \), based on observations \( y_{\delta, i}, i = 1, \ldots, n \), as in (2).

In Section 4 we study this issue. As it will turn out, under some restriction on the approximative power of the design, the best possible order of accuracy can be obtained. We indicate some cardinality \( n = n(\delta) \), depending on these properties of the design, which guarantees best possible order of accuracy of recovering \( \langle f, x \rangle \), as \( \delta \to 0 \). The crucial point here is that \( \Phi \) cannot adapt to the functional. So, another issue seems to be important and will be studied in Section 5.
Problem 1.C. What is, asymptotically as $\delta \to 0$, the least size of any design, which is independent of the functional $f$ and still allows methods of optimal accuracy?

The main objective of the paper is to study Eq. (1) under Gaussian white noise. Deterministic noise will also be briefly discussed in Section 6.

2. SETUP

As mentioned above it will be convenient for us to study the inverse problems in Hilbert scales.

A Hilbert scale $\{X^l\}_{l \in \mathbb{R}}$ is a family of Hilbert spaces $X^l$ with inner products $\langle u, v \rangle_l := \langle L^lu, L^lv \rangle$, where $L$ is a given unbounded strictly positive self-adjoint operator in a dense domain of some initial Hilbert space, say $X$. To be more precise, $X^l$ is defined as the completion of the intersection of domains of the operators $L^l$, $v \geq 0$, accomplished with norm $\|\cdot\|_l := \|L^l\cdot\|_0$, where $\|\cdot\|_0 = \|\cdot\|$ is the norm in $X$. The following interpolation property will be important below: For any triple of indices $r < s < t$ let $\theta = \frac{t-r}{t-s}$. Then we have $\|x\|_s \leq \|x\|_r \|x\|_t^{1-\theta}$, whenever $x \in X^t$.

Hilbert scales $\{X^l\}$ are invariant with respect to rescaling $\lambda \to a\lambda + b$, for $a > 0$, $b \in \mathbb{R}$. Since usually $\{X^l\}$ are specific Sobolev spaces, say $H^l(0, 1)$, for definiteness of scaling we assume $\lambda$ be chosen to fit the usual smoothness as e.g., in $H^l(0, 1)$. This goal is achieved by assuming that the canonical embeddings $J_l: X^l \to X$, $\lambda > 0$, obey

$$a_n(J_l) := \inf \{\|J_l - U\|_{X^l \to X}, \text{rank}(U) < n\} \asymp n^{-l},$$  

where $\asymp$ means equivalent in order, and $a_n$ denotes the $n$th approximation number; see [21], for example. We note that as a consequence $a_n(J^l_{\mu}: X^\mu \to X^l) \asymp n^{-\mu(\mu-\nu)}$ whenever $\mu > \nu$. By studying mathematical problems like (1) in Hilbert scales, for an operator $A$, which initially acts within $X$, we mean that both, the domain $X$ and the target $Y$ belong to appropriate Hilbert scales $X \in \{X^l\}$ and $Y \in \{Y^l\}$, which are linked by assuming $X^0 = Y^0 = X$ (actually only a finite segment of parameters $\lambda \in \mathbb{R}$ will be involved). Moreover we assume that the scaling for $\{Y^l\}$ is the same as for $\{X^l\}$. When indicating norms in spaces $X^l$ or $Y^l$, we shall often suppress the symbol for the space and just mention the parameter. It will be transparent from the context, which scale is meant.
The basic assumption concerning the operator $A$ is as follows: There exist $a > 0$ and constants $D, d > 0$ such that for all $\lambda \in \mathbb{R}$ and all $x \in X^{1-a}$ we have
\begin{equation}
 d \| x \|_{1-a} \leq \| Ax \| \leq D \| x \|_{1-a}.
\end{equation}
In other words, the operator $A$ acts along the Hilbert scales with step $a$. There are many examples of such operators; see, e.g., [15, 17, 20].

We continue this section with the assumptions on the noise. For stochastic noise (Gaussian white noise) we assume that $\xi$ is a weak or generalized random element in $Y^0$, such that for any $f \in Y^0$ the inner product $\langle \xi, f \rangle$ is a zero mean Gaussian random variable on a probability space $(\Omega, \Sigma, \mathbb{P})$ with variance $\|f\|^2$. Denoting $E$ the expectation with respect to $\mathbb{P}$, we have as a consequence
\begin{equation}
 E\langle f, \xi \rangle \langle g, \xi \rangle = \langle f, g \rangle, \quad \text{for any } f, g \in Y^0.
\end{equation}

So far, the description of the problem (1) as an ill-posed problem is not complete. We shall assume the following a priori knowledge on the exact solution $x = A^{-1}y$, namely that it belongs to the unit ball $U^\mu := \{x \in X^\mu, \|x\| \leq 1\}$ of $X^\mu$ for a certain value $\mu > 0$. Since by our assumptions the dual of $X^\mu$ is $X^{-\mu}$, it is natural to assume the admissible solution functionals belong to $X^\nu$ for some $\nu \geq -\mu$. We even assume the functional to be normalized, i.e., $f \in U^\nu$.

Given any functional $f$, let $S$ be any measurable mapping to the reals, which may be considered as approximating $\langle f, x \rangle$ by $S(y_\delta)$. Its error is then measured as
\begin{equation}
 e_{\delta}(f, S, \delta) := \sup_{x \in U^\nu} (E |\langle f, x \rangle - S(y_\delta)|^2)^{1/2}.
\end{equation}

It is transparent that any approximation has error at least proportional to $\delta$. Moreover, this best possible accuracy can be achieved for smoothness $\nu \geq a$. Therefore we assume $-\mu \leq \nu \leq a$ throughout.

This paper focuses at asymptotic considerations. Therefore, throughout the paper $c$ denotes a generic constant and may vary from appearance to appearance.

Remark 2.1. Regularization of ill-posed problems in Hilbert scales has been introduced by Natterer [19]. Some pertinent references are Neubauer [20], Mair [14], Hegland [10], and Tautenhahn [23]. These authors studied only deterministic noise.
Statistical inverse estimation in Hilbert scales has been studied by Mair and Ruymgaart [15], though these authors did not study observations in the form (1) or (2). The case of discretized white noise observations in Hilbert scales was studied only recently by the present authors in [17].

3. THE OPTIMAL DATA-FUNCTIONAL STRATEGY

We complete the description of Problem 1.A by introducing the respective error criterion. We are looking for optimal accuracy methods \( S \) and their accuracy as in (6), uniformly for \( f \in U^* \), which means we aim at determining the asymptotic behavior as \( \delta \to 0 \) of

\[
\varepsilon_{\mu, \delta} := \sup_{f \in U^*} \inf_S e_{\mu}(f, S, \delta).
\]

The main result of this section is

**Theorem 3.1.** Under assumption (4), the minimal error for solving Problem 1.A under the presence of white noise is

\[
\varepsilon_{\mu, \delta} \asymp \delta^{\alpha+\mu}, \quad \text{as} \quad \delta \to 0.
\]

The optimal order is attained by the (linear) optimal data-functional strategy \( \langle z, y_a \rangle \) with

\[
z = (\delta^2 I + AL^{-2\mu}A^*)^{-1} AL^{-2\mu} f. \tag{7}
\]

Before turning to the proof let us establish some norm bounds for the functional \( z \) from (7), which will be useful below.

**Proposition 3.1.** Let \( z \) be as in (7). Then \( z \) belongs to \( X^{\alpha+\mu} \) and there is a constant \( c < \infty \), such that for \( 0 \leq s \leq \alpha + \mu \) the following bound holds true:

\[
\|z\|_s \leq c \|f\|, \quad \delta^{\frac{s + \alpha + \mu}{s + \alpha + \mu}}. \tag{8}
\]

**Proof.** We start with the following assertion from [19]. For the operator \( B := AL^{-\mu} : X \to X \) we have

\[
X^{\mu(\alpha+\mu)} = \text{Im}((B^*B)^{\mu/2}), \quad |t| \leq 1.
\]
Consequently, if \( f \in X^\nu, \nu \leq a \), then \( L^{-\mu}f \in X^{\mu+}, \) which guarantees the existence of some \( v_f \in X \), for which

\[
L^{-\mu}f = (B^*B)^{(\mu+)/(2(a+\rho))} v_f.
\]

(9)

Also, there is a constant \( c < \infty \) for which \( \|v_f\| \leq c \|f\| \). In terms of the operator \( B \) we can rewrite \( z = B(\delta^2 I + B^*B)^{-1} L^{-\mu}f \). Since \( L^{-\mu}f \in X \) and \( B \) maps \( X \) to \( X^{\mu+} \), the first assertion is proven. Using (9) we can bound the extremal norms,

\[
\|z\|_0 = \|B(\delta^2 I + B^*B)^{-1} (B^*B)^{(\mu+)/(2(a+\rho))} v_f\|
\leq \|v_f\| \sup_{t > 0} t^\frac{\mu+\rho}{2\mu+2\rho}(\delta^2 + t)^{-1} \leq \|f\| \delta^{(\nu-\mu)/(\mu+\rho)}.
\]

(10)

Similarly,

\[
\|z\|_+ \leq D \|B(\delta^2 I + B^*B)^{-1} (B^*B)^{(\mu+)/(2(a+\rho))} v_f\|
\leq c \|v_f\| \sup_{t > 0} t^\frac{\mu+\rho}{2\mu+2\rho}(\delta^2 + t)^{-1} \leq c \|f\| \delta^{(\nu-\mu)/(\mu+\rho)}.
\]

Now, interpolation readily provides estimate (8).

We turn to the

**Proof of Theorem 3.1.** The right asymptotics is provided using a result by Donoho [2], who proved that for each functional \( f \) on \( X^\nu \),

\[
\inf_S e_\mu(f, S, \delta) \asymp \sup_{x \in U^\nu} |\langle f, x \rangle| \quad \text{as} \quad \delta \to 0.
\]

Uniformly for \( f \in U^\nu \) this implies

\[
\delta'_\mu(\delta) \asymp \sup_{f \in U^\nu} \sup_{\|x\|_{-\nu} \leq \delta} |\langle f, x \rangle| = \sup_{x \in U^\nu} \|x\|_{-\nu}.
\]

For \( \nu < \alpha \) the right hand side quantity is well studied within the framework of regularization in Hilbert scales; see, e.g., [23]. We have

\[
\sup_{x \in U^\alpha} \|x\|_{-\nu} \asymp \delta'^{\nu\alpha}_{\mu\alpha},
\]

\[
\|x\|_{-\nu} \asymp \delta'^{\nu\alpha}_{\mu\alpha},
\]

\[
\|x\|_{-\nu} \asymp \delta'^{\nu\alpha}_{\mu\alpha},
\]
which completes the determination of the right asymptotics of $\mathcal{E}_\nu,\delta(\delta)$ as $\delta \to 0$.

It remains to show that the functional $z$ from (7) attains this asymptotics. To see this we first mention that $z$ from (7) is the unique minimizer of the following Tikhonov functional $T(z) := \|A^*z - f\|_\mu^2 + \alpha \|z\|_2^2$ for $\alpha := \delta^2$, which appears in the minimization problem

$$
\sup_{x \in U^\alpha} \mathbb{E} |\langle f, x \rangle - \langle z, y_d \rangle|^2 \to M1N(z),
$$

(11)

which in turn minimizes $e_\mu(f, S, \delta)$ over all linear functionals $S: X \to \mathbb{R}$. This can be seen, using (5), from

$$
\sup_{x \in U^\alpha} \mathbb{E} |\langle f, x \rangle - \langle z, y_d \rangle|^2 = \sup_{x \in U^\alpha} \mathbb{E} |\langle f - A^*z, x \rangle - \delta \langle z, \xi \rangle|^2
\leq \|A^*z - f\|_\mu^2 + \delta^2 \|z\|_2^2.
$$

In particular $e_\mu(f, z, \delta) \leq \|A^*z - f\|_\mu^2 + \delta \|z\|_0$. Now we estimate the accuracy of the optimal data-functional strategy $z$. Since

$$
\|A^*z - f\|_\mu = \|L^{-\mu}f - L^{-\mu}A^*z\|
= \|(I - B^*B(\delta^2 + \delta^2 B^*B)^{-1})B^*B(\mu + 1)/(2(\mu + \mu)) v_f\|
\leq \delta^2 \|v_f\| \sup_{t > 0} t^{\frac{\mu+1}{\mu+1}} \|e_{\mu}^{\delta^2 + \delta^2}\|
\leq \delta^2 \|v_f\| \sup_{t > 0} \|e_{\mu}^{\delta^2 + \delta^2}\|
\leq c \|f\| \delta^{\frac{\mu}{\mu+1}}.
$$

and using the bound from (8), we arrive at $\sup_{f \in U^\mu} e_\mu(f, z, \delta) \leq c\delta^{\frac{\mu}{\mu+1}}$.

**Remark 3.1.** The above Tikhonov functional arises in Tikhonov regularization of the problem $A^*z = f$ in the space $X^{-\mu}$. It is however interesting to note that within this context the parameter choice $\alpha = \delta^2$ is not optimal, since $A^*z = f$ is ill-posed with exactly given right hand side.

We observe that in Hilbert scales the optimal data-functional strategy (11) automatically leads to Tikhonov regularization of the equation $A^*z = f$ in contrast to the mollifies method from [13], where the same equation is regularized by means of the least-square method.

4. THE OPTIMAL DATA-FUNCTIONAL STRATEGY
WHEN THE DESIGN IS GIVEN

Here we aim at studying Problem 1.B when a design $\Phi := \{\phi_1, ..., \phi_n\}$ is given. Throughout we assume that it consists of orthonormal (in $Y^0 = X$)
functionals. Thus each design induces an orthogonal projection \(Q_n\) in \(X\) by
\[Q_n x := \sum_{j=1}^{n} \langle x, \varphi_j \rangle \varphi_j, \quad x \in X.\]
Properties are given later in terms of Jackson-type inequalities.

Since the design \(\Phi\) is given beforehand, approximate solutions to Problem 1.B, using data (2) are described by (measurable) mappings
\(S: \mathbb{R}^n \to \mathbb{R}\) and the respective error is
\[e_{\mu}(f, S, \Phi, \delta) := \sup_{x \in U^\delta} (\mathbb{E} |\langle f, x \rangle - S(y_{\delta,1}, \ldots, y_{\delta,n})|^2)^{1/2}.\]

Since this setup is more restrictive as in Section 3, we conclude
\[\mathcal{E}_{\mu,\gamma}(\Phi, \delta) := \sup_{f \in U^\delta} \inf_{S: \mathbb{R}^n \to \mathbb{R}} e_{\mu}(f, S, \Phi, \delta) \geq \mathcal{E}_{\mu,\gamma}(\delta),\]
and the asymptotic behavior of the left-hand side \(\mathcal{E}_{\mu,\gamma}(\Phi, \delta)\) is determined by properties of the design \(\Phi\).

As in Section 3 we are going to find the optimal data-functional strategy, now restricted to those based on observations (2), which means minimizing
\[\sup_{x \in U^\delta} \mathbb{E} |\langle f, x \rangle - \langle z, Q_n y_{\delta} \rangle|^2 \to \text{MIN}(z).\] (12)

With an argument like in (11), which led to the representation (7), we arrive at the following representation for the unique minimizes, say \(z_n\), as
\[z_n = \arg \min \{ \|A^*Q_n z - f\|^2_{\mu} + \delta^2 \|Q_n z\|^2; \ z \in X \}
= (\delta^2 I + B_n B_n^*)^{-1} B_n L^{-\mu} f,\] (13)
where \(B_n := Q_n B = Q_n A L^{-\mu}\). This means that \(z_n\) solves the equation
\[\delta^2 z + B_n B_n^* z = B_n L^{-\mu} f,\] (14)
which arises as Euler equation for Galerkin methods applied to the Tikhonov functional \(\|A^*z - f\|^2_{\mu} + \delta^2 \|z\|^2\). In the following proposition we derive, as an intermediate step, error bounds for this data-functional strategy \(\langle z_n, y_{\delta} \rangle\) in terms of certain norm bounds, which describe approximation properties of the design.
The data-functional strategy \( \langle z_n, y_d \rangle \) is based on observations (2). Under assumptions (4) on the operator, we have the following error bound (with \( D \) from (4))

\[
\mathcal{E}_{\infty}(\Phi, \delta) \leq \sup_{f \in V^*} e_{\infty}(f, z_n, \delta)
\leq c\delta^{\frac{m+n}{a+m}} + \delta \| (I + Q_n) z \| (1 + \delta^{-1}D \| (I - Q_n) \|_{\tau^{*}, \tau^*}).
\] (15)

**Proof.** It follows just from the construction of \( z_n \), see (14), that the strategy \( \langle z_n, y_d \rangle \) is based on observations (2). For the proof of estimate (15) we auxiliary use the optimal data-functional \( z \) from (7) in Section 3. By construction of \( z_n \) we have

\[
\sup_{x \in U^n} (E |\langle f, x \rangle - \langle z_n, y_d \rangle|^2)^{1/2} \leq \sup_{x \in U^n} (E |\langle f, x \rangle - \langle Q_n z, y_d \rangle|^2)^{1/2}.
\]

The latter can be estimated by

\[
\sup_{x \in U^n} (E |\langle f, x \rangle - \langle Q_n z, y_d \rangle|^2)^{1/2}
\leq \sup_{x \in U^n} (E |\langle f, x \rangle - \langle z, y_d \rangle|^2)^{1/2} + \sup_{x \in U^n} (E |\langle z, y_d \rangle - \langle Q_n z, y_d \rangle|^2)^{1/2},
\]

with \( z \) from (7). The first term on the right is bounded by \( c\delta^{(m+n)/(a+m)} \) in view of Theorem 3.1. For each \( x \) the second term is estimated as

\[
(E |\langle z, y_d \rangle - \langle Q_n z, y_d \rangle|^2)^{1/2} \leq |\langle (I - Q_n) z, Ax \rangle| + \delta \| (I - Q_n) z \|
\leq \| (I - Q_n) z \| (\delta + \| (I - Q_n) Ax \|),
\]

completing the proof of the proposition after sup-ing over \( x \in U^n \). □

As can be drawn from (15), the quality of the design will determine how large its size has to be, in order to yield the best possible order of accuracy \( \delta^{(m+n)/(a+m)} \). This leads to studying sequences \( (\Phi_n)_{n \in \mathbb{N}} \) of designs.

For \( s > 0 \) we say that \( (\Phi_n)_{n \in \mathbb{N}} \) has power \( s \) if there is \( c < \infty \) such that

\[
\| (I - Q_n) \|_{\tau^*, \tau^n} \leq cn^{-s}, \quad n \in \mathbb{N}.
\] (16)

Since by assumption (3), \( a_n(J; Y^* \to Y^n) \approx n^{-s} \), any sequence \( (\Phi_n)_{n \in \mathbb{N}} \) of designs with power \( s \) achieves the best possible order of approximation of \( Y^* \) in \( Y \). We shall discuss this point in more detail in the example at the end of this section.
As can be seen from the bound in Proposition 4.1, the highest smoothness to be taken into account is \( s = a + \mu \). If \( (\Phi_n)_{n \in \mathbb{N}} \) has this maximal power, then we are able to prove the following

**Theorem 4.1.** Suppose that the operator \( A \) obeys (4) and that the sequence \( (\Phi_n)_{n \in \mathbb{N}} \) has power \( a + \mu \). Then the data-functional strategy \( \langle z_n, y_d \rangle \) satisfies for \( n \gg \delta^{-1/(a+\mu)} \) the estimate

\[
\delta_{\mu, a}(\Phi, \delta) \leq \sup_{f \in \mathcal{U}} e_{\mu}(f, z_n, \delta) \leq c\delta^{\frac{s}{2s+a}}. \tag{17}
\]

**Proof.** We use Proposition 3.1 to bound \( \delta \| (I - Q_n) z \| \leq \delta \| z \|_0 \leq c\delta^{(\mu+\nu)/(a+\mu)} \). By the above choice of \( n \), the expressions \( \delta^{-1} \| (I - Q_n) \|_{\infty, \ldots, \infty} \) are uniformly bounded, such that the estimate from Proposition 4.1 completes the proof of the theorem.

Next we shall discuss the situation when the given sequence of designs \( (\Phi_n)_{n \in \mathbb{N}} \) allows approximation as in (16) for some \( s < a + \mu \). We will analyze the error of the strategy \( z_n \) from (13) in this special situation. As we will see, the best possible accuracy still can be achieved for such designs, but only on account of a larger size of it.

**Theorem 4.2.** Under assumptions (4) on the operator and (16) on the design, the data-functional strategy \( \langle z_n, y_d \rangle \) from (13) realizes for \( n \gg \delta^{-(a+\mu)/(2a+\mu)} \) the estimate \( \sup_{f \in \mathcal{U}} e_{\mu}(f, z_n, \delta) \gg \delta^{(\mu+\nu)/(a+\mu)} \).

**Proof.** The proof will follow from the bound provided in Proposition 4.1. Since for \( (\Phi_n)_{n \in \mathbb{N}} \) only estimate (16) is available, we can rely only on \( \| (I - Q_n) z \| \leq c n^{-1} \| z \|_0 \) and \( \| (I - Q_n) \|_{\infty, \ldots, \infty} \ll c n^{-1} \). By the choice of \( n \) and the bound from Proposition 3.1 on \( \| z \|_0 \), we arrive at

\[
\delta^{-1} \| (I - Q_n) \|_{\infty, \ldots, \infty} \ll c \delta^{\frac{s+\mu-\nu}{2s+a+\mu}}
\]

and also

\[
\| (I - Q_n) z \| \ll c \delta^{\frac{2s+\mu-\nu}{2s+a+\mu}}.
\]

Inserting these estimates into (15) yields the order of accuracy \( \delta^{(\mu+\nu)/(a+\mu)} \), accomplishing the proof of the theorem.

As we will see in the next section, the asymptotics \( n \gg \delta^{-1/(a+\mu)} \) from Theorem 4.1 is sharp. We do not know whether the asymptotics for \( n \) from Theorem 4.2 is also sharp.
We complete this section with an

**Example 1.** We consider the problem of estimating the value of some function $x(t)$, $t \in (0, 1)$ at some point $t_0 \in (0, 1)$ based on noisy observations (1) for the identity operator $A: L^2(0, 1) \to L^2(0, 1)$. We assume that the unknown function $x$ belongs to $W^\mu$, for some $\mu > 1/2$, which for integer $m > 0$ is

$$W^\mu = \{ x \in L^2(0, 1), x^{(\mu-1)} \text{ is absolutely cont., } x^{(\mu)} \in L^2(0, 1) \},$$

and boundary conditions $x^{(2l)}(0) = x^{(2l)}(1) = 0$, $l = 0, \ldots, \lfloor \frac{\mu-1}{2} \rfloor$, where $\lfloor u \rfloor$ denotes the largest integer less than $u$.

In general $W^\mu$ belongs to the scale $\{ W^s \}_{s \in \mathbb{R}}$ of Sobolev Hilbert spaces, generated from $L^2(0, 1)$ by the operator $Lx = \sum_{j=1}^\infty a_j \langle x, e_j \rangle e_j$, $x \in L^2(0, 1)$ with basis $e_j(t) = \sqrt{2} \sin(\pi j t)$.

The optimal data-functional strategy $z$ as given in (7) takes the form

$$\langle z, y_d \rangle = \sum_{k=1}^\infty \frac{(\pi k)^{-2a}}{\delta^2 + (\pi k)^{-2a}} \langle z, y_d \rangle e_k(t_0).$$

Being an infinite sum this is numerically not feasible. Its truncation at some $n$ corresponds to $z_n$ from (13) with design $\{ e_1(t), \ldots, e_n(t) \}$. We note that this Fourier design has (maximal) power $\mu$; the choice of $n \asymp \delta^{\frac{1}{2}}$ guarantees best possible accuracy, independently of $n$, for which the functional $\langle f_0, x \rangle := x(t_0)$ belongs to $W^\mu$; but the order of the accuracy is determined by $n$. Function evaluation does not directly fit to the Hilbert scale setup. We can however guarantee $f_0 \in W^\mu$ whenever $\nu < -1/2$, yielding accuracy $\delta^{\frac{1}{2} + \frac{\nu+1}{\mu}}$ for any $\epsilon > 0$. It is interesting to note, that for this particular problem of function evaluation, the data-functional strategies $\langle z, y_d \rangle$ and $\langle z_n, y_d \rangle$ from above can be seen to provide a slightly better accuracy $\delta^{\frac{1}{2} + \frac{\nu+1}{\mu}}$, which was proven by Ibragimov and Khas’minskii [11].

Now, if the observations are based on noisy histograms rather than Fourier coefficients, i.e.,

$$y_{d,i} = \sqrt{n} \int_{(i-1)/n}^{i/n} x(s) \, ds + \delta \xi_i, \quad i = 1, \ldots, n,$$

with i.i.d. standard normal variates $\xi_i$, then we can guarantee power $s = 1$ only. For $\mu \leq 1$ (and $a = 0$) this design has maximal power and the same cardinality as for the Fourier design allows best possible accuracy. For $\mu > s = 1$ the estimate from Theorem 4.2 provides the same accuracy as above, but for a larger size $n \asymp \delta^{-(\mu+1)/2\mu}$ of the design.
5. COMPLEXITY ISSUES

We have seen in Section 3 that for solving Problem 1.A one linear functional on the observations suffices to achieve the optimal order of accuracy \( \delta^{(\mu + \nu)/(a + \rho)} \). In Section 4 we studied the case when the design \( \Phi \) is given independently of the functional \( f \) to be computed. Theorem 4.1 implies that the same accuracy \( \delta^{(\mu + \nu)/(a + \rho)} \) can be achieved, but the size of the design was of order \( n \approx \delta^{-1/(a + \rho)} \). Problem 1.C therefore raises the question, which amount of information is necessary to ensure the best possible order of accuracy. In mathematical terms this results in the study of the

\[
\begin{align*}
\rho_n(\delta) &= \inf_{\text{Card}(\Phi) \leq n} \mathcal{E}_{\mu, \nu}(\Phi, \delta),
\end{align*}
\]

and in the limiting case \( \rho(\delta) := \lim_{n \to \infty} \rho_n(\delta) \). As in [17] we introduce the following quantity for any fixed \( 1 < C < \infty \).

\[
\mathcal{N}(\delta) := \inf \{ n : \rho_n(\delta) \leq C\rho(\delta) \},
\]

called information complexity. The study of information complexity, but within the framework of classical (well-posed) problems, is fundamental in Information-Based Complexity [24]. Our variant as presented in (18) reflects the particular circumstance that \( \rho_n(\delta) \) will not converge to 0, as this is typically the case for problems studied so far; see [24].

Our previous analysis readily provides us with the exact asymptotics of \( \rho(\delta) \) as \( \delta \to 0 \). First we note that Theorem 3.1 can be seen as recovering \( \langle f, x \rangle \) with complete information on \( y \), which implies \( \rho(\delta) \geq c\delta^{(\mu + \nu)/(a + \rho)} \). On the other hand, Theorem 4.1 shows that this order can be achieved with some design, given beforehand, thus actually \( \rho(\delta) \approx \delta^{(\mu + \nu)/(a + \rho)} \). Theorem 4.1 also yields an upper bound on \( \mathcal{N}(\delta) \), since it provides a method of optimal accuracy, which uses asymptotically \( n \approx \delta^{-1/(a + \rho)} \) functionals. Thus if \( C \) from (18) is large enough, which means, larger than the quotient, say \( \tilde{C} > 0 \), of the respective upper bound from Theorem 4.1 by the lower bound from Theorem 3.1, then

\[
\mathcal{N}(\delta) \leq \inf \{ n : \rho_n(\delta) \leq \tilde{C}\delta^{\frac{\mu + \nu}{a + \rho}} \} \leq c\delta^{-1/(a + \rho)}.
\]  

Below we will establish that this is the right order as \( \delta \to 0 \).

**Theorem 5.1.** Let us assume (4) on the operator and a priori smoothness \( \mu \) of the solution. Then if in the definition of \( \mathcal{N}(\delta) \) the constant \( C \) is larger than \( \tilde{C} \) from above, then we have \( \mathcal{N}(\delta) \approx \delta^{-1/(a + \rho)} \), independently of the smoothness \( \nu \) of the solution-functional \( f \).
The above asymptotics will follow from

**Proposition 5.1.** \( r_n(\delta) \geq c(n^{-\mu+\nu} + \delta^{(\mu+\nu)/(\alpha+\beta)}) \).

**Proof.** Given any collection \( \mathcal{Y} = \{\psi_1, \ldots, \psi_n\} \subset X \) and \( \lambda \geq 0 \) we let

\[
U^\lambda(\mathcal{Y}) := \{x \in U^\lambda, \langle \psi_i, x \rangle = 0, i = 1, \ldots, n\}.
\]

We also abbreviate \( A^*\Phi = \{A^*\phi_i, i = 1, \ldots, n\} \) for any design \( \Phi \). With this notation we have

\[
\delta_{\mu, n}(\Phi, \delta) \geq \sup_{f \in U^\delta} \inf_{x \in U^{\delta}(A^*\Phi)} (E |\langle f, x \rangle - S(\langle \phi_i, \delta z \rangle)|^2)^{1/2}
\]

\[
\geq \sup_{f \in U^\delta} \inf_{x \in U^{\delta}(A^*\Phi)} |\langle f, x \rangle|,
\]

since \( S(\langle \phi_i, \delta z \rangle) \) does not depend on \( x \) and \( U^{\delta}(A^*\Phi) \) is centrally symmetric. Interchanging the sup-s above implies

\[
\delta_{\mu, n}(\Phi, \delta) \geq \sup_{x \in U^{\delta}(A^*\Phi)} \|x\|_{\lambda,n},
\]

and consequently

\[
r_n(\delta) = \inf_{\text{Card}(\Phi) \leq n} \delta_{\mu, n}(\Phi, \delta) \geq \inf_{\text{Card}(\mathcal{Y}) \leq n} \sup_{x \in U^{\delta}(A^*\Phi)} \|x\|_{\lambda,n}
\]

\[
\geq c_{n+1}(J^{-\mu}: X^\mu \to X^{-\mu}),
\]

where the last quantity is just the \((n+1)\)st Gelfand number of the embedding \( J^{-\mu}: X^\mu \to X^{-\mu} \); see, e.g., [21]. Since in Hilbert spaces all \( s \)-numbers coincide, we arrive at

\[
c_{n+1}(J^{-\mu}: X^\mu \to X^{-\mu}) = a_{n+1}(J^{-\mu}: X^\mu \to X^{-\mu})
\]

\[
= a_{n+1}(J_{\mu^{-\mu}}: X^{\mu+\nu} \to X) \approx n^{-\mu+\nu}.
\]

Thus \( r_n(\delta) \geq cn^{-\mu+\nu} \). Since on the other hand \( r_n(\delta) \geq r(\delta) \geq c\delta^{(\mu+\nu)/(\alpha+\beta)} \) by the reasoning above, the proof of the proposition is complete. \( \square \)

We turn to the

**Proof of Theorem 5.1.** By the definition of \( \mathcal{N}(\delta) \) and the lower bound in Proposition 5.1 we conclude

\[
\mathcal{N}(\delta) \geq \inf\{n: (n^{-\mu+\nu} + \delta^{(\mu+\nu)/(\alpha+\beta)}) \leq c\delta^{(\mu+\nu)/(\alpha+\beta)}\},
\]
which can easily be resolved to \( \mathcal{N}(\delta) \geq c\delta^{-1/(a+\rho)} \). Together with the respective upper bound (19) this completes the proof of the theorem.

6. SUMMARY

As already pointed out, the advantage of the data-functional strategies (7) and (13), respectively, consists in the possibility of precomputing the data-functional, regardless of later observations. The strategies attain the best possible accuracy only when the smoothness properties are known and can be reflected within the Hilbert scale framework. One particular example for this is satellite geodesy, where the smoothness of the gravitational potential is known to be 3/2 with respect to the Hilbert scale of spherical Sobolev spaces; see [6]. However, the optimal data-functional strategy cannot adapt to unknown smoothness. If this is the case, then the solution-functional strategy is an alternative, since it allows to achieve, up to logarithmic factors, the same order of accuracy, applying an adaptive procedure, as investigated in [9].

We also mention that all results are valid for bounded deterministic noise when the individual error \( e_{\mu}(f, S, \delta) \) is replaced by

\[
e_{\mu}^{\text{det}}(f, S, \delta) := \sup_{x \in \mathbb{U}^n} \sup_{|\xi| \leq 1} |\langle f, x \rangle - S(y_\delta)|,
\]

and proceeding for the deterministic versions of \( \delta_{\mu, \nu}(\delta) \), \( \delta_{\mu, \nu}(\Phi, \delta) \), \( r_\mu(\delta) \) and \( \mathcal{N}(\delta) \), analogously. In particular, the required lower bound for \( \delta_{\mu, \nu}^{\text{det}}(\delta) \) follows from [18]. For deterministic noise the error cannot be exactly written in terms of a Tikhonov functional as in Section 3, instead there is only an estimate, which is sufficient for upper bounds.

The case of deterministic noise is very convenient for a discussion of the efficiency of \( z_n \) when the design \( \Phi \) does not have maximal power. In this framework, it is tempting to choose the size \( n \) of the design, such that \( \|y - Q_n y_\delta\| \) is of the same order as \( \delta \), because it allows to keep the noise level after discretization. This would require \( n \asymp \delta^{-1/\gamma} \), since

\[
\sup_{y - Ax \|y - y_\delta\| < \delta} \sup_{x \in \mathbb{U}^\rho} \|y - Q_n y_\delta\| \geq \sup_{y - Ax \|y - y_\delta\| < \delta} \sup_{x \in \mathbb{U}^\rho} \|y - Q_n (y - y_\delta)\| \geq c n^{-s} - \delta,
\]

provided that the power of \( \Phi \) cannot be improved. For a design with maximal power it is indeed true that \( \|y - Q_n y_\delta\| \) is of the order \( \delta \); this leads to a strategy with minimal amount of information \( \delta^{-1/(a+\rho)} \). For a design...
with power $s < \mu + a$, as stated in Theorem 4.2, a size $n \asymp \delta^{-\frac{(a+\mu+s)}{2s(a+\mu)}}$ is sufficient, which means that this method is more economical. By this choice of $n$ we easily bound

$$\sup_{y \in \mathcal{D}} \sup_{x \in U^a} \|y - Q_n y\| \geq cn^{\frac{a}{2s(a+\mu)}}$$

which is less accurate than $\delta$ as long as $s < a + \mu$.

REFERENCES


