JOURNAL OF INTEGRAL EQUATIONS
AND APPLICATIONS
Volume 14, Number 3, Fall 2002

METRIC AND PROBABILISTIC INFORMATION ASSOCIATED WITH FREDHOLM INTEGRAL EQUATIONS OF THE FIRST KIND

ENRICO DE MICHELI AND GIOVANNI ALBERTO VIANO

ABSTRACT. The problem of evaluating the information associated with Fredholm integral equations of the first kind, when the integral operator is self-adjoint and compact, is considered here. The data function is assumed to be perturbed *gently* by an additive noise so that it still belongs to the range of the operator. First we estimate upper and lower bounds for the ε-capacity (and then for the *metric information*), and explicit computations in some specific cases are given; then the problem is reformulated from a probabilistic viewpoint and use is made of the probabilistic information theory. The results obtained by these two approaches are then compared.

1. Introduction. Let us consider the following class of Fredholm integral equations of the first kind:

$$Af = g, \tag{1.1}$$

where $A : X \to Y$ is a self-adjoint compact operator, X and Y being the solution and the data space, respectively. Hereafter we set $X = Y = L^2[a, b]$.

Solving Equation (1.1) presents two problems:

a) The Range (A) is not closed in the data space Y. Therefore, given an arbitrary function $g \in Y$, it does not follow necessarily that there exists a solution $f \in X$.

b) Even if two data functions g_1 and g_2 belong to Range (A), and their distance in Y is small, nevertheless the distance between $A^{-1}g_1$ and $A^{-1}g_2$ can be unlimitedly large, in view of the fact that the inverse of the compact operator A is not bounded (X and Y being infinite dimensional space).

In the numerical applications, g is perturbed by a noise n which can represent either round-off numerical error or measurement error if g describes experimental data. Assuming in both cases that the perturbation produced by the noise is additive, the data function actually known is $\bar{g} = g + n$ (instead of the noiseless data function g). Then, in order to recover f one is forced to use the so-called *regularization methods*; the literature on these topics is very extensive, and we shall return later on this point.

Since the operator A is self-adjoint it admits a set of eigenfunctions $\{\psi_k\}_1^\infty$ and, accordingly, a countably infinite set of eigenvalues $\{\lambda_k\}_1^\infty$. The eigenfunctions form an orthonormal basis of the orthogonal complement of the null space of the operator A, and therefore an orthonormal basis of $L^2[a,b]$ when A is injective. For the sake of simplicity we consider hereafter only this case. The Hilbert-Schmidt theorem guarantees that $\lim_{k\to\infty} \lambda_k = 0$. We shall suppose hereafter that the eigenvalues are ordered as follows: $\lambda_1 > \lambda_2 > \lambda_3 > \cdots$; furthermore, we assume for simplicity that they are bounded by 1: i.e., $\lambda_1 \le 1$. If we consider the noiseless data function g, we can associate to the integral equation (1.1) the eigenfunction expansion

$$f(x) = \sum_{k=1}^{\infty} \frac{g_k}{\lambda_k}\, \psi_k(x), \tag{1.2}$$

where $g_k = (g, \psi_k)$, $(\cdot,\cdot)$ denoting the scalar product in $L^2[a,b]$. The series (1.2) converges in the sense of the L^2-norm. Unfortunately this series is not useful since, in practice, the noiseless data function g is unknown. If we take into account the additive noise n, instead of Equation (1.1), we have

$$Af + n = \bar{g}. \tag{1.3}$$

Therefore, instead of expansion (1.2), we have to deal with an expansion of the type

$$\sum_{k=1}^{\infty} \frac{\bar{g}_k}{\lambda_k}\, \psi_k(x), \quad \bar{g}_k = (\bar{g}, \psi_k), \tag{1.4}$$

which either diverges if $\bar{g} \notin \text{Range}\,(A)$, or converges to a function whose distance in norm from the true solution f (corresponding to the noiseless data) can be quite large. One is then forced to use regularization procedures as mentioned above.

The mathematical framework outlined so far is only a schematization of reality; in particular, if the data g describes experimental data, then it obviously will be an element of a finite dimensional space, while the solution f can still be considered an element of an infinite-dimensional function space; in general, the data space Y and the solution space X may differ. In this case the analysis would require the use of singular values and singular functions of the operator A [**2, 18**], instead of the eigenvalues λ_k and eigenvectors ψ_k. For the sake of clarity, here it is convenient to identify data with an element g of $L_2[a,b]$ and deal with a self-adjoint operator A; in this way, the analysis is technically simpler and becomes more transparent for our purposes.

Several methods of regularization have been proposed [**4, 9**]: all of them modify one of the elements of the triplet $\{A, X, Y\}$ [**18**]. Among these methods, the procedure which is probably the most popular consists in admitting only those solutions which belong to a compact subset of the solution space X. The key theorem used in this method reads as follows: let σ be a continuous map from a compact topological space into a Hausdorff topological space; if σ is one-to-one, then its inverse map σ^{-1} is continuous [**11**]. The condition of compactness can be realized by the use of a priori bounds [**14, 22**], which require some prior knowledge or some constraints on the solution. Then the procedure works by taking into account two bounds, one on the solutions and one on the noise n:

$$\|Bf\|_X \leq 1, \tag{1.5}$$

$$\|n\|_Y \leq \varepsilon, \tag{1.6}$$

where B is a suitable constraint operator. Let us suppose that the eigenfunctions $\{\psi_k\}_1^\infty$ diagonalize the operator B^*B, i.e., A^*A and B^*B commute. In such a case we have $B^*Bf = \sum_{k=1}^\infty \beta_k^2 f_k \psi_k$, where $f_k = (f, \psi_k)$, and β_k^2 are the eigenvalues of B^*B. The constraint operator B has compact inverse if and only if $\lim_{k\to\infty} \beta_k^2 = +\infty$; under such a condition, the solution obtained by truncating expansion (1.4) at the largest integer k such that $\lambda_k \geq \varepsilon\beta_k$, converges to the solution f, as $\varepsilon \to 0$, in the sense of the L^2-norm. In several cases a much milder constraint can be conveniently used, i.e., $B = I$ for all k, $\beta_k = 1$. In this case the compactness condition, required by the theorem quoted above, is not satisfied; however, we shall prove in Section 2 that the approximation f_* obtained by truncating expansion (1.4) at the largest

k such that $\lambda_k \geq \varepsilon$ is convergent, though in weak sense, to the solution f as $\varepsilon \to 0$.

Hereafter we shall only consider this last truncation method, and we denote by $k_0(\varepsilon)$ the largest integer k such that $\lambda_k \geq \varepsilon$; further, we assume that $\bar{g} \in \text{Range}\,(A)$. Since A is compact, $Y_0 \equiv \text{Range}\,(A)$ is a compact subset of Y, and then finite coverings of Y_0 can be constructed. By adopting the language of the communication theory [**17**], and regarding the inverse problem of approximating f from a given $\bar{g}$ as a *communication channel* problem, one can compute the *maximal length* of the messages conveyed back from $\bar{g}$ to f. We are thus led to find a relationship between the maximal length of these messages, which is related to the truncation number $k_0(\varepsilon)$, and the *massiveness* (or *degree of compactness*) of the set Y_0. It turns out that the *degree of compactness* of Y_0 is related to the *smoothness* of the kernel of the integral operator A. In fact, the asymptotic behavior of the eigenvalues λ_k, for large k, is strictly related to the regularity properties of the kernel: Hille and Tamarkin [**10**] have systematically explored the relationship between the regularity properties of the kernel and the distribution of the eigenvalues of the Fredholm integral equation of the first kind. We can say that, as the regularity of the kernel increases, e.g., passing from the class of functions C^0 to C^∞ and then to the class of analytic functions, the eigenvalues λ_k decrease more and more rapidly for $k \to \infty$. Thus the minimum number of balls in a covering of Y_0, or the maximum number of balls in a packing of Y_0 [**20**], which give a numerical estimate of the *degree of compactness* of Y_0, decreases as the *smoothness* of the kernel increases. Finally, the type of restored continuity in reconstructing f from a given $\bar{g}$ depends on the *a priori* global bounds imposed on the solution (see formula (1.5)), and also on the *degree of compactness* of Y_0 and, accordingly, it is related to the length of the messages conveyed back from $\bar{g}$ to reconstruct f. Since we are concerned with the maximal length of these messages, we are led to consider a weak-type convergence in the reconstruction of the solution f; accordingly, we will define $k_0(\varepsilon)$ as the largest integer such that $\lambda_k \geq \varepsilon$. By adopting a more restrictive constraint we could achieve strong-type convergence, but at the same time we would have shorter messages conveyed back from $\bar{g}$ for reconstructing f.

The problem of reconstructing f from $\bar{g}$ can be reformulated as well in probabilistic terms, in view of the fact that the data function g

is perturbated by the noise n, which can be properly regarded as a random variable. With this in mind one can rewrite equation (1.3) in probabilistic form as

$$A\xi + \zeta = \eta, \tag{1.7}$$

where ξ, ζ and η, which correspond to f, n and $\bar{g}$ respectively, are Gaussian weak random variables [**1**] in the Hilbert space $L^2[a,b]$. Next, Equation (1.7) can be turned into an infinite sequence of one-dimensional equations by means of orthogonal projections, i.e.,

$$\lambda_k \xi_k + \zeta_k = \eta_k, \quad k = 1, 2, \dots , \tag{1.8}$$

where $\xi_k = (\xi, \psi_k)$, $\zeta_k = (\zeta, \psi_k)$, $\eta_k = (\eta, \psi_k)$ are Gaussian random variables. Using this approach it is possible to evaluate the amount of information $J(\xi_k, \eta_k)$ about the variable ξ_k, which is contained in the variable η_k. From this approach then another method of truncation emerges, which is based on neglecting all those components for which $J(\xi_k, \eta_k)$ is less than $(1/2)\ln 2$. As illustrated in Section 3, this criterion leads to a truncation number which is very close to the number $k_0(\varepsilon)$ introduced previously. One can thus conclude that the two procedures, the deterministic one, based on the evaluation of the maximal length of the messages conveyed back from $\bar{g}$ to f, and the probabilistic one, based on the information theory, yield essentially the same result.

Information theory, or the theory of coding arose from the fundamental paper of Shannon in 1948 [**21**]. Perhaps it should be more correctly referred to as statistical communication theory. The information source is any producer of information according to some known probability law, and this information has to be communicated to the destination by means of a transmission channel. Noise can be regarded as anything which impairs the ability of the channel to transmit with complete reliability. Information theory is concerned with the methods for achieving high reliability without reducing the transmission rate too drastically. Successively the mathematical theory of information was extended by several authors, notably Kolmogorov and Gelfand (see, in particular, [**8**] and the papers quoted therein). One question quite naturally arises: On the one hand information theory is formulated in the framework and uses language and tools of the probability theory, on the other hand the concept of *information* can be thought of as more basic and independent of probability [**13**]. Then the problem becomes how to construct a nonprobabilistic theory of information. To this purpose Kolmogorov and his school introduced and developed an alternative approach to the quantitative definition of information, which is logically independent of probabilistic assumptions: the measure of information is given in purely combinatorial terms [**13**]. This combinatorial, or metric, approach finally results in the theory of the ε-entropy and ε-capacity of sets in metric spaces [**12**].

The connection between ideas and concepts of Shannon's information theory, with particular attention to the notion of *length of a message in binary units*, and those of ε-entropy and ε-capacity are illustrated in detail in [**12**], to which the interested reader is referred (to this purpose, let us also mention [**23**], where the ε-entropy plays a crucial role in connection with empirical processes estimation). With a small abuse of language we call *metric information* that induced by the ε-capacity, which is, indeed, defined as the number of binary signs that can be reliably transmitted. Finally, the problem of comparing the results of probabilistic and nonprobabilistic, or metric, information theory remains. The main aim of this paper consists precisely in trying to give a partial answer to this question in the specific case of Fredholm integral equations of the first kind.

The paper is organized as follows. In Section 2 we first prove that the approximation f_* converges weakly to f as $\varepsilon \to 0$. Then we find an upper and a lower bound for the ε-entropy associated with the mapping of the unit ball, in the solution space, induced by the operator A. Next, we evaluate explicitly an upper bound for the maximal length of the messages conveyed back from $\bar{g}$ to reconstruct f, and this provides an estimate of what we call *metric information*. Explicit calculations are given in three specific cases: harmonic continuation, backward solution of the heat equation, first kind Fredholm integral equation with continuous kernel. In Section 3 we reconsider the problem from a probabilistic viewpoint. We introduce another truncation method based on probabilistic information theory, and accordingly we derive an approximation which converges to the solution, in the sense of the probabilistic theory, under suitable conditions on the covariance operator of the solution.

2. Metric information associated with Fredholm integral equations of the first kind.

A. *Weak convergence of the f_* approximation.* Let us consider the approximation $f_* = \sum_{k=1}^{k_0(\varepsilon)} (\bar{g}_k/\lambda_k)\psi_k$ where $k_0(\varepsilon)$ is the largest integer such that $\lambda_k \geq \varepsilon$. We want to prove the weak convergence of f_* to f as $\varepsilon \to 0$ and, accordingly, the *weak continuity* in the restored solution; for this purpose we need the following auxiliary lemma.

Lemma 1. *For any function f which satisfies the following bounds*

$$\|Af - \bar{g}\|_{Y \equiv L^2[a,b]} \leq \varepsilon, \tag{2.1}$$

$$\|f\|_{X \equiv L^2[a,b]} \leq 1, \tag{2.2}$$

the following inequalities hold:

$$\|A(f - f_*)\|_Y \leq \sqrt{2}\varepsilon, \tag{2.3}$$

$$\|f - f_*\|_X \leq \sqrt{2}, \tag{2.4}$$

$$\|A(f - f_*)\|_Y^2 + \varepsilon^2 \|f - f_*\|_X^2 \leq 4\varepsilon^2. \tag{2.5}$$

Proof. a) From the inequality $\lambda_k < \varepsilon$ for $k > k_0$ and the bound $\|f\|_X \leq 1$ it follows:

$$\sum_{k=k_0+1}^{\infty} \lambda_k^2 |f_k|^2 < \varepsilon^2. \tag{2.6}$$

From $\|Af - \bar{g}\|_Y \leq \varepsilon$ we get:

$$\sum_{k=1}^{k_0} \lambda_k^2 \left| f_k - \frac{\bar{g}_k}{\lambda_k} \right|^2 \leq \varepsilon^2. \tag{2.7}$$

Therefore we have

$$\|A(f - f_*)\|_Y^2 = \sum_{k=1}^{k_0} \lambda_k^2 \left| f_k - \frac{\bar{g}_k}{\lambda_k} \right|^2 + \sum_{k=k_0+1}^{\infty} \lambda_k^2 |f_k|^2 \leq 2\varepsilon^2, \tag{2.8}$$

and inequality (2.3) is proved.

b) From the inequality $\lambda_k \geq \varepsilon$ for $k \leq k_0$ and the bound $\|Af - \bar{g}\|_Y \leq \varepsilon$ we obtain

$$\sum_{k=1}^{k_0} \left| f_k - \frac{\bar{g}_k}{\lambda_k} \right|^2 = \sum_{k=1}^{k_0} \frac{1}{\lambda_k^2} \, |\lambda_k f_k - \bar{g}_k|^2 \leq 1. \tag{2.9}$$

From $\|f\|_X \leq 1$ it follows:

$$\sum_{k=k_0+1}^{\infty} |f_k|^2 \leq 1. \tag{2.10}$$

Therefore we have:

$$\|f - f_*\|_X^2 = \sum_{k=1}^{k_0} \left| f_k - \frac{\bar{g}_k}{\lambda_k} \right|^2 + \sum_{k=k_0+1}^{\infty} |f_k|^2 \leq 2, \tag{2.11}$$

and inequality (2.4) is proved. Next, from (2.8) and (2.11) we obtain:

$$\|A(f - f_*)\|_Y^2 + \varepsilon^2 \, \|f - f_*\|_X^2 \leq 4\varepsilon^2, \tag{2.12}$$

that is, inequality (2.5). □

Let us note that $\lim_{\varepsilon \to 0} k_0(\varepsilon) = +\infty$. The latter equality follows from the definition itself of $k_0(\varepsilon)$ and from the fact that $\lim_{k\to\infty} \lambda_k = 0$. Next we prove the following theorem.

Theorem 1. *For any function f which satisfies bounds* (2.1) *and* (2.2)*, the following limit holds true*:

$$\lim_{\varepsilon \to 0} \, (f - f_*, v)_X = 0, \quad \forall\, v \in X; \|v\|_X \leq 1. \tag{2.13}$$

Proof. Let us put: $x_k = f_k - (f_*)_k$; then we have:

$$(f - f_*, v)_X = \sum_{k=1}^{\infty} x_k v_k, \quad \left(\sum_{k=1}^{\infty} |v_k|^2 \leq 1 \right). \tag{2.14}$$

Next, by the Schwarz inequality and bound (2.5), we have:

$$
\begin{aligned}
|(f-f_*,v)_X| &\le \sum_{k=1}^{\infty} |x_k v_k| = \sum_{k=1}^{\infty} \left(\frac{\lambda_k^2+\varepsilon^2}{\lambda_k^2+\varepsilon^2} \right)^{1/2} |x_k v_k| \\
&\le \left(\|A(f-f_*)\|_Y^2 + \varepsilon^2 \, \|f-f_*\|_X^2 \right)^{1/2} \\
&\quad \cdot \left(\sum_{k=1}^{\infty} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2} \right)^{1/2} \le \left(4\varepsilon^2 \sum_{k=1}^{\infty} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2} \right)^{1/2}.
\end{aligned}
\tag{2.15}
$$

Next we split the sum $\sum_{k=1}^{\infty} |v_k|^2/(\lambda_k^2+\varepsilon^2)$ into two parts, i.e.,

$$
\sum_{k=1}^{k_0} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2} + \sum_{k=k_0+1}^{\infty} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2}. \tag{2.16}
$$

The first term of the sum (2.16) can be majorized as follows:

$$
\sum_{k=1}^{k_0} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2} \le \frac{1}{2\varepsilon^2} \sum_{k=1}^{\infty} |v_k|^2 \le \frac{1}{2\varepsilon^2}. \tag{2.17}
$$

From formulae (2.15) and (2.17) we have

$$
4\varepsilon^2 \sum_{k=1}^{k_0} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2} \le 2 \sum_{k=1}^{\infty} |v_k|^2 \le 2. \tag{2.18}
$$

Furthermore, $\lim_{\varepsilon\to 0} \varepsilon^2 |v_k|^2/(\lambda_k^2+\varepsilon^2) = 0$ for $k \le k_0$. Therefore we have

$$
\lim_{\varepsilon\to 0} 4\varepsilon^2 \sum_{k=1}^{k_0} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2} = 0. \tag{2.19}
$$

Let us now consider the second term of sum (2.16); we can write

$$
\sum_{k=k_0+1}^{\infty} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2} \le \frac{1}{\varepsilon^2} \sum_{k=k_0+1}^{\infty} |v_k|^2. \tag{2.20}
$$

Therefore from formulae (2.15) and (2.20) we get

$$4\varepsilon^2 \sum_{k=k_0+1}^{\infty} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2} \leq 4 \sum_{k=k_0+1}^{\infty} |v_k|^2. \tag{2.21}$$

Then, taking into account that $\lim_{\varepsilon\to 0} k_0(\varepsilon) = +\infty$, we can conclude:

$$\lim_{\varepsilon\to 0} 4\varepsilon^2 \sum_{k=k_0+1}^{\infty} \frac{|v_k|^2}{\lambda_k^2+\varepsilon^2} = 0. \tag{2.22}$$

From (2.19) and (2.22) we then obtain:

$$\lim_{\varepsilon\to 0} (f - f_*, v)_X = 0, \quad \forall\, v \in X; \|v\|_X \leq 1, \tag{2.23}$$

and the theorem is proved. □

B. ***ε*-entropy and *ε*-Capacity associated with the operator *A*.** Let us consider the unit ball in the solution space $X \equiv L^2[a,b]$, i.e., $\{f \in X \mid \|f\|_X \leq 1\}$. The operator A maps the unit ball onto a compact ellipsoid $\mathcal{E} \in \text{Range}\,(A)$ contained in $Y \equiv L^2[a,b]$, whose semi-axes' lengths are the eigenvalues λ_k of the operator A. In order to give a numerical estimate of the *massiveness* of the set $\mathcal{E}$, let us first recall some basic definitions [**12, 16**]:

(a) A family $Y_1, \cdots, Y_n$ of subsets of Y is an ε-covering of $\mathcal{E}$ if the diameter of each Y_k does not exceed 2ε and if the sets Y_k cover $\mathcal{E}$: $\mathcal{E} \subset \cup_{k=1}^n Y_k$.

(b) Points $y_1, \cdots, y_m$ of $\mathcal{E}$ are called ε-distinguishable if the distance between each two of them exceeds ε.

Since $\mathcal{E}$ is compact, then a finite ε-covering exists for each $\varepsilon > 0$ and, moreover, $\mathcal{E}$ can contain only finitely many ε-distinguishable points. For a given $\varepsilon > 0$, the number n of sets Y_k in a covering family depends on the family, but the minimal value of n, $N_\varepsilon(\mathcal{E}) = \min n$, is an invariant of the set $\mathcal{E}$, which depends only on ε. Its logarithm (throughout the paper $\log x$ will always denote the logarithm of the number x to the base 2), that is the function $H_\varepsilon(\mathcal{E}) = \log N_\varepsilon(\mathcal{E})$ is the ε-entropy of the set $\mathcal{E}$. Analogously, the number m in definition (b) depends on the choice of points, but its maximum $M_\varepsilon(\mathcal{E}) = \max m$

is an invariant of the set $\mathcal{E}$. Its logarithm, that is the function $C_\varepsilon(\mathcal{E}) = \log M_\varepsilon(\mathcal{E})$ is called the ε-capacity of the set $\mathcal{E}$. This quantity represents the maximum number of ε-distinguishable signals that can be received, that is, those data which satisfy the following inequalities $\|\bar{g}^{(i)} - \bar{g}^{(k)}\|_Y > \varepsilon$, for all $i \neq k$, $\bar{g}^{(i)}$, $\bar{g}^{(k)} \in \mathcal{E}$.

A general result about ε-entropy and ε-capacity are the following inequalities [**16**]:

$$H_\varepsilon(\mathcal{E}) \leq C_\varepsilon(\mathcal{E}) \leq H_{\varepsilon/2}(\mathcal{E}). \tag{2.24}$$

To obtain estimates for the ε-capacity $C_\varepsilon(\mathcal{E})$, our aim now is to look for a lower bound for $H_\varepsilon(\mathcal{E})$ and an upper bound for $H_{\varepsilon/2}(\mathcal{E})$. For this purpose, let us consider the finite dimensional subspace Y_{k_0} of Y spanned by the first k_0 axes of $\mathcal{E}$, and put $\mathcal{E}_{k_0} = \mathcal{E} \cap Y_{k_0}$. Then $\mathcal{E}_{k_0}$ is a finite dimensional ellipsoid whose volume is just $\prod_{k=1}^{k_0} \lambda_k$ times the volume Ω_{k_0} of the unit ball in Y_{k_0}. Since the volume of an ε-ball in Y_{k_0} is just $\varepsilon^{k_0}\Omega_{k_0}$, we see that in order to cover the ellipsoid $\mathcal{E}$ by ε-balls we shall need at least $\prod_{k=1}^{k_0} \lambda_k/\varepsilon$ such balls. From this it follows that:

$$\prod_{k=1}^{k_0} \frac{\lambda_k}{\varepsilon} \leq N_\varepsilon(\mathcal{E}), \tag{2.25}$$

and therefore we have the following lower bound for the ε-entropy $H_\varepsilon(\mathcal{E})$:

$$\sum_{k=1}^{k_0} \log \frac{\lambda_k}{\varepsilon} \leq \log N_\varepsilon(\mathcal{E}) = H_\varepsilon(\mathcal{E}). \tag{2.26}$$

An upper bound for $H_{\varepsilon/2}(\mathcal{E})$ can be found in the following way [**8, 19**]: Let us construct in Y_{k_0} the cubical lattice with mesh width $\varepsilon_1 = \varepsilon/(2\sqrt{k_0})$, and with coordinate axes the axes of $\mathcal{E}_{k_0}$. In view of the choice of ε_1 any point of Y_{k_0}, and in particular of $\mathcal{E}_{k_0}$, lies within a distance not exceeding $(1/2)\varepsilon_1\sqrt{k_0} = (\varepsilon/4)$ from the nearest point of this lattice. In particular, it will lie at a distance not exceeding $(\varepsilon/4)$ from one of the lattice points which are contained in the parallelepiped P_{k_0} defined by:

$$-\frac{\varepsilon}{4} - \lambda_k \leq x_k \leq \frac{\varepsilon}{4} + \lambda_k, \quad 1 \leq k \leq k_0. \tag{2.27}$$

Now, if $k_0 = k_0(\varepsilon/4)$, that is, k_0 represents the number of terms in the sequence $\{\lambda_k\}$ which are greater than $(\varepsilon/4)$, then every point $x \in \mathcal{E}$ lies at a distance not exceeding $(\varepsilon/4)$ from a point of $\mathcal{E}_{k_0}$. In fact, let us write $x = \sum_k x_k \psi_k$, $\{\psi_k\}$ being the orthonormal basis for Y made of the eigenvectors of the operator A. Since x belongs to $\mathcal{E}$, then evidently $\sum_{k=1}^{\infty} |(x_k/\lambda_k)|^2 \leq 1$. Hence the square of the distance from x to $\mathcal{E}_{k_0}$ is

(2.28)

$$d^2(x, \mathcal{E}_{k_0}) = \sum_{k=k_0+1}^{\infty} |x_k|^2 = \sum_{k=k_0+1}^{\infty} \lambda_k^2 \left|\frac{x_k}{\lambda_k}\right|^2 \leq \lambda_{k_0+1}^2 \sum_{k=1}^{\infty} \left|\frac{x_k}{\lambda_k}\right|^2 \leq \left(\frac{\varepsilon}{4}\right)^2.$$

Now, the balls of radius $(\varepsilon/2)$ with centers at those lattice points within P_{k_0} cover the ellipsoid $\mathcal{E}$. In fact, from (2.28) each point of $\mathcal{E}$ is at a distance not exceeding $(\varepsilon/4)$ from $\mathcal{E}_{k_0}$, and each point of $\mathcal{E}_{k_0}$ is at a distance not exceeding $(\varepsilon/4)$ from some point of the lattice belonging to P_{k_0}; then each point of $\mathcal{E}$ lies at a distance not exceeding $(\varepsilon/2)$ from some point of the lattice belonging to P_{k_0}. Obviously the number of lattice points in P_{k_0} is not greater than

$$\prod_{k=1}^{k_0} 2\left(\frac{\lambda_k}{\varepsilon_1} + 1\right) = \prod_{k=1}^{k_0} \frac{2}{\varepsilon}\,(2\lambda_k\sqrt{k_0} + \varepsilon) \leq \left(\frac{6\sqrt{k_0}}{\varepsilon}\right)^{k_0}, \tag{2.29}$$

where we used the assumption $\varepsilon < \lambda_1 \leq 1 \leq k_0$. Then the number of elements in this ε-covering is no more than $[(6\sqrt{k_0(\varepsilon/4)})/\varepsilon]^{k_0(\varepsilon/4)}$ since $k_0 = k_0(\varepsilon/4)$. Taking the logarithm, we finally obtain

$$\begin{aligned} H_{\varepsilon/2}(\mathcal{E}) &\leq k_0\left(\frac{\varepsilon}{4}\right) \log \frac{6\sqrt{k_0(\varepsilon/4)}}{\varepsilon} \\ &= k_0\left(\frac{\varepsilon}{4}\right)\left[\log\left(\frac{1}{\varepsilon}\right) + \log 6 + \frac{1}{2}\log k_0\left(\frac{\varepsilon}{4}\right)\right]. \end{aligned} \tag{2.30}$$

For the next step we note that $H_\varepsilon(\mathcal{E})$ is a nondecreasing function as $\varepsilon \to 0$, then we can introduce the *order of growth* $\rho(\mathcal{E})$ of the entropy $H_\varepsilon(\mathcal{E})$ as follows:

$$\rho(\mathcal{E}) = \lim_{\varepsilon\to 0} \sup \frac{\log H_\varepsilon(\mathcal{E})}{\log(1/\varepsilon)}, \tag{2.31}$$

or, in the case $\rho(\mathcal{E}) = 0$, the *logarithmic order of growth* $\sigma(\mathcal{E})$ of $H_\varepsilon(\mathcal{E})$ which reads

$$\sigma(\mathcal{E}) = \limsup_{\varepsilon\to 0} \frac{\log H_\varepsilon(\mathcal{E})}{\log\log(1/\varepsilon)}. \tag{2.32}$$

Since we are interested in relating the asymptotic behavior of $H_\varepsilon(\mathcal{E})$ as $\varepsilon \to 0$ with the asymptotic behaviour of the *semi-axes* $\{\lambda_k\}$ of $\mathcal{E}$ as $k \to \infty$, we are led to introduce the *exponent of convergence* λ and the *logarithmic exponent of convergence* μ of the sequence $\{1/\lambda_k\}$, see [**15**]:

$$\lambda = \limsup_{\varepsilon\to 0} \frac{\log k_0(\varepsilon)}{\log(1/\varepsilon)}, \tag{2.33}$$

$$\mu = \limsup_{\varepsilon\to 0} \frac{\log k_0(\varepsilon)}{\log\log(1/\varepsilon)}, \tag{2.34}$$

where $k_0(\varepsilon)$ denotes the number of elements of the sequence λ_k which are greater than ε. The following relationship is proved in [**19**]: $\rho(\mathcal{E}) = \lambda$, and if $\rho(\mathcal{E}) = \lambda = 0$, then $\sigma(\mathcal{E}) = \mu + 1$. Finally, we can define the *degree of compactness* d_c associated with the range of the operator A as $d_c = (1/\rho)$ (if $\rho \neq 0$), and the *exponential degree of compactness* of Range$\,(A)$ as $d_c^e = 2^{1/\sigma}$ (if $\rho = 0$).

By using bounds (2.26) and (2.30), we can now evaluate the degree of compactness of Range$\,(A)$ in three specific examples: harmonic continuation, backward solution of the heat equation, and a convolution equation with continuous kernel; in all these examples the behavior with k of the eigenvalues is uniform, in the sense that the relative rate of decaying of the eigenvalues follows, for all k, a uniform law in k.

1. *Harmonic continuation.* Let us consider a family $\mathcal{F}$ of functions $u(r,\theta)$ which satisfy the Laplace equation at the interior of the unit disk. We want to determine $u(b,\theta)$, $b < 1$, assuming that $u(a,\theta)$, $a < b$, is known within a certain approximation. The solution to the problem is obtained by solving the following integral equation of Fredholm-type:

$$u(a,\theta) = \frac{1}{2\pi}\int_0^\pi P(\theta-\phi)\,u(b,\phi)\,d\phi, \quad -\pi < \theta \leq \pi, \tag{2.35}$$

where $P(\theta-\phi)$ is the Poisson kernel given by:

$$P(\theta-\phi) = \sum_{k=-\infty}^{+\infty} \left(\frac{a}{b}\right)^{|k|} e^{ik(\theta-\phi)}. \tag{2.36}$$

We can put Equation (2.35) into the form (1.1): $Af = g$, where $f(\phi) \equiv u(b,\phi)$, $g(\theta) = u(a,\theta)$, $(b > a)$; $u(b,\phi)$ is the restriction to the circle of radius b of a function harmonic in the unit disk, which belongs to $L^2[-\pi,\pi]$; then the following expansion converges in the sense of the L^2-norm:

$$u(1,\theta) = \sum_{k=-\infty}^{+\infty} u_k e^{ik\theta}, \quad \left(\sum_{k=-\infty}^{+\infty} |u_k|^2 < \infty\right). \tag{2.37}$$

Furthermore, we have:

$$u(b,\theta) = \sum_{k=-\infty}^{+\infty} b^{|k|} u_k e^{ik\theta}, \tag{2.38}$$

which is uniformly convergent. The eigenvalues of the operator A are $\lambda_k = (a/b)^{|k|}$, $b > a$, and the eigenfunctions are given by $\psi_k(\theta) = e^{-ik\theta}$; evidently, $\lim_{k\to\infty} \lambda_k = 0$. The Range (A) is not closed in $L^2[-\pi,\pi]$; in fact, only those functions u which satisfy the following bound:

$$\sum_{k=-\infty}^{+\infty} \left(u_k a^{|k|}\right)^2 < \infty, \tag{2.39}$$

belong to the Range (A). Now, if a noise n is added to the data function g, the function actually known is $\bar{g} = g + n$ which, in general, does not belong to Range (A); nevertheless hereafter we still assume that $\bar{g} \in$ Range (A). Next we restrict the solution space to those functions which satisfy the following bound:

$$\sum_{k=-\infty}^{+\infty} \left(u_k b^{|k|}\right)^2 \le 1. \tag{2.40}$$

It is now easy to evaluate the truncation number $k_0(\varepsilon)$, which is given by the largest integer such that $\lambda_k \ge \varepsilon$, i.e.,

$$k_0(\varepsilon) = \left[\frac{\log(1/\varepsilon)}{\log(b/a)}\right], \tag{2.41}$$

where $[\cdot]$ stands for the integral part. Now we split the sums (2.36)–(2.40) into two parts: the first is obtained by varying k from zero to $+\infty$; the second by varying k from -1 to $-\infty$. We denote the ε-entropy (ε-capacity) associated with the truncation of the first sum by $H_\varepsilon^{(+)}(\mathcal{E})$ ($C_\varepsilon^{(+)}(\mathcal{E})$); accordingly, the ε-entropy (ε-capacity) associated with the truncation of the second sum by $H_\varepsilon^{(-)}(\mathcal{E})$, $C_\varepsilon^{(-)}(\mathcal{E})$. Then using formula (2.41) and inequality (2.30) we obtain:

(2.42)

$$\begin{aligned}\sum_{k=1}^{k_0(\varepsilon)} \log\left(\frac{\lambda_k}{\varepsilon}\right) &\leq H_\varepsilon^{(+)}(\mathcal{E}) \leq C_\varepsilon^{(+)}(\mathcal{E}) \\ &\leq H_{\varepsilon/2}^{(+)}(\mathcal{E}) \leq k_0\left(\frac{\varepsilon}{4}\right)\left[\log\left(\frac{1}{\varepsilon}\right) + \log 6 + \frac{1}{2}\log k_0\left(\frac{\varepsilon}{4}\right)\right] \\ &\leq \frac{2+\log(1/\varepsilon)}{\log(b/a)}\left[\log\left(\frac{1}{\varepsilon}\right) + \log 6 + \frac{1}{2}\log k_0\left(\frac{\varepsilon}{4}\right)\right].\end{aligned}$$

The leading term on the righthand side of (2.42) as $\varepsilon \to 0$ is given by

$$\frac{\log(1/\varepsilon)}{\log(b/a)}\log\left(\frac{1}{\varepsilon}\right) \sim k_0(\varepsilon)\log\left(\frac{1}{\varepsilon}\right), \tag{2.43}$$

while the leading term on the lefthand side of (2.42) becomes

$$\frac{1}{2}\,k_0(\varepsilon)\log\left(\frac{1}{\varepsilon}\right). \tag{2.44}$$

We thus obtain, for ε sufficiently small, fairly sharp inequalities for the ε-capacity:

$$\frac{1}{2}\,k_0(\varepsilon)\log\left(\frac{1}{\varepsilon}\right) \lesssim C_\varepsilon^{(+)}(\mathcal{E}) \lesssim k_0(\varepsilon)\log\left(\frac{1}{\varepsilon}\right) \leq \frac{\log^2(1/\varepsilon)}{\log(b/a)}. \tag{2.45}$$

We thus have an upper bound for the maximal length, in binary units, of the messages conveyed back from $\bar{g}$ to reconstruct f, associated with the truncation of the positive sum; we obtain, with obvious notation:

$$L_{\max}^{(+)}(\varepsilon) \lesssim 2^{k_0(\varepsilon)\log(1/\varepsilon)} \sim 2^{(\log^2(1/\varepsilon)/\log(b/a))}. \tag{2.46}$$

Finally, for the total maximal length we obtain:

$$L_{\max}(\varepsilon) = L_{\max}^{(+)}(\varepsilon) + L_{\max}^{(-)}(\varepsilon) \lesssim 2^{k_0(\varepsilon)\log(1/\varepsilon)+1} \underset{\varepsilon\to 0}{\sim} 2^{k_0(\varepsilon)\log(1/\varepsilon)} \sim 2^{(\log^2(1/\varepsilon)/\log(b/a))}, \tag{2.47}$$

which can be taken as a quantitative estimate of the *metric information*.

Remark. Let us note that $\log(b/a) = \text{Cons.} \cdot L\{C\}$, where $L\{C\}$ is the extremal length of $\{C\}$, the latter expressing the set of curves in the ring domain $0 < a < r < b < \infty$, which join $r = a$ to $r = b$. $L\{C\}$ is a conformal invariant [**6**]. The righthand side of (2.45) may be regarded as a particular case of a more general result due to Erohin, see [**16**], which shows that for general sets of analytic functions:

$$H_{\varepsilon}^{(+)} \sim C_{\varepsilon}^{(+)} \sim \gamma \log^2\left(\frac{1}{\varepsilon}\right), \tag{2.48}$$

γ depending on some conformal invariant.

Concerning the *order of growth* $\rho(\mathcal{E})$ of the ε-entropy and the *exponent of convergence* λ: from (2.41) it follows that $\rho(\mathcal{E}) = \lambda = 0$. We then move on to the *logarithmic order of growth* $\sigma(\mathcal{E})$ and, correspondingly, to the *logarithmic exponent of convergence* μ; we have $\sigma(\mathcal{E}) = 2$ and, consequently, the *exponential degree of compactness* $d_c^e = 2^{1/\sigma} = 2^{1/2}$.

2. *Backward solution of the heat equation.* Let us consider a heat conducting ring of radius 1. One can pose two problems:

i) *Direct problem.* Determine the temperature distribution $h(t,\theta)$ at time t, when $h(0,\theta)$ is given. The solution is obtained by solving the Cauchy problem for the heat equation:

$$h_t = D\, h_{\theta\theta}, \quad D > 0, \tag{2.49}$$

$$h(0,\theta) = h_0(\theta), \quad 0 \le \theta < 2\pi. \tag{2.50}$$

ii) *Inverse problem.* Determine the temperature distribution $h(b,\theta) = f(\theta)$, at time $t = b$, when $h(a,\theta) \equiv g(\theta)$, $a > b$, is given.

The solution to the inverse problem is obtained by solving the Fredholm integral equation of the first kind:

$$h(a,\theta) \equiv g(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{K}(\theta-\phi) f(\phi)\, d\phi, \tag{2.51}$$

where the kernel $\mathcal{K}(\theta - \phi)$ is the elliptic Jacobi theta function:

$$\mathcal{K}(\theta - \phi) = \sum_{k=-\infty}^{+\infty} e^{-Dk^2(a-b)} e^{ik(\theta-\phi)}. \tag{2.52}$$

The eigenfunctions and the eigenvalues of the integral operator A are respectively $\psi_k(\theta) = e^{-ik\theta}$, $\lambda_k = \exp(-Dk^2(a-b))$; moreover, $\lim_{k\to\infty} \lambda_k = 0$. Once again we assume that the solution and the data space X and Y are both $L^2[-\pi, \pi]$. We may now consider the following expansion

$$h(t, \theta) = \sum_{k=-\infty}^{+\infty} h_k e^{-Dk^2 t} e^{ik\theta}, \tag{2.53}$$

which converges in the sense of the L^2-norm.

Again the Range (A) is not closed in $L^2[-\pi, \pi]$; in fact, only those functions h which satisfy the following bound:

$$\sum_{k=-\infty}^{+\infty} \left(h_k\, e^{-Dk^2 a}\right)^2 < \infty, \tag{2.54}$$

belong to Range (A). If a noise n is added to the data function g, only the function $\bar{g} = g + n$ is known and, in general, it does not belong to Range (A). Nevertheless we assume even in this case that $\bar{g} \in$ Range (A). Next we restrict the solution space to a subspace composed of those functions which satisfy the following a priori constraint:

$$\|h\|_{L^2}^2 = \sum_{k=-\infty}^{+\infty} \left(h_k\, e^{-Dk^2 b}\right)^2 \leq 1. \tag{2.55}$$

The truncation number $k_0(\varepsilon)$, which is given by the largest integer such that $\lambda_k \geq \varepsilon$ can be easily evaluated, i.e.,

$$k_0(\varepsilon) = \left[\left(\frac{\log(1/\varepsilon)}{D(a-b)} \right)^{1/2} \right]. \tag{2.56}$$

Based on considerations analogous to those developed in the case of harmonic continuation, and by splitting the sums (2.52)–(2.55) into two sums as done before, we obtain:

(2.57)

$$\sum_{k=1}^{k_0(\varepsilon)} \log\left(\frac{\lambda_k}{\varepsilon}\right) \leq C_\varepsilon^{(+)}(\mathcal{E}) \leq k_0\left(\frac{\varepsilon}{4}\right)\left[\log\left(\frac{1}{\varepsilon}\right) + \log 6 + \frac{1}{2}\log k_0\left(\frac{\varepsilon}{4}\right)\right]$$
$$\leq \left(\frac{2+\log(1/\varepsilon)}{D(b-a)}\right)^{1/2}\left[\log\left(\frac{1}{\varepsilon}\right) + \log 6 + \frac{1}{2}\log k_0\left(\frac{\varepsilon}{4}\right)\right].$$

The leading term on the righthand side of (2.57), as $\varepsilon \to 0$, is given by

$$\left(\frac{\log(1/\varepsilon)}{D(a-b)}\right)^{1/2} \log\left(\frac{1}{\varepsilon}\right) \sim k_0(\varepsilon)\log(1/\varepsilon), \tag{2.58}$$

while the leading term on the lefthand side of (2.57), as $\varepsilon \to 0$, is

$$\left(1 - \frac{1}{3}\log e\right) k_0(\varepsilon)\log(1/\varepsilon). \tag{2.59}$$

We therefore have quite sharp bounds on the ε-capacity, i.e.,

$$\left(1 - \frac{1}{3}\log e\right) k_0(\varepsilon)\log(1/\varepsilon) \lesssim C_\varepsilon^{(+)}(\mathcal{E}) \lesssim k_0(\varepsilon)\log(1/\varepsilon). \tag{2.60}$$

Then, we have an upper bound for the maximal length, in binary units, of the messages conveyed back from the data for reconstructing the solution, i.e.,

$$L_{max}^{(+)}(\varepsilon) \lesssim 2^{k_0(\varepsilon)\log(1/\varepsilon)} \leq 2^{\frac{\text{Cons.}}{(a-b)^{1/2}}[\log(1/\varepsilon)]^{3/2}}. \tag{2.61}$$

Then the final result referring to the total maximal length of the messages is:

$$\begin{aligned} L_{\max}(\varepsilon) &= L_{\max}^{(+)}(\varepsilon) + L_{\max}^{(-)}(\varepsilon) \\ &\lesssim 2^{k_0(\varepsilon)\log(1/\varepsilon)+1} \\ &\underset{\varepsilon\to 0}{\sim} 2^{k_0(\varepsilon)\log(1/\varepsilon)} \\ &\sim 2^{\frac{\text{Cons.}}{(a-b)^{1/2}}[\log(1/\varepsilon)]^{3/2}}. \end{aligned} \tag{2.62}$$

Accordingly, the *exponential degree of compactness* is given by $d_c^e = 2^{2/3}$.

3. *First kind Fredholm integral equation with continuous kernels.* Let us consider the following Fredholm integral equation of the first kind:

$$Af \equiv \int_0^1 \mathcal{K}(x,y)\, f(y)\, dy = g(x), \tag{2.63}$$

where the kernel $\mathcal{K}(x,y)$ is the continuous function

$$\mathcal{K}(x,y) = (1-x)y, \quad 0 \le y \le x \le 1, \tag{2.64}$$

$$\mathcal{K}(x,y) = x(1-y), \quad 0 \le x \le y \le 1. \tag{2.65}$$

Eigenfunctions and eigenvalues of operator A in Equation (2.63) can be easily evaluated: the eigenvalues are: $\lambda_k = 1/(k^2\pi^2)$. Once again, following considerations analogous to those developed in the previous examples, we obtain $k_0(\varepsilon) = [1/(\pi\sqrt{\varepsilon})]$ and, for ε sufficiently small, $(2\log e)\, k_0(\varepsilon) \lesssim C_\varepsilon \lesssim \frac{5}{2} k_0(\varepsilon) \log(1/\varepsilon)$. Consequently, we have $\rho = (1/2)$, $d_c = 2$ and

$$L_{max}(\varepsilon) \lesssim 2^{k_0(\varepsilon)\log(1/\varepsilon)} \le 2^{1/(\pi\sqrt{\varepsilon})\log(1/\varepsilon)}. \tag{2.66}$$

Remark. With reference to this last example, the reader interested in sharp bounds on the ε-capacity in the general setting of Sobolev spaces is referred to [**3**], see also Section 6 of [**12**].

Summarizing, we have the following table:

Behavior of λ_k	$\log L_{\max}(\varepsilon)$	d_c	d_c^e
$e^{-c_1 k}$	$c_1' \,[\log(1/\varepsilon)]^2$	—	$2^{1/2}$
$e^{-c_2 k^2}$	$c_2' \,[\log(1/\varepsilon)]^{3/2}$	—	$2^{2/3}$
c_3/k^2	$c_3' \,\varepsilon^{-1/2}\log(1/\varepsilon)$	2	—

3. Probabilistic information. Here we want to reconsider Equation (1.1) from a probabilistic point of view, adding explicitly the term representing the noise. With this in mind we pass from Equation (1.1) to Equation (1.3), and then to the probabilistic form of the latter, i.e., Equation (1.7), where ξ, ζ and η are Gaussian weak random variables (w.r.v.) in the Hilbert space $L^2[a,b]$ [**1**]. A Gaussian w.r.v. is uniquely defined by its mean element and its covariance operator; in the present case we denote by $R_{\xi\xi}$, $R_{\zeta\zeta}$ and $R_{\eta\eta}$ the covariance operators of ξ, ζ and η respectively. Next, we make the following assumptions:

i) ξ and ζ have zero mean, i.e., $m_\xi = m_\zeta = 0$;

ii) ξ and ζ are uncorrelated, i.e., $R_{\xi\zeta} = 0$;

iii) $R_{\zeta\zeta}^{-1}$ exists.

Regarding assumption (i), if it is known that $m_\xi \neq 0$ and $m_\zeta \neq 0$, then the problem can be easily reformulated in terms of the variables $(\xi - m_\xi)$ and $(\zeta - m_\zeta)$. The second hypothesis simply states that the signal-process ξ and the noise-process ζ are independent. Finally, the third assumption is the mathematical formulation of the fact that all the components of the data function are affected by noise or, in other words, that no components of the noise are equal to zero with probability one. As shown by Franklin, see formula (3.11) of [**5**], if assumptions (i) and (ii) are satisfied, then

$$R_{\eta\eta} = A R_{\xi\xi} A^\star + R_{\zeta\zeta}, \tag{3.1}$$

and the cross-covariance operator is given by:

$$R_{\xi\eta} = R_{\xi\xi} A^\star. \tag{3.2}$$

We also assume that $R_{\zeta\zeta}$ depends on a parameter ε that tends to zero when the noise vanishes, i.e.,

$$R_{\zeta\zeta} = \varepsilon^2 N, \tag{3.3}$$

where N is a given operator, e.g., $N = I$ for the white noise.

Now we are faced with the following problem:

Problem. *Given a value $\bar{g}$ of the* w.r.v. *η, find an estimate of the* w.r.v. *ξ.*

In order to give an answer to this problem, we turn Equation (1.7) into an infinite sequence of one-dimensional equations by means of the orthogonal projections, obtaining Equations (1.8), where $\xi_k = (\xi, \psi_k)$, $\zeta_k = (\zeta, \psi_k)$, $\eta_k = (\eta, \psi_k)$ are Gaussian random variables. Accordingly we introduce the variances $\rho_k^2 = (R_{\xi\xi}\psi_k, \psi_k)$, $\varepsilon^2\nu_k^2 = (R_{\zeta\zeta}\psi_k, \psi_k)$, $\lambda_k^2\rho_k^2 + \varepsilon^2\nu_k^2 = (R_{\eta\eta}\psi_k, \psi_k)$. Next we evaluate the amount of information on the variable ξ_k which is contained in the variable η_k; we have [**7**]:

$$J(\xi_k, \eta_k) = -\frac{1}{2}\ln(1 - r_k^2), \tag{3.4}$$

where

$$r_k^2 = \frac{|E\{\xi_k\eta_k\}|^2}{E\{|\xi_k|^2\}\,E\{|\eta_k|^2\}} = \frac{(\lambda_k\rho_k)^2}{(\lambda_k\rho_k)^2 + (\varepsilon\nu_k)^2}. \tag{3.5}$$

Thus

$$J(\xi_k, \eta_k) = \frac{1}{2}\ln\left(1 + \frac{\lambda_k^2\rho_k^2}{\varepsilon^2\nu_k^2}\right). \tag{3.6}$$

From equality (3.6) it follows that $J(\xi_k, \eta_k) < (1/2)\ln 2$ if $\lambda_k\rho_k < \varepsilon\nu_k$, that is, if the signal-to-noise ratio of the kth component is small. Thus, we are naturally led to introduce the following two sets: one, denoted by $\mathcal{I}$, which accounts for the components in which the signal dominates the noise; the other one, denoted by $\mathcal{N}$, which is instead related to the components in which the noise prevails; precisely, we define:

$$\mathcal{I} = \{k : \lambda_k\rho_k \geq \varepsilon\nu_k\}, \tag{3.7}$$

$$\mathcal{N} = \{k : \lambda_k\rho_k < \varepsilon\nu_k\}. \tag{3.8}$$

Remark. Let us note that the sets $\mathcal{I}$ and $\mathcal{N}$ are not equipped, in general, with any order relation. However, we can rearrange and renumber the terms $\lambda_k\rho_k$ and $\varepsilon\nu_k$ in such a way as to introduce an order relationship. Furthermore, for the sake of simplicity and without loss of generality, we hereafter assume that there do not exist two identical terms $\lambda_k\rho_k/\nu_k$ corresponding to different values of k. In this situation there exists a unique value of k, denoted by k_I, which separates set $\mathcal{I}$ from set $\mathcal{N}$.

Since ξ_k and ζ_k are supposed to be Gaussian random variables, we can assume the following probability densities:

(3.9)
$$p_{\xi_k}(x) = \frac{1}{\sqrt{2\pi}\,\rho_k}\ \exp\left\{-\left(\frac{x^2}{2\rho_k^2}\right)\right\}, \quad k = 1, 2, \dots ,$$

(3.10)
$$p_{\zeta_k}(x) = \frac{1}{\sqrt{2\pi}\,\varepsilon\nu_k}\ \exp\left\{-\left(\frac{x^2}{2\varepsilon^2\nu_k^2}\right)\right\}, \quad k = 1, 2, \dots .$$

By equations (1.8) we can also introduce the conditional probability density $p_{\eta_k}(y|x)$ of the random variable η_k for fixed $\xi_k = x$, which reads:

(3.11)
$$\begin{aligned} p_{\eta_k}(y|x) &= \frac{1}{\sqrt{2\pi}\,\varepsilon\nu_k}\ \exp\left\{-\ \frac{(y-\lambda_k x)^2}{2\varepsilon^2\nu_k^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}\,\varepsilon\nu_k}\ \exp\left\{-\ \frac{\lambda_k^2}{2\varepsilon^2\nu_k^2}\left(x - \frac{y}{\lambda_k}\right)^2\right\}. \end{aligned}$$

Let us now apply the Bayes formula, which provides the conditional probability density of ξ_k given η_k through the following expression:

$$p_{\xi_k}(x|y) = \frac{p_{\xi_k}(x)p_{\eta_k}(y|x)}{p_{\eta_k}(y)}. \tag{3.12}$$

Thus, if a realization of the random variable η_k is given by $\bar{g}_k$, formula (3.12) becomes

(3.13)
$$p_{\xi_k}(x|\bar{g}_k) = A_k \exp\left\{-\frac{x^2}{2\rho_k^2}\right\} \exp\left\{-\frac{\lambda_k^2}{2\varepsilon^2\nu_k^2}\left(x - \frac{\bar{g}_k}{\lambda_k}\right)^2\right\}, \quad A_k = \text{Cons.}$$

The conditional probability density (3.13) can be regarded as the product of two Gaussian probability densities: $p_1(x) = A_k^{(1)} \exp\left\{-x^2/2\rho_k^2\right\}$ and $p_2(x) = A_k^{(2)} \exp\left\{-(\lambda_k^2/2\varepsilon^2\nu_k^2)\,(x - (\bar{g}_k/\lambda_k))^2\right\}$, $A_k = A_k^{(1)} \cdot A_k^{(2)}$, whose variances are respectively given by ρ_k^2 and $(\varepsilon\nu_k/\lambda_k)^2$. Let us note that if $k \in \mathcal{I}$, the variance associated with the density $p_2(x)$ is smaller than the corresponding variance of $p_1(x)$, and vice versa if $k \in \mathcal{N}$. Therefore, it is reasonable to consider as an acceptable approximation of $\langle \xi_k \rangle$ the mean value given by the density $p_2(x)$ if $k \le k_I$, (i.e., if $k \in \mathcal{I}$), whereas the mean value is given by the density $p_1(x)$ if $k > k_I$, (i.e., if $k \in \mathcal{N}$). We can write the following approximation

$$\langle \xi_k \rangle = \begin{cases} \dfrac{\bar{g}_k}{\lambda_k} & \text{if } k \le k_I, \\ 0 & \text{if } k > k_I. \end{cases} \tag{3.14}$$

Consequently, given the value $\bar{g}$ of the w.r.v. η, we are led to consider the following estimate of ξ: $\sum_{k \in \mathcal{I}} (\bar{g}_k/\lambda_k)\psi_k \equiv \xi_I$. Next, we introduce the operator $B_{\mathcal{I}} : L^2[a,b] \to L^2[a,b]$, defined as follows

$$B_{\mathcal{I}} \psi_k = \begin{cases} \dfrac{1}{\lambda_k}\, \psi_k & \text{if } k \le k_I, \\ 0 & \text{if } k > k_I, \end{cases} \tag{3.15}$$

then $\xi_I = B_{\mathcal{I}} \bar{g} = \sum_{k=1}^{k_I} (\bar{g}_k/\lambda_k)\psi_k$. We can now evaluate the global mean square error; taking into account formulae (3.1) and (3.2), we can formally write:

$$\begin{aligned} E\left\{\|\xi - B_{\mathcal{I}}\eta\|^2\right\} &= \mathrm{Tr}\ (R_{\xi\xi} - R_{\xi\xi}AB^\star - BAR_{\xi\xi} + BR_{\eta\eta}B^\star) \\ &= \sum_{k=k_I+1}^{\infty} \rho_k^2 + \sum_{k=1}^{k_I} \left(\frac{\varepsilon\nu_k}{\lambda_k}\right)^2. \end{aligned} \tag{3.16}$$

The sum (3.16) is finite if and only if $\mathrm{Tr}\, R_{\xi\xi} = \sum_{k=1}^{\infty} \rho_k^2 < \infty$, i.e., if the covariance operator $R_{\xi\xi}$ is of trace class. In the following we assume that this condition is satisfied. Hereafter we also suppose that $\lim_{k\to\infty}(\lambda_k\rho_k/\nu_k) = 0$, and therefore the set $\mathcal{I}$ exists and its cardinality is finite for any given $\varepsilon > 0$. Next, we prove the following lemma.

Lemma 2. *If* $\mathrm{Tr}\, R_{\xi\xi} = \Gamma < \infty$ *and moreover* $\lim_{k\to\infty}(\lambda_k\rho_k/\nu_k) = 0$, *then we can introduce a number* $k_\alpha(\varepsilon)$ *defined as follows*

$$k_\alpha(\varepsilon) = \max\left\{ m \in \mathbf{N} : \sum_{k=1}^{m} \left(\rho_k^2 + \frac{\varepsilon^2\nu_k^2}{\lambda_k^2}\right) \le \Gamma \right\}. \tag{3.17}$$

We can then prove

$$\text{i)} \quad \lim_{\varepsilon\to 0} k_\alpha(\varepsilon) = +\infty, \tag{3.18}$$

$$\text{ii)} \quad \lim_{\varepsilon\to 0} \left\{ \sum_{k=k_\alpha+1}^{\infty} \rho_k^2 + \sum_{k=1}^{k_\alpha(\varepsilon)} \left(\frac{\varepsilon\nu_k}{\lambda_k}\right)^2 \right\} = 0. \tag{3.19}$$

Proof. (i) Let us denote by k_{α_1} the sum $k_\alpha + 1$. If equality (3.18) is not true, then there should exist a finite number M, which does not depend on ε and such that, for any sequence $\{\varepsilon_i\}$ converging to zero, $k_{\alpha_1} < M$. From formula (3.17) it then follows:

$$\Gamma < \sum_{k=1}^{k_{\alpha_1}(\varepsilon_i)} \left(\rho_k^2 + \frac{\varepsilon^2\nu_k^2}{\lambda_k^2}\right) \le \sum_{k=1}^{M} \left(\rho_k^2 + \frac{\varepsilon^2\nu_k^2}{\lambda_k^2}\right). \tag{3.20}$$

For any sequence $\{\varepsilon_i\}$ tending to zero, we have

$$\Gamma < \sum_{k=1}^{M} \rho_k^2 \le \sum_{k=1}^{\infty} \rho_k^2 = \Gamma, \tag{3.21}$$

and the contradiction is explicit.

(ii) Since $\lim_{\varepsilon\to 0} k_\alpha(\varepsilon) = +\infty$, and $\sum_{k=1}^{\infty} \rho_k^2 < \infty$, then

$$\lim_{\varepsilon\to 0} \sum_{k=k_\alpha(\varepsilon)+1}^{\infty} \rho_k^2 = 0. \tag{3.22}$$

Regarding the term $\sum_{k=1}^{k_\alpha(\varepsilon)} (\varepsilon\nu_k/\lambda_k)^2$, we can proceed as follows. From formula (3.17) we have

$$\sum_{k=1}^{k_\alpha(\varepsilon)} \left(\frac{\varepsilon\nu_k}{\lambda_k}\right)^2 + \sum_{k=1}^{k_\alpha(\varepsilon)} \rho_k^2 \le \Gamma = \sum_{k=1}^{\infty} \rho_k^2, \tag{3.23}$$

and therefore

$$\sum_{k=1}^{k_\alpha(\varepsilon)} \left(\frac{\varepsilon\nu_k}{\lambda_k}\right)^2 \le \sum_{k=k_\alpha(\varepsilon)+1}^{\infty} \rho_k^2. \tag{3.24}$$

Since $\lim_{\varepsilon\to 0}\sum_{k=k_\alpha+1}^{\infty}\rho_k^2 = 0$ (see formula (3.22)), we also have $\lim_{\varepsilon\to 0}\sum_{k=1}^{k_\alpha(\varepsilon)}(\varepsilon\nu_k/\lambda_k)^2 = 0$. □

Finally, we can prove the following theorem.

Theorem 2. *If the covariance operator* $R_{\xi\xi}$ *is of trace class, and moreover* $\lim_{\varepsilon\to 0}\lambda_k\rho_k/\nu_k = 0$*, then the following limit holds true:*

$$\lim_{\varepsilon\to 0} E\left\{\|\xi - B_{\mathcal{I}}\eta\|^2\right\} = \lim_{\varepsilon\to 0}\left\{\sum_{k=k_I+1}^{\infty}\rho_k^2 + \sum_{k=1}^{k_I}\left(\frac{\varepsilon\nu_k}{\lambda_k}\right)^2\right\} = 0. \tag{3.25}$$

Proof. The proof proceeds in two steps.

a) We want to prove that $\lim_{\varepsilon\to 0}\sum_{k=k_I+1}^{\infty}\rho_k^2 = 0$. We have two possibilities: either $k_I \geq k_\alpha$, or $k_I < k_\alpha$. In the former case the statement follows from the fact that $\lim_{\varepsilon\to 0}\sum_{k=k_\alpha+1}^{\infty}\rho_k^2 = 0$. In the latter case, if $k_I < k_\alpha$, then we have

(3.26)

$$\sum_{k=k_I(\varepsilon)+1}^{\infty}\rho_k^2 \leq \sum_{k=k_\alpha+1}^{\infty}\rho_k^2 + \sum_{k=k_I(\varepsilon)+1}^{k_\alpha(\varepsilon)}\left(\frac{\varepsilon\nu_k}{\lambda_k}\right)^2 \leq \sum_{k=k_\alpha+1}^{\infty}\rho_k^2 + \sum_{k=1}^{k_\alpha(\varepsilon)}\left(\frac{\varepsilon\nu_k}{\lambda_k}\right)^2.$$

But in Lemma 2 we have proved that the righthand side of formula (3.26) tends to zero as $\varepsilon\to 0$, and the statement follows.

b) We want to prove that $\lim_{\varepsilon\to 0}\sum_{k=1}^{k_I(\varepsilon)}(\varepsilon\nu_k/\lambda_k)^2 = 0$. Now again either $k_I \leq k_\alpha$ or $k_I > k_\alpha$. In the first case the statement follows from the fact that $\lim_{\varepsilon\to 0}\sum_{k=1}^{k_\alpha(\varepsilon)}(\varepsilon\nu_k/\lambda_k)^2 = 0$, as proved in Lemma 2. If, on the contrary, $k_I > k_\alpha$, then we have, for $k \leq k_I$, $\rho_k \geq \varepsilon\nu_k/\lambda_k$, and therefore

$$\sum_{k=k_\alpha+1}^{k_I}\left(\frac{\varepsilon\nu_k}{\lambda_k}\right)^2 \leq \sum_{k=k_\alpha+1}^{k_I}\rho_k^2 \leq \sum_{k=k_\alpha+1}^{\infty}\rho_k^2. \tag{3.27}$$

Since $\lim_{\varepsilon\to 0}\sum_{k=k_\alpha+1}^{\infty}\rho_k^2 = 0$, it follows that

$$\lim_{\varepsilon\to 0}\sum_{k=k_\alpha+1}^{k_I}\left(\frac{\varepsilon\nu_k}{\lambda_k}\right)^2 = 0. \tag{3.28}$$

Now the statement follows recalling that $\lim_{\varepsilon\to 0}\sum_{k=1}^{k_\alpha}(\varepsilon\nu_k/\lambda_k)^2 = 0$, as proved in Lemma 2. □

If we now sum up the information carried by the set $\{\eta_k\}_{k\in\mathcal{I}}$ on the corresponding set $\{\xi_k\}_{k\in\mathcal{I}}$ we obtain the quantity:

$$\sum_{k=1}^{k_I} J(\xi_k,\eta_k) = \sum_{k=1}^{k_I}\ln\left(1+\frac{\lambda_k^2\rho_k^2}{\varepsilon^2\nu_k^2}\right)^{1/2} \simeq \sum_{k=1}^{k_I}\ln\left|\frac{\lambda_k\rho_k}{\varepsilon\nu_k}\right|, \tag{3.29}$$

which could be called the *probabilistic information* associated with equation (1.7). For the approximation on the righthand side of (3.29) we used $\lambda_k\rho_k \geq \varepsilon\nu_k$ for $k\in\mathcal{I}$. Now, in order to compare the probabilistic information with the *metric information*, we may consider two somehow extremal approximations:

α) If $\rho_k\sim\nu_k$, $k\in\mathcal{I}$, we have

$$\sum_{k=1}^{k_I}\ln\left|\frac{\lambda_k\rho_k}{\varepsilon\nu_k}\right| \sim \sum_{k=1}^{k_I}\ln\left(\frac{\lambda_k}{\varepsilon}\right) = \sum_{k=1}^{k_0}\ln\left(\frac{\lambda_k}{\varepsilon}\right), \tag{3.30}$$

since $k_I = k_0$. Let us note that the righthand side of formula (3.30) coincides (up to an immaterial conversion factor between logarithm types) with the lower bound for $H_\varepsilon(\mathcal{E})$.

β) If $\lambda_k\rho_k\sim\nu_k$, $k\in\mathcal{I}$, we have

$$\sum_{k=1}^{k_I}\ln\left|\frac{\lambda_k\rho_k}{\varepsilon\nu_k}\right| \sim k_I(\varepsilon)\ln\left(\frac{1}{\varepsilon}\right) \sim k_0(\varepsilon)\ln\left(\frac{1}{\varepsilon}\right), \tag{3.31}$$

which coincides with the upper bound for $H_{\varepsilon/2}(\mathcal{E})$, which we have computed in the various examples of the previous section.

It is interesting to note that the metric information provides the limits of the range over which the probabilistic information varies when the signal-to-noise ratio ranges between the extrema given by the two previous approximations. The results given by approximations (α) and (β) allow us to look at the analogy and parallelism between metric and probabilistic information on a more precise and quantitative ground.

REFERENCES

1. A.V. Balakrishnan, *Applied functional analysis*, Springer-Verlag, New York, 1976.

2. M. Bertero, C. De Mol and G.A. Viano, *The stability of inverse problems*, in *Inverse scattering problems in optics*, Springer-Verlag, Berlin, 1980.

3. M.S. Birman and M.Z. Solomjak, *Piecewise-polynomial approximations of functions of the classes* W_p^α, Math. USSR, Sbornik **2** (1967), 295–317.

4. H.W. Engl, *Regularization methods for the stable solution of inverse problems*, Surveys Math. Indust. **3** (1993), 71–143.

5. J.N. Franklin, *Well-posed stochastic extensions of ill-posed linear problems*, J. Math. Anal. Appl. **31** (1970), 682–716.

6. W.H.J. Fuchs, *Topics in the theory of functions of one complex variable*, Van Nostrand, Princeton, 1967.

7. I.M. Gel'fand and N.Ya. Vilenkin, *Generalized functions*, Vol. 4, Academic Press, New York, 1964.

8. I.M. Gel'fand and A.M. Yaglom, *Calculation of the amount of information about a random function contained in another such function*, Amer. Math. Soc. Transl. **12** (1959), 199–246.

9. C.W. Groetsch, *The theory of Tikhonov regularization for Fredholm integral equations of the first kind*, Pitman, Boston, 1984.

10. E. Hille and J.D. Tamarkin, *On the characteristic values of linear integral equations*, Acta Math. **57** (1931), 1–76.

11. F. John, *Continuous dependence on data for solutions of partial differential equations with a prescribed bound*, Comm. Pure Appl. Math. **13** (1960), 551–585.

12. J.L. Kelley, *General topology*, Van Nostrand, Princeton, 1955.

13. A.N. Kolmogorov, *Three approaches to the quantitative definition of information*, Problemy Peredachi Informatsii **1** (1965), 3–11.

14. A.N. Kolmogorov and V.M. Tihomirov, *ϵ-entropy and ϵ-capacity of sets in functional spaces*, Amer. Math. Soc. Transl. **17** (1961), 277–364.

15. B.Ya. Levin, *Distribution of zeros of entire functions*, Amer. Math. Soc., Providence, 1964.

16. G.G. Lorentz, *Approximation of functions*, Holt, Rinehart and Winston, New York, 1966.

17. D. Middleton, *An introduction to statistical communication theory*, McGraw-Hill, New York, 1960.

18. M.Z. Nashed, *Generalized inverses and applications*, Academic Press, New York, 1976.

19. R.T. Prosser, *The ε-entropy and ε-capacity of certain time-varying channels*, J. Math. Anal. Appl. **16** (1966), 553–573.

20. C.A. Rogers, *Packing and covering*, Cambridge Univ. Press, Cambridge, 1964.

21. C.E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423.

22. A. Tikhonov and V. Arsenine, *Méthodes de résolution des problèmes mal posés*, Mir, Moscow, 1976.

23. S. van de Geer, *Applications of empirical process theory*, Cambridge Univ. Press, Cambridge, 2000.

Istituto di Cibernetica e Biofisica, Consiglio Nazionale delle Ricerche, Via De Marini, 6–16149 Genova, Italy
E-mail address: `demicheli@icb.ge.cnr.it`

Dipartimento di Fisica - Università di Genova, Istituto Nazionale Di Fisica Nucleare - sez. di Genova, Via Dodecaneso, 33–16146 Genova, Italy
E-mail address: `viano@ge.infn.it`