JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **197,** 188–206 (1996)
ARTICLE NO. 0015

On the Eigenfunction Expansions Associated with Fredholm Integral Equations of First Kind in the Presence of Noise

N. Magnoli* and G. A. Viano*

Dipartimento di Fisica, Università di Genova, Istituto Nazionale di Fisica Nucleare, sez. di Genova Via Dodecaneso, 33-16146 Genoa, Italy

Submitted by Joyce R. McLaughlin

Received June 7, 1994

In this paper we consider the eigenfunction expansions associated with Fredholm integral equations of first kind when the data are perturbed by noise. We prove that these expansions are asymptotically convergent, in the sense of L^2-norm, when the bound of the noise tends to zero. This result allows us to construct a continuous mapping from the data space to the solution space, without using any constraint or *a priori* bound. We can also show a probabilistic version of this result, which is based on the order–disorder transition in the Fourier coefficients of the noisy data. From these results one can derive algorithms and in particular statistical methods able to furnish approximations of the solution without any use of prior knowledge.

1. INTRODUCTION

The signal recovery problem can be formulated in the language of communication channel theory: to determine the identity of an input signal from the knowledge (within a certain noise) of the output signal and the identity of the channel. In several problems of interest a linear channel may be described as an operator of the Hilbert–Schmidt class, i.e., an operator of the form

$$(Af)(x) = \int_a^b K(x, y) f(y)\, dy = g(x) (a \le x \le b), \tag{1}$$

where $K(x, y)$ is supposed to be a Hermitian and square integrable kernel; i.e.,

$$K(x, y) = \overline{K(y, x)} \tag{2}$$

*E-mail: Decnet 39166; Bitnet @GENOVA.INFN.IT.

0022-247X/96 $12.00

and

$$\int_a^b \left\{ \int_a^b |K(x,y)|^2 \, dx \right\} dy < \infty. \tag{3}$$

Then $A: L^2(a,b) \to L^2(a,b)$ is a self-adjoint compact operator.

For the sake of simplicity we shall suppose hereafter that the kernel K, the function g, and the unknown function f are real-valued functions; moreover we assume that the interval $[a,b]$ is a bounded and closed subset of the real line. These assumptions do not limit the generality of our reasoning.

The Hilbert–Schmidt theorem guarantees that the integral operator A 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 also that $\lim_{k\to\infty} \lambda_k = 0$. Furthermore we shall suppose hereafter that the eigenvalues are ordered as follows: $\lambda_1 > \lambda_2 > \lambda_3 > \cdots$. The reason for all these assumptions, which can be simply removed, is to make more transparent all the arguments below. We could suppose, for instance, that the supports of the data and those of the solutions belong to intervals which do not coincide; i.e., we could assume that the operator A transforms functions which belong to $L^2(a,b)$ into functions which belong to $L^2(c,d)$, where the interval $[c,d]$ does not coincide with the interval $[a,b]$. In this case we can work out the problem in terms of singular values and singular functions of the operator A (see [BDV]) and all the results obtained below can be easily reformulated.

We can associate with integral equation (1) the eigenfunction expansion

$$f(x) = \sum_{k=1}^{\infty} \left(\frac{g_k}{\lambda_k} \right) \psi_k, \tag{4}$$

where $g_k = (g, \psi_k)$ ($(\cdot,\cdot)$ denotes the scalar product in $L^2(a,b)$). The series (4) converges in the sense of L^2.

As we said above, in the signal recovery problem we must take into account the "disturbance" which can be represented by an additive noise $n(x)$. Then, instead of Eq. (1), we have

$$Af + n = \bar{g}, \tag{5}$$

where $\bar{g}(x) = g(x) + n(x)$. Therefore instead of expansion (4) we have to deal with an expansion of the type

$$\sum_{k=1}^{\infty} \left(\frac{\bar{g}_k}{\lambda_k} \right) \psi_k(x), \tag{6}$$

where $\bar{g}_k = (\bar{g}, \psi_k)$. The expansion (6) is generally diverging because $\bar{g}$ does not belong, in general, to the range of the operator A. This is a manifestation of the ill-posedness of the Fredholm integral equation of the first kind.

Several methods of regularization have been proposed; all of them modify one of the elements of the triplet $\{A, X, Y\}$, where A is the integral operator defined by Eq. (1), and X and Y are, respectively, the solution and the data space (in our case $X \equiv Y \equiv L^2(a, b)$). 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 on a compact topological space into a Hausdorff topological space; if σ is one-to-one, then its inverse map σ^{-1} is continuous [Ke]. The condition of compactness can be realized by the use of *a priori* bounds [Jo], which require some prior knowledge or some constraints on the solution. Then the procedure works taking into account two bounds, one on the solutions, the other on the noise $n(x)$ (see [TA]),

$$\|Bf\|_{L^2(a,b)} \leq 1 \tag{7}$$

$$\|n\|_{L^2(a,b)} \leq \epsilon, \tag{8}$$

where B is a suitable constraint operator. Let us suppose that the eigenfunctions $\{\psi_k\}_1^\infty$ diagonalize the operator B^*B; 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 a compact inverse iff $\lim_{k \to \infty} \beta_k^2 = +\infty$; under such a condition, the solution obtained, by truncating expansion (6) at the largest integer k such that $\lambda_k > \epsilon \beta_k$, converges to the solution f, as $\epsilon \to 0$, in the sense of L^2-norm. In several cases a much milder constraint is used, i.e., $B = I(\beta_k = 1, \forall k)$. In this case the compactness condition is not satisfied and the approximation obtained, by truncating expansion (6) at the largest integer k such that $\lambda_k > \epsilon$, is only weakly convergent to the solution f as $\epsilon \to 0$ [BDV].

At this point we must stress the remarkable difference between synthesis problems and signal recovery: whereas in the first class of problems suitable constraints on the solutions enter in a natural way (consider, for instance, the constraint on the ohmic loss in the antenna synthesis), this is not always the case in signal recovery. Furthermore truncations of expansion (6), as those illustrated above, obtained comparing the eigenvalues λ_k with the bound ϵ on the noise, appear quite unnatural from the viewpoint of experimental or physical science. It would seem much more appropriate that a truncation be obtained by comparing the Fourier coefficient of the

data (i.e., g_k) with the noise. Therefore the following question arises quite naturally:

Problem. Is it possible to find a regularization procedure for Fredholm integral equation of first kind, with Hilbert–Schmidt kernel, that does not require any *a priori* information?

We can give a positive answer to this question if we are able to exhibit a continuous mapping of the noisy data into the solution space X without any requirement of compactness or any use of *a priori* bounds on the solution. We shall prove the existence of this continuous mapping in Section 2 by showing that even if expansion (6) is divergent when n does not belong to the range of the operator A, it is nevertheless asymptotically convergent, in the sense of L^2-norm, as the noise tends to zero. Unfortunately from this result a general algorithm able to furnish approximations of the unknown solution f does not necessarily follow. As we shall discuss in Section 2, these approximations can be obtained only in some specific and very peculiar situations. However, in Section 3 we shall show that the main result proved in Section 2 admits a probabilistic version. From this fact it follows that some probabilistic and statistical methods (like that of correlogram) can be used for an approximate determination of the solution f, without any specific and *a priori* information on f.

2. ASYMPTOTIC CONVERGENCE OF EIGENFUNCTION EXPANSIONS

In this section the only assumption we use is that the disturbance is represented by a bounded and integrable function $n(x)$. For the sake of simplicity we will assume the simplest and very realistic bound

$$\sup |n(x)| \leq \epsilon, \qquad x \in [a, b] \tag{9}$$

We do not require hereafter in this section an explicit knowledge of ϵ. Then we can prove the following lemma.

LEMMA 1. *The following statements hold true*:

(i) $$\sum_{k=1}^{\infty} \left(\frac{g_k}{\lambda_k} \right)^2 = \|f\|^2_{L^2(a,b)} = C \tag{10}$$

(ii) $$\sum_{k=1}^{\infty} \left(\frac{\bar{g}_k}{\lambda_k} \right)^2 = \infty, \qquad \text{if } \bar{g} \notin \text{Range}(A) \tag{11}$$

(iii) $$\lim_{\epsilon \to 0} \bar{g}_k = g_k, \qquad \forall k \tag{12}$$

(iv) If $k_0(\epsilon)$ is defined as

$$k_0(\epsilon) = \max\left\{ m \in \mathbf{N} : \sum_{k=1}^{m} \left(\frac{\bar{g}_k}{\lambda_k} \right)^2 \leq C \right\} \tag{13}$$

$$\text{then } \lim_{\epsilon \to 0} k_0(\epsilon) = +\infty. \tag{14}$$

Proof. Statements (i) and (ii) are obvious; in particular statement (i) is simply the Parseval equality associated with the eigenfunction expansion (4). For what concerns statement (iii) it is an immediate consequence of the bound (9); indeed we have

$$\bar{g}_k = \int_a^b [g(x) + n(x)] \psi_k \, dx. \tag{15}$$

Since

$$\left| \int_a^b n(x) \psi_k \, dx \right| \leq \epsilon \int_a^b |\psi_k| \, dx, \tag{16}$$

$\lim_{\epsilon \to 0} \bar{g}_k = g_k, \forall k$. Finally we can prove statement (iv) as follows. Let us introduce $k_1(\epsilon) = k_0(\epsilon) + 1$. Then from definition (13) it follows that

$$\sum_{k=1}^{k_1(\epsilon)} \left(\frac{\bar{g}_k}{\lambda_k} \right)^2 > C. \tag{17}$$

(Hereafter, for the sake of simplicity and without loss of generality, we shall suppose that the Fourier coefficients $\bar{g}_k$ $(k \geq k_0(\epsilon))$ are not zero. Indeed if $\bar{g}_{k_1(\epsilon)}$ is zero, then it is sufficient to replace the sum at the l.h.s. of (17) by a similar sum extended up to a Fourier coefficient $\bar{g}_k$ $(k > k_1(\epsilon))$ different from zero.) Now statement (iv) is proved if we are able to show that $\lim_{\epsilon \to 0} k_1(\epsilon) = \infty$. We claim that for any sequence $\{\epsilon_i\}$ tending to zero, we have that $\lim_{\epsilon_i \to 0} k_1(\epsilon_i) = \infty$. Suppose that this is not the case; then there should exist a finite number M which does not depend on ϵ and such that $k_1(\epsilon_i) \leq M$ even if the sequence $\{\epsilon_i\}$ tends to zero. Then from inequality (17) it follows that

$$C < \sum_{k=1}^{k_1(\epsilon_i)} \left(\frac{\bar{g}_k}{\lambda_k} \right)^2 \leq \sum_{k=1}^{M} \left(\frac{\bar{g}_k}{\lambda_k} \right)^2. \tag{18}$$

But as $\epsilon_i \to 0$, we have for statement (iii)

$$C < \sum_{k=1}^{M} \left(\frac{g_k}{\lambda_k} \right)^2 \leq \sum_{k=1}^{\infty} \left(\frac{g_k}{\lambda_k} \right)^2 = C, \tag{19}$$

which leads to a contradiction, and statement (iv) is proved. ∎

Now we introduce the following approximation:

$$f_0^{(\epsilon)}(x) = \sum_{k=1}^{k_0(\epsilon)} \left(\frac{\bar{g}_k}{\lambda_k} \right) \psi_k. \tag{20}$$

Then we can prove the following theorem.

THEOREM 1. *The following equality holds true*:

$$\lim_{\epsilon \to 0} \| f - f_0^{(\epsilon)} \|_{L^2(a,b)} = 0. \tag{21}$$

Proof. From the Parseval equality it follows that

$$\| f - f_0^{(\epsilon)} \|^2_{L^2(a,b)} = \sum_{k_0(\epsilon)+1}^{\infty} \left(\frac{g_k}{\lambda_k} \right)^2 + \sum_{k=1}^{k_0(\epsilon)} \left(\frac{\bar{g}_k - g_k}{\lambda_k} \right)^2. \tag{22}$$

Since $\lim_{\epsilon \to 0} k_0(\epsilon) = +\infty$ and $\sum_{k=1}^{\infty} (g_k/\lambda_k)^2$ converges,

$$\lim_{\epsilon \to 0} \sum_{k=k_0(\epsilon)+1}^{\infty} \left(\frac{g_k}{\lambda_k} \right)^2 = 0.$$

Concerning the second term at the r.h.s. of formula (22), let us introduce the functions

$$\mathscr{G}(u) = \sum_{k=1}^{\infty} \left(\frac{g_k}{\lambda_k} \right) I_{[k,k+1[}(u) \tag{23}$$

$$\overline{\mathscr{G}}(u) = \sum_{k=1}^{\infty} \left(\frac{\bar{g}_k}{\lambda_k} \right) I_{[k,k+1[}(u), \tag{24}$$

where I_A is the characteristic function of the set A. We have

$$\int_0^\infty [\mathscr{G}(u)]^2\, du = \sum_{k=1}^{\infty} \left(\frac{g_k}{\lambda_k}\right)^2 = C \tag{25}$$

$$\int_0^\infty \left[\overline{\mathscr{G}}(u)\right]^2 du = \sum_{k=1}^{\infty} \left(\frac{\bar{g}_k}{\lambda_k}\right)^2 = \infty \tag{26}$$

and at each fixed value of $u (u \geq 0)$

$$\overline{\mathscr{G}}(u) \to \mathscr{G}(u) \qquad \text{for } \epsilon \to 0. \tag{27}$$

Let us denote by $U(\epsilon)$ the unique root of the equation $\int_0^U [\overline{\mathscr{G}}(u)]^2\, du = C$ (indeed $0 \leq U \to \int_0^U [\overline{\mathscr{G}}(u)]^2\, du$ is a continuous function which is zero for $U = 0$, and $+\infty$ for $U \to +\infty$; furthermore let us recall that we have assumed that all the coefficients $\bar{g}_k$ are different from zero). From statement (iv) of Lemma 1, it follows that $\lim_{\epsilon \to 0} U(\epsilon) = +\infty$. Next we have

$$\int_0^{U(\epsilon)} \left[\overline{\mathscr{G}}(u) - \mathscr{G}(u)\right]^2 du$$

$$= \int_{U(\epsilon)}^\infty [\mathscr{G}(u)]^2\, du - 2\int_0^{U(\epsilon)} \mathscr{G}(u)\left[\overline{\mathscr{G}}(u) - \mathscr{G}(u)\right] du. \tag{28}$$

Concerning the first term at the r.h.s. of formula (28), we have

$$\lim_{\epsilon \to 0} \int_{U(\epsilon)}^\infty [\mathscr{G}(u)]^2\, du = 0. \tag{29}$$

Concerning the second term, we introduce the following function:

$$H^{(\epsilon)}(u) = \left[\overline{\mathscr{G}}(u) - \mathscr{G}(u)\right] \qquad \text{if } 0 \leq u \leq U(\epsilon) \tag{30}$$

$$H^{(\epsilon)}(u) = 0 \qquad \text{if } u > U(\epsilon). \tag{31}$$

Using Schwartz inequality we obtain

$$\int_0^\infty \left[H^{(\epsilon)}(u)\right]^2 du \leq 4C \qquad (\epsilon > 0); \tag{32}$$

moreover we have the following limit

$$\lim_{\epsilon \to 0} H^{(\epsilon)}(u) \to 0 \qquad (u \geq 0). \tag{33}$$

The family of functions $H^{(\epsilon)}(u)$ is bounded in $L^2(0,\infty)$; then it has a subsequence which is weakly convergent in $L^2(0,\infty)$. The limit of this subsequence is zero. We claim that the whole family $\{H^{(\epsilon)}\}$ converges weakly to zero in $L^2(0,\infty)$. In fact, if this statement were not true, then there would exist two subsequences weakly convergent to different limits. But this case can be excluded, since the weak limit of any subsequence of $\{H^{(\epsilon)}\}$ is necessarily zero. Therefore it follows that

$$\lim_{\epsilon\to 0}\int_0^\infty \mathcal{G}(u)H^{(\epsilon)}(u)\,du = 0. \tag{34}$$

Then from formula (28) we derive

$$\lim_{\epsilon\to 0}\int_0^{U(\epsilon)}\left[\bar{\mathcal{G}}(u) - \mathcal{G}(u)\right]^2 du = 0 \tag{35}$$

and therefore

$$\lim_{\epsilon\to 0}\sum_{k=1}^{k_0(\epsilon)}\left(\frac{\bar{g}_k - g_k}{\lambda_k}\right)^2 = 0 \tag{36}$$

and the theorem is proved. ∎

Let us note that the operator defined by Eq. (1) maps the space $X \equiv L^2(a,b)$ into the space $Y \equiv L^2(a,b)$, but the elements $(Af)(x)$ do not fill up Y. Adding to the noiseless data $g(x)$ the noise $n(x)$, we have the functions $\bar{g}(x)$ which do not belong in general to the range of A, but still belong to Y if the noise satisfies some appropriate bound, like that expressed by inequality (9). The set of functions $\bar{g}$, which does not necessarily fill up the space $Y \equiv L^2(a,b)$, shall be denoted hereafter by $\bar{Y}(\bar{Y} \subseteq Y)$. Then we can associate to the approximation (20) an operator $\bar{B}$ defined as follows:

$$\bar{B}\bar{g} = \sum_{k=1}^{k_0(\epsilon)}\left(\frac{\bar{g}_k}{\lambda_k}\right)\psi_k. \tag{37}$$

The operator $\bar{B}$ maps $\bar{Y}$ into X. If $\epsilon = 0$, then

$$\|\bar{B}\bar{g}\|^2_{L^2(a,b)} = \|\bar{B}g\|^2_{L^2(a,b)} = \sum_{k=1}^{\infty}\left(\frac{g_k}{\lambda_k}\right)^2 = C; \tag{38}$$

if $\epsilon > 0$, then $k_0(\epsilon) < \infty$ and

$$\|\bar{B}\bar{g}\|^2_{L^2(a,b)} = \sum_{k=1}^{k_0(\epsilon)} \left(\frac{\bar{g}_k}{\lambda_k} \right)^2 \leq C, \tag{39}$$

and therefore $\bar{B}\bar{g} \in L^2(a, b)$.

Then we have the following corollary of Theorem 1:

COROLLARY. *The mapping* $\bar{B}: \bar{Y} \to X$ (*of* $\bar{Y}$ *into* X) *is continuous*.

Proof. Let us recall that continuous mappings of one metric space into another are precisely those which send convergent sequences into convergent sequences or, in other words, which preserve convergence. Then let us consider a sequence of elements of $\bar{Y}$, i.e., a sequence $\bar{g}_{\epsilon_i} \in \bar{Y}$, which tends to $g \in Y$ for $\epsilon_i \to 0$, in the norm of $\bar{Y}$. This amounts to saying that $\lim_{\epsilon_i \to 0} \|\bar{g}_{\epsilon_i} - g\|_{\bar{Y}} \to 0$. To the sequence $\bar{g}_{\epsilon_i}$ we can associate the sequence $\bar{B}\bar{g}_{\epsilon_i}$ through equality (37).

Then by Theorem 1 it follows that

$$\lim_{\epsilon_i \to 0} \|\bar{B}\bar{g}_{\epsilon_i} - \bar{B}g\|_X = 0 \tag{40}$$

and the Corollary is proved. ∎

Unfortunately this result does not furnish a numerical algorithm able to determine the unknown function f. Indeed we do not know the constant C (i.e., the square of the norm of the function f), and therefore we are not able to find the truncation number $k_0(\epsilon)$ and accordingly to determine the approximation (20). It is possible, however, to work out different methods which are able to solve this problem in several situations of practical interest. One of this procedures is a probabilistic method and it shall be illustrated in the next section; the other is of deterministic nature and it is essentially based on the following remarkable fact: the asymptotic behavior of the eigenvalues λ_k, for large k, is strictly related to the regularity properties of the kernel $K(x, y)$. Hille and Tamarkin [HT] have systematically explored the relationship between the regularity properties of the kernel such as integrability, continuity, differentiability, analyticity, and the like, and the distributions of the eigenvalues of the Fredholm integral equation of the first kind. We will limit ourselves to illustrating the situation with two examples of opposite types:

(a) let there be $K(x, y) = \sin(x - y)/(x - y)$, and $a = -1$, $b = 1$ in Eq. (1); $K(x, y)$ is an entire analytical function and the corresponding

eigenvalues λ_k are characterized by an asymptotic behaviour of the following type: $\lambda_k = O(\exp(-Dk \log k))$ ($D = \text{const}$).

(b) let there be

$$K(x, y) = (1 - x)y \qquad (0 \le y \le x \le 1) \tag{41}$$

$$K(x, y) = x(1 - y) \qquad (0 \le x \le y \le 1) \tag{42}$$

and $a = 0$ and $b = 1$ in Eq. (1); in this case we can explicitly evaluate the eigenvalues λ_k, which are given by $\lambda_k = 1/(k^2\pi^2)$ and decrease only as the inverse of a power.

In conclusion we can say that, as the regularity of the kernel increases passing from the class of functions C^0 to C^∞ and then to the class of analytic entire functions, accordingly the eigenvalues λ_k decrease more and more rapidly for $k \to \infty$. The Hille–Tamarkin analysis allows us to elaborate a method which, in some cases, can be used in order to find out the truncation number $k_0(\epsilon)$ introduced in the approximation (20). To this purpose we need to explore the properties of the following function $M(m) = \sum_{k=1}^{m} (\bar{g}_k/\lambda_k)^2$. The relevant properties of M are:

(1) It is an increasing function of m.

(2) If ϵ is sufficiently small (practically ϵ is smaller than $\|g\|_{L^2(a,b)}$ for several orders of magnitude), and the values of $|g_k|$ are decreasing for increasing values of k, then $M(m)$ presents a plateau when it reaches the value C (i.e., the square of the norm of f). Indeed from statement (iv) of Lemma 1 (i.e., $\lim_{\epsilon \to 0} k_0(\epsilon) = \infty$) it follows that $M(m)$ remains constant (or nearly constant) when it attains the value C.

(3) The plateau discussed under point (2) corresponds to the order–disorder transition in the coefficients $\bar{g}_k = g_k + n_k (n_k = \int_a^b n(x)\psi_k(x)\,dx)$. For $k \le k_0$ the "data" g_k are prevailing on the Fourier coefficients of the noise n_k; for $k > k_0(\epsilon)$ the "noise" n_k is prevailing on the Fourier coefficients of the noiseless data. This property will play a relevant role in the next section.

Now let us suppose that the kernel K has such analytical properties that the eigenvalues λ_k decrease asymptotically like an exponential, i.e., $\lambda_k = O(\exp(-k))$, while the noise coefficients decrease like the inverse of a power, for instance $n_k = O(k^{-1})$. Then the behaviour of $M(m)$ for $m > k_0(\epsilon)$ is dominated by the asymptotic behaviour of the eigenvalues. In the specific example mentioned above, we have for $m > k_0$

$$M(m) \sim C_1 + C_2 \exp(2m)/(m^2) \qquad (C_1, C_2 = \text{const}). \tag{43}$$

Therefore from the analytical properties of the kernel K we can discover, plotting numerically $M(m)$ versus m, the plateau of $M(m)$ corresponding to C and therefore determine the truncation number k_0. Worked-out numerical examples of this procedure have been given by one of the authors in Refs. [Vi, SV1]. This method, however, requires a few words of comment. First let us observe that the numerical plot of $M(m)$ versus m can still present one or several plateau for $m > k_0$. This is essentially due to the erratic behaviour of the noise and accordingly of the Fourier coefficients n_k. This fact could obscure a clear-cut separation between the plateau of $M(m)$, corresponding to the value C, and the others. Furthermore the procedure requires that the eigenvalues λ_k decrease rapidly for $k \to \infty$; therefore for a kernel $K(x, y)$, like that of example (b), the method could be inappropriate. In all the cases when this procedure fails, we have to consider a statistical method like that illustrated in the next section.

REMARKS. (i) Let us recall that (as was remarked in the Introduction) the truncation methods suggested by the standard regularization techniques are obtained by comparing the eigenvalues λ_k with the error bound ϵ, and this procedure is rather unsatisfactory from the viewpoint of experimental sciences. On the contrary, in our case, if one is able to determine the number $k_0(\epsilon)$, it does correspond to the plateau of $M(m)$, which is attained when the Fourier coefficients of noiseless data g_k are comparable in magnitude with n_k (see properties (2) and (3) of $M(m)$). In other words the true data are compared with the noise and this is exactly what is done in the analysis of experimental data.

(ii) Let us observe that the behaviour of $k_0(\epsilon)$ is related to the behaviour of the ratio n_k/λ_k as a function of k and therefore that it depends essentially on the type of noise and on the properties of the kernel $K(x, y)$. Returning once more to the results of Hille and Tamarkin [HT] relating the regularity of the kernel to the behaviour of the eigenvalues λ_k, we must distinguish between the following two situations:

(a) The kernel is an analytic function and the eigenvalues λ_k decrease very rapidly for $k \to \infty$. In such a case n_k/λ_k may increase very fast, for k sufficiently large, and even if we were able to decrease considerably the value of ϵ, the increase of $k_0(\epsilon)$ would be nonetheless quite small. We say that the restored continuity is weak.

(b) The kernel is of class C^0 and accordingly the eigenvalues λ_k do not decrease rapidly for $k \to \infty$ (but, for instance, like the inverse of a power). In this case the increase of $k_0(\epsilon)$, for $\epsilon \to 0$, could be rather fast and the restored continuity is more robust. In this case the number of Fourier coefficients which transmit information increases in a significant way as we lower ϵ (see also the next section for an interpretation of these results from the point of view of information theory).

3. PROBABILISTIC APPROACH AND INFORMATION THEORY

Now we want to formulate a probabilistic version of Theorem 1; therefore we must regard f, n, and $\bar{g}$ as values of random variables in appropriate Hilbert spaces. To this purpose let us briefly fix our notations and ideas on the introduction of random variables (and weak random variables) in Hilbert spaces (see [Ba]).

Let us consider the triplet $(\Omega, \mathscr{F}, P)$, where Ω is an abstract point set (e.g. the set of outcomes ω of an experiment), $\mathscr{F}$ is a σ-algebra of subsets of ω, and P is a measure on $\mathscr{F}$ with $P(\Omega) = 1$. This triplet is a probabilistic space. Let $\mathscr{H}$ be a separable Hilbert space: then a Hilbert space random variable h is an application of Ω on $\mathscr{H}$; $h\colon \Omega \to \mathscr{H}$; i.e., $h(\omega)$ is an element of $\mathscr{H}$.

In order to include processes like white noise, one has to consider weak random variables (w.r.v.). To this purpose we must introduce the cylinder sets in Hilbert spaces. Let us consider any finite-dimensional subspace $\mathscr{H}_m$ in $\mathscr{H}$. By a cylinder set we mean any set of the form $B + \mathscr{H}_m^{\perp}$, where B is a Borel subset of $\mathscr{H}_m$ and $\mathscr{H}_m^{\perp}$ is the orthogonal complement of $\mathscr{H}_m$. Then a function $h(\omega)$ $(\omega \in \Omega)$, which maps Ω into $\mathscr{H}$, is a weak random variable if (i) the inverse image of any cylinder set of $\mathscr{H}$ is a set of the σ-algebra of $\mathscr{F}$; (ii) the probability measure induced on the Borel sets of an arbitrary finite dimensional subspace of $\mathscr{H}$ is countably additive. In other words if $\{u_k\}_1^{\infty}$ is a basis in $\mathscr{H}$, then $(h, u_k)_{\mathscr{H}_m}$ $(k = 1, 2, \ldots, m)$, for each m defines an ordinary random variable.

We shall denote by μ_h the cylinder probability measure induced by h on $C(\mathscr{H})$, where $C(\mathscr{H})$ is the class of the cylinder sets of $\mathscr{H}$; i.e., if $\Gamma \in C(\mathscr{H})$, then $\mu_h(\Gamma) = P\{\omega\colon h(\omega) \in \Gamma\}$.

A w.r.v. $h\colon \Omega \to \mathscr{H}$ is called Gaussian (or normal) if, for any element $w \in \mathscr{H}$, the random variable $h_w = (h, w)_{\mathscr{H}}$ is Gaussian. A Gaussian w.r.v. is uniquely defined by its mean element m_h and covariance operator R_{hh}. The mean element is the unique vector of $\mathscr{H}$ such that, for any $w \in \mathscr{H}$,

$$(m_h, w)_{\mathscr{H}} = E\{(h, w)_{\mathscr{H}}\} = \int_{\Omega} (h(\omega), w)_{\mathscr{H}} P(d\omega). \tag{44}$$

The covariance operator R_{hh} is the unique, bounded, linear, self-adjoint, non-negative operator on $\mathscr{H}$ such that, for any $w \in \mathscr{H}$,

$$(R_{hh} w, w)_{\mathscr{H}} = E\{|(h - m_h, w)_{\mathscr{H}}|^2\} = \int_{\Omega} |(h(\omega) - m_h, w)_{\mathscr{H}}|^2 P(d\omega). \tag{45}$$

The Gaussian cylinder measure μ_h can be extended to be countably additive on $B(\mathscr{H})$ (where $B(\mathscr{H})$ is the class of Borel sets of $\mathscr{H}$) iff R_{hh} is an operator of the trace class, i.e.,

$$E\{\|h - m_h\|_{\mathscr{H}}^2\} = \mathrm{Tr}(R_{hh}) = \sum_{k=1}^{\infty} (R_{hh}u_k, u_k)_{\mathscr{H}} < \infty, \tag{46}$$

where $\{u_k\}_1^\infty$ is an arbitrary orthonormal basis in $\mathscr{H}$. Any cylindrical measure with this property will be termed countably additive.

As a typical example of w.r.v. we keep in mind the white noise, a Gaussian process, which we denote by ζ. Its covariance operator $R_{\zeta\zeta}$ is given by $R_{\zeta\zeta} = \epsilon^2 I$, I being the identity operator in $\mathscr{H}$. Each Fourier component $\zeta_k = (\zeta, u_k)_{\mathscr{H}}, \{u_k\}_1^\infty$ being an orthonormal basis in $\mathscr{H}$, is an ordinary random variable of zero mean and ϵ^2 variance.

As we said above, we regard f, n, and $\bar{g}$ as values of the w.r.v. ξ, ζ, and η, respectively. Then Equation (5) must be written as

$$A\xi + \zeta = \eta. \tag{47}$$

We assume hereafter that the w.r.v. ξ and ζ are Gaussian and independent and that they have zero mean. Recall now that the eigenfunctions $\{\psi_k\}_1^\infty$ of the operator A form an orthonormal basis of $L^2(a, b)$. Then the Fourier components $\xi_k = (\xi, \psi_k)_{L^2(a,b)}$ and $\zeta_k = (\zeta, \psi_k)_{L^2(a,b)}$ are ordinary random variables with zero mean and variance ρ_k^2 and $\epsilon^2\nu_k^2$, respectively. At this point let us remark that in the following we shall essentially work with the trace of operators (like the covariance operators), which is independent on the particular basis chosen. Therefore we can very well use as a basis the eigenfunctions $\{\psi_k\}_1^\infty$ of the operator A. From the assumption that ξ and ζ are uncorrelated and have zero mean it follows that the convariance operator $R_{\eta\eta}$ of η is given by (see also [Fr])

$$R_{\eta\eta} = AR_{\xi\xi}A + R_{\zeta\zeta}, \tag{48}$$

where $R_{\xi\xi}$ and $R_{\zeta\zeta}$ are the covariance operators of ξ and ζ, respectively. Moreover the w.r.v. ξ and η are not independent and their cross-covariance operator is given by (see [Fr])

$$R_{\xi\eta} = E[\xi, \eta^*] = E[\xi, \xi^*A^* + \zeta^*] = R_{\xi\xi}A^* = R_{\xi\xi}A. \tag{49}$$

Finally we write $R_{\zeta\zeta}$ in the form

$$R_{\zeta\zeta} = \epsilon^2 N, \tag{50}$$

where N is a given operator (it is the identity I in the case of white noise) and ϵ is a parameter that tends to zero when the noise vanishes. 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.

We solve this problem under the additional assumption that the covariance operator $R_{\xi\xi}$ is of the trace class; i.e.,

$$\operatorname{Tr} R_{\xi\xi} = \sum_{k=1}^{\infty} \left(R_{\xi\xi}\psi_k, \psi_k\right) = \sum_{k=1}^{\infty} \rho_k^2 = \Gamma < \infty. \tag{51}$$

Remark. Let us focus our attention on the following point: we are obliged to assume that $\operatorname{Tr} R_{\xi\xi}$ is finite; nevertheless we do not require a knowledge of the constant Γ, which plays a role similar (but not identical) to that of the constant $C = \|f\|_{L^2}^2$ introduced in Section 1.

If B: $L^2(a,b) \to L^2(a,b)$ is any linear continuous operator, we call $\xi_B = B\eta$ a linear estimate of the random variable ξ. Then the reliability of the estimate is measured by

$$E\{\|\xi - B\eta\|^2\} = \operatorname{Tr}\left(R_{\xi\xi} - R_{\xi\xi}AB^* - BAR_{\xi\xi} + BR_{\eta\eta}B^*\right), \tag{52}$$

where Eq. (49) has been used.

Now let us consider the class of linear bounded operators $\hat{B}$: $L^2(a,b) \to L^2(a,b)$ defined by

$$\hat{B}\bar{g} = \sum_{k\in I} \frac{\bar{g}_k}{\lambda_k}\psi_k, \tag{53}$$

where $\bar{g}_k = (\bar{g}, \psi_k)_{L^2}$ and I is an arbitrary finite set.

Next we introduce the set I_α, which is defined as

$$I_\alpha\{k \le k_\alpha(\epsilon)\}, \tag{54}$$

where

$$k_\alpha(\epsilon) = \max\left\{m \in \mathbf{N}: \sum_{k=1}^{m}\left(\rho_k^2 + \frac{\epsilon^2\nu_k^2}{\lambda_k^2}\right) \le \Gamma\right\}. \tag{55}$$

When $\hat{B}$ is given by $\hat{B}_\alpha$ (i.e., the operator $\hat{B}$ associated with the set I_α), Eq. (52) reads

$$E\{\|\xi - \hat{B}_\alpha \eta\|^2\} = \sum_{k=k_\alpha(\epsilon)+1}^{\infty} \rho_k^2 + \sum_{k=1}^{k_\alpha(\epsilon)} \left(\frac{\epsilon \nu_k}{\lambda_k}\right)^2, \tag{56}$$

where Eq. (49) has been used.

Now we may prove the following lemma.

LEMMA 2. *The following equalities hold true*:

$$\text{(i)} \quad \lim_{\epsilon \to 0} k_{\alpha(\epsilon)} = +\infty \tag{57}$$

$$\text{(ii)} \quad \lim_{\epsilon \to 0} \left\{ \sum_{k=k_\alpha(\epsilon)+1}^{\infty} \rho_k^2 + \sum_{k=1}^{k_\alpha(\epsilon)} \left(\frac{\epsilon \nu_k}{\lambda_k}\right)^2 \right\} = 0. \tag{58}$$

Proof. (i) The proof of (i) is strictly similar to the proof of statement (iv) of Lemma 1. Indeed if we assume that this statement is not true, we have a contradiction. In fact let us denote by k_{α_1} the sum $k_{\alpha_1} = k_\alpha + 1$. If equality (57) is not true, then there should exist a finite number M, which does not depend on ϵ and such that, for any sequence $\{\epsilon_i\}$ converging to zero, $k_{\alpha_1}(\epsilon_i) < M$. From formula (55) it follows that

$$\Gamma < \sum_{k=1}^{k_{\alpha_1}(\epsilon_i)} \left(\rho_k^2 + \frac{\epsilon^2 \nu_k^2}{\lambda_k^2}\right) \leq \sum_{k=1}^{M} \left(\rho_k^2 + \frac{\epsilon^2 \nu_k^2}{\lambda_k^2}\right). \tag{59}$$

For $\epsilon_i \to 0$, we have

$$\Gamma < \sum_{k=1}^{M} \rho_k^2 \leq \sum_{k=1}^{\infty} \rho_k^2 = \Gamma \tag{60}$$

and the contradiction is explicit.

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

$$\lim_{\epsilon \to 0} \sum_{k=k_\alpha(\epsilon)+1}^{\infty} \rho_k^2 = 0. \tag{61}$$

Regarding the term $\sum_{k=1}^{k_\alpha(\epsilon)} (\epsilon \nu_k / \lambda_k)^2$ we can proceed as follows: from formula (55) we have

$$\sum_{k=1}^{k_\alpha(\epsilon)} \left(\frac{\epsilon \nu_k}{\lambda_k}\right)^2 + \sum_{k=1}^{k_\alpha(\epsilon)} \rho_k^2 \leq \Gamma = \sum_{k=1}^{\infty} \rho_k^2 \tag{62}$$

and therefore

$$\sum_{k=1}^{k_\alpha(\epsilon)} \left(\frac{\epsilon \nu_k}{\lambda_k} \right)^2 \leq \sum_{k=k_\alpha(\epsilon)+1}^{\infty} \rho_k^2. \tag{63}$$

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

We are looking for a truncation point, which shall be denoted hereafter by k_t, which corresponds to the order–disorder transition point in the Fourier coefficients $\bar{g}_k$ (see property (3) of the function $M(m)$ in Section 2). Unfortunately we do not have any evidence that the truncation point k_α coincides with k_t. A quantitative criterium can be derived from the information theory. Indeed we know that the amount of information contained in the random variable $\eta_k = (\eta, \psi_k)$ regarding the variable $\xi_k = (\xi, \psi_k)$ is given by [GY]

$$J_k(\xi_k, \eta_k) = -\tfrac{1}{2} \log(1 - r_k^2), \tag{64}$$

where

$$r_k^2 = \frac{|E[\xi_k, \eta_k]|^2}{E\left[|\xi_k|^2\right] E\left[|\eta_k|^2\right]} = \frac{(\lambda_k \rho_k)^2}{(\lambda_k \rho_k)^2 + (\epsilon \nu_k)^2} \tag{65}$$

and finally

$$J_k(\xi_k, \eta_k) = \tfrac{1}{2} \log\left(1 + \frac{(\lambda_k \rho_k)^2}{(\epsilon \nu_k)^2}\right). \tag{66}$$

From (66) it follows that if $\lambda_k \rho_k < \epsilon \nu_k$, then $J_k(\xi_k, \eta_k) < \frac{1}{2} \log 2$ and the amount of information on ξ_k contained in η_k is very small. Therefore we can conclude that the transition point k_t is the value of k such that

$$\text{for } k \leq k_t : \lambda_k \rho_k > \epsilon \nu_k \tag{67}$$

$$\text{for } k > k_t : \lambda_k \rho_k < \epsilon \nu_k. \tag{68}$$

The Hilbert–Schmidt theorem guarantees that this value certainly does exist in the case of white noise (i.e., $\nu_k = 1, \forall k$), supposing that not only the eigenvalues λ_k but also the terms ρ_k are ordered in a monotonous sequence. We can say that this transition point exists also in the case of colored noise and more generally even if the covariance operator of the noise is of trace class, if $\lim_{k \to \infty} \lambda_k \rho_k / \nu_k = 0$, assuming once more that also the terms ν_k are ordered in a monotonous sequence.

Now we can associate with the transition point k_t a set I_t defined as $I_t = \{k \leq k_t(\epsilon)\}$. Accordingly we can introduce the operator $\hat{B}_t$ which can be defined in the following way:

$$\hat{B}_t \bar{g} = \sum_{k \in I_t} \frac{\bar{g}_k}{\lambda_k} \psi_k. \tag{69}$$

Then we can compute the estimate $E[\| \xi - \hat{B}_t \eta \|^2]$, obtaining

$$E\left[\| \xi - \hat{B}_t \eta \|^2\right] = \sum_{k=k_t+1}^{\infty} \rho_k^2 + \sum_{k=1}^{k_t} \left(\frac{\epsilon \nu_k}{\lambda_k} \right)^2. \tag{70}$$

Next we prove the following theorem.

THEOREM 2. *The following limit holds true*:

$$\lim_{\epsilon \to 0} \left\{ \sum_{k=k_t(\epsilon)+1}^{\infty} \rho_k^2 + \sum_{k=1}^{k_t(\epsilon)} \left(\frac{\epsilon \nu_k}{\lambda_k} \right)^2 \right\} = 0. \tag{71}$$

Proof. The proof proceeds in two steps

(a) We want to prove that $\lim_{\epsilon \to 0} \sum_{k=k_t(\epsilon)+1}^{\infty} \rho_k^2 = 0$. Now either $k_t \geq k_\alpha$ or $k_t < k_\alpha$. In the first case the statement follows from the fact that $\lim_{\epsilon \to 0} \sum_{k=k_\alpha+1}^{\infty} \rho_k^2 = 0$. On the contrary if $k_t < k_\alpha$, then we have

$$\begin{aligned} \sum_{k=k_t(\epsilon)+1}^{\infty} \rho_k^2 &\leq \sum_{k=k_\alpha+1}^{\infty} \rho_k^2 + \sum_{k=k_t(\epsilon)+1}^{k_\alpha(\epsilon)} \left(\frac{\epsilon \nu_k}{\lambda_k} \right)^2 \\ &\leq \sum_{k=k_\alpha+1}^{\infty} \rho_k^2 + \sum_{k=1}^{k_\alpha(\epsilon)} \left(\frac{\epsilon \nu_k}{\lambda_k} \right)^2. \end{aligned} \tag{72}$$

But in Lemma 2 we have proved that the r.h.s. of formula (72) tends to zero as $\epsilon \to 0$, and then the statement follows.

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

$$\sum_{k=k_{\alpha(\epsilon)+1}}^{k_t(\epsilon)} \left(\frac{\epsilon \nu_k}{\lambda_k} \right)^2 \leq \sum_{k=k_{\alpha(\epsilon)+1}}^{k_t(\epsilon)} \rho_k^2 \leq \sum_{k=k_{\alpha(\epsilon)+1}}^{\infty} \rho_k^2. \tag{73}$$

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

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

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

Theorem 2 can be regarded as the probabilistic version of Theorem 1, and accordingly the operator $\hat{B}_t$ corresponds to the operator $\bar{B}$ introduced in Section 2. This fact allows us to use statistical methods for the identification of the plateau of $M(m)$ which corresponds to the order–disorder transition in the coefficients $\bar{g}_k$. Indeed we can separate the coefficients $\bar{g}_k$ into two classes:

(a) the Fourier coefficients $\bar{g}_k$ $(k \leq k_t)$, from which a significant amount of information can be extracted;

(b) the Fourier coefficients $\bar{g}_k$ $(k > k_t)$, which can be regarded as random numbers, since the noise is prevailing on the information.

A remarkable statistical method which can be used is the so-called "correlogram method." Let us suppose that we know a certain number N of Fourier coefficients $\bar{g}_k$. Then we can compute the correlation between pairs of them, n units apart, as [KS]

$$\delta_n = \frac{\sum_{k=1}^{N-n} \left(\bar{g}_k - \bar{\bar{g}}_k \right) \left(\bar{g}_{k+n} - \bar{\bar{g}}_{k+n} \right)}{\left\{ \sum_{k=1}^{N-n} \left(\bar{g}_k - \bar{\bar{g}}_k \right)^2 \sum_{k=1}^{N-n} \left(\bar{g}_{k+n} - \bar{\bar{g}}_{k+n} \right)^2 \right\}^{1/2}}, \tag{74}$$

where $n = 0, 1, 2, \ldots, N-1$ and

$$\bar{\bar{g}}_k = \frac{1}{N-n} \sum_{k=1}^{N-n} \bar{g}_k \tag{75}$$

$$\bar{\bar{g}}_{k+n} = \frac{1}{N-n} \sum_{k=1}^{N-n} \bar{g}_{k+n}. \tag{76}$$

Then one plots $|\delta_n|$ versus n. In a series of uncorrelated random numbers all the terms δ_n, apart from $\delta_0 = 1$, are equal to zero within sampling limits. Therefore one can use the departure of the correlation from zero to test the departure of the terms $\bar{g}_k$ from the randomness. Only those values of $|\delta_n|$ that are lying outside a strip of width $2/\sqrt{N}$ (N = total number of sampling data) must be regarded as different from zero. Let us assume that δ_n drops to zero, within statistical accuracy, for a certain value n_0 of n (i.e., $|\delta_n| > 2/\sqrt{N}$ for $n < n_0$, $|\delta_n| < 2/\sqrt{N}$ for $n > n_0$);

then we can identify the truncation point $k_t(\epsilon)$ with $n_0 + 1$. This method has been illustrated in [SV2], with several numerical examples in the specific case of an integral equation whose eigenvalues λ_k are decreasing as the inverse of a power (i.e., $\lambda_k = 1/(k^2\pi^2)$): the kernel of this equation is exactly that considered in Section 2 (see example (b)). Let us stress once more that this statistical procedure does not require any *a priori* information on the solution f.

REFERENCES

[Ba] A. V. Balakrishnan, "*Applied Functional Analysis*", Chapter 6, Springer, New York, 1976.

[BDV] M. Bertero, C. De Mol, and G. A. Viano, The stability of inverse problems, *in* "Inverse Scattering Problems in Optics," Springer-Verlag, Berlin/New York, 1980.

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

[GY] 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. Ser.* 2 **12** (1959), 199–246.

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

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

[Ke] J. Kelley, "General Topology," Van Nostrand, Princeton, 1955.

[KS] M. G. Kendall and A. Stuart, "The Advanced Theory of Statistics," Vol. 3, Griffin, London, 1966.

[SV1] E. Scalas and G. A. Viano, The Hausdorff moments in statistical mechanics, *J. Math. Phys.* **34**, No. 12 (1993), 5781–5800.

[SV2] E. Scalas and G. A. Viano, Resolving power and information theory in signal recovery, *J. Opt. Soc. Amer. A.* **10**, No. 5 (1993), 991–996.

[TA] A. Tichonov and V. Arsenine, "*Méthodes de Rèsolution de Problémes Mal Posès*," Mir, Moscow, 1976.

[Vi] G. A. Viano, Solution of the Hausdorff moment problem by the use of Pollaczek polynomials, *J. Math. Anal. Appl.* **156** (1991), 410–427.